NEW CONSTRUCTIONS OF SATISFACTION CLASSES

ALI ENAYAT AND ALBERT VISSER

Abstract. We use model-theoretic ideas to present a perspicuous and versatile method of constructing full satisfaction classes on models of Peano arithmetic. We also comment on the ramifications of our work on issues related to conservativity and interpretability.

1. Introduction

In our forthcoming paper [2] we explore satisfaction classes over a wide variety of 'base theories' ranging from weak fragments of arithmetic to systems of ZF set theory and beyond. This note provides a synopsis of some of this work in the context of the most popular base theory adopted in investigations of axiomatic theories of truth, namely PA (Peano Arithmetic).

The notion of a satisfaction class was first introduced and investigated by Krajewski in his 1976 paper [11]. Two noteworthy accomplishments of [11] are the following results:

(1) If a countable model of a 'base theory' (such as PA) carries at least one full satisfaction class, then it carries continuum-many full satisfaction classes.

(2) Every model of ZF has an elementary extension that carries a full satisfaction class.

The question whether the analogue of (2) holds for PA remained open until the appearance of the joint work [10] of Kotlarski, Krajewski, and Lachlan in 1981, in which the rather exotic proof-theoretic technology of '$\mathcal{M}$-logic' (an infinitary logical system based on a nonstandard model $\mathcal{M}$), was invented to construct 'truth classes' over countable recursively saturated models of PA.[1] This model-theoretic result can be used to show that the analogue of (2) does indeed hold for PA, which in turn can be used to show that $\mathsf{PA}^{\mathsf{FT}}$

Date: November 24, 2012.

This research was partially supported by a grant from the Descartes Center of Utrecht University, which supported the first author's visit to Utrecht to work closely with the second author.

[1] As explained in Section 4, a truth class is essentially a well-behaved kind of satisfaction class. The $\mathcal{M}$-logic methodology was further elaborated to establish refined constructions of full truth classes by Smith [15,16,17], Kaye [9], and Engström [3].

is conservative over PA, where $\mathsf{PA}^{\mathsf{FT}} = \mathsf{PA} +$ "T is a full truth class". The conservativity of $\mathsf{PA}^{\mathsf{FT}}$ over PA has attracted considerable philosophical attention, especially in relation to the grand debate concerning deflationism.[2]

In this paper we present a perspicuous method for the construction of full satisfaction classes that is dominantly based on *model-theoretic techniques* (e.g., expanding the language, compactness, and elementary chains). As we shall see, our construction method is quite versatile and can be used to construct many (if not all) of the results that have hitherto been only possible to establish with the use of $\mathcal{M}$-logic machinery. Furthermore, the method can also be employed to build new types of full satisfaction classes (see Section 6).

We present the necessary preliminaries in Section 2, and then in Section 3 we concentrate on the basic form of our new construction of full satisfaction classes, where it is used to show that every model of PA has an elementary extension that carries a full satisfaction class. The versatility of the methodology of Section 3 is illustrated in Section 4, in which an appropriate modification of the method is used to construct truth classes for models of PA. As explained in Section 5, certain arithmetizations of our construction can also be employed to establish that (1) $\mathsf{PA}^{\mathsf{FT}}$ is interpretable in PA; and (2) the conservativity of $\mathsf{PA}^{\mathsf{FT}}$ over PA can be verified in PRA (Primitive Recursive Arithmetic). Finally, in Section 6 we briefly describe further applications of the methods introduced in this paper.

Acknowledgments. We are grateful to the editors of this volume for their interest in our work. Thanks also to Volker Halbach, Fredrik Engström, and James Schmerl for helpful feedback on preliminary drafts of this paper. We are particularly indebted to Schmerl for catching an inaccuracy in an earlier formulation of Lemma 3.1, and for his suggestion to distill the results of our paper [2] in this form for wider dissemination.

2. Preliminaries

2.1. Definition. Throughout this paper PA refers to Peano arithmetic formulated in a *relational language* $\mathcal{L}_{\mathsf{PA}}$ using the logical constants $\{\neg, \vee, \exists, =\}$. Note that in this formulation PA has no constant symbols; the arithmetical operations of addition and multiplication are construed as ternary relations; and conjunction, universal quantification, and other logical constants are taken as defined notions in the usual way.

[2] A recent noteworthy paper in this connection is McGee's [13].

It is well known that PA has more than sufficient expressive machinery to handle syntactic notions. The following list of $\mathcal{L}_{\mathsf{PA}}$-formulae will be useful here.[3]

- $\mathsf{Form}(x)$ is the formula expressing "x is the code of an $\mathcal{L}_{\mathsf{PA}}$-formula using variables $\{v_i : i \in \mathbb{N}\}$, and the non-logical symbols available in $\mathcal{L}_{\mathsf{PA}}$".
- $\mathsf{Asn}(x)$ is the formula expressing "x is the code of an assignment", where an assignment here simply refers to a function whose domain consists of a (finite) set of variables. We use α and its variants (α', α_0, etc.) to range over assignments.
- $y \in \mathsf{FV}(x)$ is the formula expressing "$\mathsf{Form}(x)$ and y is a free variable of x".
- $y \in \mathsf{Dom}(\alpha)$ is the formula expressing "the domain of α includes y".
- $\mathsf{Asn}(\alpha, x)$ is the following formula expressing "α is an assignment for x":

$$(\mathsf{Form}(x) \wedge \mathsf{Asn}(\alpha) \wedge \forall y(y \in \mathsf{Dom}(\alpha) \leftrightarrow y \in \mathsf{FV}(x))\,.$$

- $x \lhd y$ is the formula expressing "x is the code of an immediate subformula of the $\mathcal{L}_{\mathsf{PA}}$-formula coded by y", i.e., $x \lhd y$ abbreviates the conjunction of $\mathsf{Form}(y)$ and the following disjunction:

$$(y = \neg x) \vee \exists z\,((y = x \vee z) \vee (y = z \vee x)) \vee \exists i\,(y = \exists v_i\ x)\,.^4$$

The theory $\mathsf{PA}^{\mathsf{FS}}$ (read as "PA with full satisfaction") is formulated in an *expansion* of the language $\mathcal{L}_{\mathsf{PA}}$ by adding a new *binary* predicate $\mathsf{S}(x, y)$. The binary/unary distinction is of course not an essential one since PA has access to a definable pairing function. However, the binary/unary distinction *at the conceptual level* marks the key difference between satisfaction classes and truth classes (the latter are discussed in Section 4). $\mathsf{PA}^{\mathsf{FS}}$ is defined below with the help of a collection of sentences $\mathsf{Tarski}(\mathsf{S}, \mathsf{F})$.

When reading the definition below it is helpful to bear in mind that $\mathsf{Tarski}(\mathsf{S}, \mathsf{F})$ expresses:

> F is a subset of Form that is closed under immediate subformulae; each member of S is an ordered pair of the form (x, α), where $x \in \mathsf{F}$ and α is an assignment for x; and S satisfies Tarski's compositional clauses.

[3] All of the formulae in the list can be arranged to be Σ_1-formulae in the sense of Definition 2.4.

[4] Technically speaking, this formula should be written so as to distinguish the logical operations of the meta-langauge with those of the object-language. For example, using Feferman's commonly used 'dot-convention', one would write:

$$\left(y = \dot{\neg} x\right) \vee \exists z\left(\left(y = x \mathbin{\dot{\vee}} z\right) \vee \left(y = z \mathbin{\dot{\vee}} x\right)\right) \vee \exists i\left(y = \dot{\exists} v_i\ x\right).$$

However, since the difference between the two kinds of operation will be always clear from the context, we have opted for the lighter notation.

2.2. Definition. $\mathsf{PA}^{\mathsf{FS}}{:=}\ \mathsf{PA} \cup \mathsf{Tarski}\,(\mathsf{S}, \mathsf{Form})$, where $\mathsf{Tarski}(\mathsf{S}, \mathsf{F})$ is the conjunction of the universal generalizations of the formulae $\mathsf{tarski}_0(\mathsf{S}, \mathsf{F})$ through $\mathsf{tarski}_4(\mathsf{S}, \mathsf{F})$ described below, all of which are formulated in $\mathcal{L}_{\mathsf{PA}} \cup \{\mathsf{F}(\cdot), \mathsf{S}(\cdot, \cdot)\}$, where S and F do not appear in $\mathcal{L}_{\mathsf{PA}}$.

In the following formulae R ranges over the *relations in* $\mathcal{L}_{\mathsf{PA}}$; $t, t_0, t_1, \cdots$ are *metavariables*, e.g, we write $R\,(t_0, \cdots, t_{n-1})$ instead of $R\left(v_{i_0}, \cdots, v_{i_{n-1}}\right)$; and $\alpha' \supseteq \alpha$ abbreviates
$$(\mathsf{Dom}(\alpha') \supseteq \mathsf{Dom}(\alpha)) \wedge \forall t \in \mathsf{Dom}(\alpha)\ \alpha(t) = \alpha'(t).$$

- $\mathsf{tarski}_0(\mathsf{S}, \mathsf{F}) := (\mathsf{F}(x) \to \mathsf{Form}(x)) \wedge (\mathsf{S}(x, \alpha) \to (\mathsf{F}(x) \wedge \mathsf{Asn}(\alpha, x))) \wedge (y \lhd x \wedge \mathsf{F}(x) \to \mathsf{F}(y))\,.$
- $\mathsf{tarski}_{1,R}(\mathsf{S}, \mathsf{F}) :=$
$$\left(\mathsf{F}(x) \wedge (x = {}^{\ulcorner} R\,(t_0, \cdots, t_{n-1})^{\urcorner}) \wedge \mathsf{Asn}(\alpha, x) \wedge \bigwedge_{i<n} \alpha(t_i) = a_i\right) \to (\mathsf{S}(x, \alpha) \leftrightarrow R\,(a_0, \cdots, a_{n-1}))\,.$$
- $\mathsf{tarski}_2(\mathsf{S}, \mathsf{F}) := (\mathsf{F}(x) \wedge (x = \neg y) \wedge \mathsf{Asn}(\alpha, x)) \to (\mathsf{S}(x, \alpha) \leftrightarrow \neg \mathsf{S}(y, \alpha))\,.$
- $\mathsf{tarski}_3(\mathsf{S}, \mathsf{F}) := (\mathsf{F}(x) \wedge (x = y_1 \vee y_2) \wedge \mathsf{Asn}(\alpha, x)) \to (\mathsf{S}(x, \alpha) \leftrightarrow (\mathsf{S}\,(y_1, \alpha \restriction \mathsf{FV}(y_1)) \vee \mathsf{S}\,(y_2, \alpha \restriction \mathsf{FV}(y_2))))\,.$
- $\mathsf{tarski}_4(\mathsf{S}, \mathsf{F}) := (\mathsf{F}(x) \wedge (x = \exists t\ y) \wedge \mathsf{Asn}(\alpha, x)) \to (\mathsf{S}(x, \alpha) \leftrightarrow \exists \alpha' \supseteq \alpha\ \mathsf{S}(y, \alpha'))\,.$

2.3. Definition. Suppose $\mathcal{M} \models \mathcal{L}_{\mathsf{PA}}$, $F \subseteq M$, and S is a binary relation on M.[5]

(a) S is an *F-satisfaction class* if $(\mathcal{M}, S, F) \models \mathsf{Tarski}(\mathsf{S}, \mathsf{F})$.[6]

(b) Let $\omega_{\mathcal{M}}$ be the well-founded initial segment of $\mathcal{M}$ that is isomorphic to the ordinal ω. We say that F is the set of *standard* $\mathcal{L}_{\mathsf{PA}}$-formulae of $\mathcal{M}$ if
$$F = \mathsf{Form}^{\mathcal{M}} \cap\ \omega_{\mathcal{M}}.$$
In this case there is a unique F-satisfaction class on $\mathcal{M}$, known as the *Tarskian satisfaction class on* $\mathcal{M}$.

(c) S is a *full satisfaction class on* $\mathcal{M}$ if S is an *F-satisfaction class* for $F := \mathsf{Form}^{\mathcal{M}}$. This is equivalent to $(\mathcal{M}, S) \models \mathsf{PA}^{\mathsf{FS}}$.

2.4. Definition. $\Sigma_0 = \Pi_0 =$ the collection of $\mathcal{L}_{\mathsf{PA}}$-formulae all of whose quantifiers are of the form $\exists x < y\ \varphi$ or $\forall x < y\ \varphi$; Σ_{n+1} consists of formulae of the form $\exists x_0 \cdots \exists x_{k-1}\ \varphi$, where $\varphi \in \Pi_n$; and Π_{n+1} consists of formulae of the form $\forall x_0 \cdots \forall x_{k-1}\ \varphi$, where $\varphi \in \Sigma_n$. Here k ranges over ω, with

[5]Throughout the paper we use the convention of using M, M_0, N, etc. to denote the universes of discourse of structures $\mathcal{M}$, $\mathcal{M}_0$, $\mathcal{N}$, etc.

[6]Note that the closure of F under direct subformulae does not guarantee that F should also contain 'infinitely deep' subformulae of a nonstandard formula in F.

the understanding that $k = 0$ corresponds to an empty block of quantifiers; this convention leads to the pleasant consequence that $\Sigma_n \subseteq \Sigma_{n+1}$ and $\Pi_n \subseteq \Pi_{n+1}$ for all n.

2.5. Theorem. (Mostowski [9], [6]) *For each nonzero $n < \omega$ there is a binary Σ_n-formula $\mathsf{Sat}_n(x, y)$ such that*

$$\mathsf{PA} \vdash \mathsf{Tarski}(\mathsf{Sat}_n, \mathbf{\Sigma}_n),$$

where $\mathbf{\Sigma}_n$ is (the arithmetization of) the set of codes of formulae in Σ_n.

3. The Basic Construction

In this section we explain the basic methodology of building satisfaction classes using tools from model theory. The following lemma lies at the heart of the main result of this section.

3.1. Lemma. *Let $\mathcal{N}_0 \models \mathsf{PA}$, $F_1 := \mathsf{Form}^{\mathcal{N}_0}$, $F_0 \subseteq F_1$, and suppose S_0 is an F_0-satisfaction class. Then there is an elementary extension $\mathcal{N}_1$ of $\mathcal{N}_0$ that carries an F_1-satisfaction class $S_1 \supseteq S_0$ and $(c, \alpha) \in S_0$ whenever $c \in F_0$, $\alpha \in N_0$, and $(c, \alpha) \in S_1$.*

Proof: Let $\mathcal{L}^+_{\mathsf{PA}}(\mathcal{N}_0)$ be the language obtained by enriching $\mathcal{L}_{\mathsf{PA}}$ with constant symbols for each member of N_0, and new *unary* predicates U_c for each $c \in \mathsf{Form}^{\mathcal{N}_0}$. It helps to have in mind that the intended interpretation of U_c is $\{\alpha \in A_c : S_1(c, \alpha)\}$, where $A_c := \{\alpha : \mathcal{N}_1 \models \mathsf{Asn}(\alpha, c)\}$.

We first wish to describe a new set of axioms

$$\Theta := \{\theta_c : c \in F_1\}$$

formulated in $\mathcal{L}^+_{\mathsf{PA}}(\mathcal{N}_0)$, where θ_c stipulates 'local Tarskian behavior' for U_c.

If $R \in \mathcal{L}_{\mathsf{PA}}$ and $\mathcal{N}_0 \models c = \ulcorner R(t_0, \cdots, t_{n-1})\urcorner$, then

$$\theta_c := \forall \alpha \, (\mathsf{U}_c(\alpha) \leftrightarrow \mathsf{Asn}(\alpha, c) \wedge R\,(\alpha(t_0), \cdots, \alpha(t_{n-1}))).$$

If $\mathcal{N}_0 \models c = \neg d$, then

$$\theta_c := \forall \alpha \, (\mathsf{U}_c(\alpha) \leftrightarrow \mathsf{Asn}(\alpha, c) \wedge \neg \mathsf{U}_d(\alpha)) \, .$$

If $\mathcal{N}_0 \models c = d_1 \vee d_2$, then

$$\theta_c := \forall \alpha \, (\mathsf{U}_c(\alpha) \leftrightarrow \mathsf{Asn}(\alpha, c) \wedge (\mathsf{U}_{d_1}(\alpha \restriction \mathsf{FV}(d_1)) \vee \mathsf{U}_{d_2}(\alpha \restriction \mathsf{FV}(d_2)))) \, .$$

If $\mathcal{N}_0 \models c = \exists v_a \; b$, then

$$\theta_c := \forall \alpha \, (\mathsf{U}_c(\alpha) \leftrightarrow \mathsf{Asn}(\alpha, c) \wedge \exists \alpha' \supseteq \alpha \; \mathsf{U}_b(\alpha') \wedge \mathsf{Asn}(\alpha', b)) \, .$$

Let

$$\Gamma := \{\mathsf{U}_c(\alpha) : c \in F_0 \text{ and } (c,\alpha) \in S_0\} \cup \{\neg\mathsf{U}_c(\alpha) : c \in F_0 \text{ and } (c,\alpha) \notin S_0\},$$

and let

$$\mathsf{Th}^+(\mathcal{N}_0) := \mathsf{Th}(\mathcal{N}_0, a)_{a\in N_0} \cup \Theta \cup \Gamma.$$

We now proceed to show that $\mathsf{Th}^+(\mathcal{N}_0)$ is consistent by demonstrating that each finite subset of $\mathsf{Th}^+(\mathcal{N}_0)$ is interpretable in $(\mathcal{N}_0, S_0)$. To this end, suppose T_0 is a finite subset of $\mathsf{Th}^+(\mathcal{N}_0)$ and let C consist of the collection of $c \in F_0$ such that U_c appears in T_0. If $C = \emptyset$, T_0 is readily seen to be consistent, so we shall assume that $C \neq \emptyset$ for the rest of the argument.

Our goal is to construct subsets $\{U_c : c \in C\}$ of N_0 such that the following two conditions hold when U_c is interpreted by U_c:

(1) $(\mathcal{N}_0, U_c)_{c\in C} \models \{\theta_c : c \in C\}$, and

(2) For $c \in C \cap F_0$, $U_c = \{\alpha \in N_0 : (c,\alpha) \in S_0\}$.

We shall construct $\{U_c : c \in C\}$ *in stages*, beginning with the simplest formulae in C, and working our way up using Tarski rules for more complex ones. Recall that $c \lhd d$ expresses "c is a direct subformula of d". Define $\lhd^*$ on C by:

$$c \lhd^* d \text{ iff } (c \lhd d)^{\mathcal{N}_0} \text{ and } \theta_d \in T_0 \cap \Theta.$$

Note that whenever $c \lhd^* d$, then for all $c' \lhd d$ we have $c' \in C$ and $c' \lhd^* d$. The finiteness of C implies that $(C, \lhd^*)$ is well-founded, which in turn helps us define a useful measure of complexity for $c \in C$ using the following recursive definition:

$$\mathsf{rank}_C(c) := \sup\{\mathsf{rank}_C(d) + 1 : d \in C \text{ and } d \lhd^* c\}.$$

Note that for $c \in C$, $\mathsf{rank}_C(c) = 0$ precisely when $\theta_c \notin T_0 \cap \Theta$. Next, let

$$C_i := \{c \in C : \mathsf{rank}_C(c) \leq i\}.$$

Observe that $C_0 \neq \emptyset$ (since C is finite and nonempty), and that if $c \in C_{i+1}$, then the codes of all immediate subformulae of the formula coded by c are in C_i. This observation ensures that the following recursive clauses yield a well-defined U_c for each $c \in C$.

- If $c \in C_0$ then $U_c := \begin{cases} \{\alpha : (c,\alpha) \in S_0\}, \text{ if } c \in F_0; \\ U_c := \emptyset, \text{ if } c \notin F_0. \end{cases}$.
- If $c \in C_{i+1}\backslash C_i$ and $c = \neg d$, then
$$U_c := \{\alpha \in A_c : \alpha \notin U_d\}.$$
- If $c \in C_{i+1}\backslash C_i$ and $c = a \vee b$, then
$$U_c := \{\alpha \in A_c : \alpha \restriction \mathsf{FV}(a) \in U_a \text{ or } \alpha \restriction \mathsf{FV}(b) \in U_b\}.$$

- If $c \in C_{i+1} \backslash C_i$ and $c = \exists v_a\ b$, then
$$U_c := \left\{\alpha \in A_c : \exists \alpha' \in N\ (\alpha \subseteq \alpha' \text{ and } \alpha' \in U_b)\right\}.$$

Note that in the first item above, the choice of $U_c := \emptyset$ when $c \in C_0$ and $c \notin F_0$ is completely arbitrary.[7] Also, in the third item above where $c = a \vee b$, both a and b will be in C_i, thanks to the properties of $\lhd^*$.

It is routine to verify, using induction on $\mathsf{rank}_C(c)$, that (1) and (2) hold for $(\mathcal{N}_0, U_c)_{c \in C}$. More specifically, if $\mathsf{rank}_C(c) = 0$, then $\theta_c \notin T_0 \cap \Theta$, so (1) is vacuously satisfied, and (2) is satisfied by design. On the other hand, when $\mathsf{rank}_C(c) > 0$ then (1) is satisfied since U_c is defined so as to comply with Tarski conditions; and (2) is satisfied since S_0 is an F_0-satisfaction class. This concludes the proof of the consistency of arbitrary finite subsets T_0 of $\mathsf{Th}^+(\mathcal{N}_0)$, which in turn shows that $\mathsf{Th}^+(\mathcal{N}_0)$ has a model, i.e., some elementary extension $\mathcal{N}_1$ of $\mathcal{N}_0$ has an expansion $\mathcal{N}_1^+$ of the form
$$\mathcal{N}_1^+ := (\mathcal{N}_1, U_c)_{c \in F_1}$$
with the property that $\mathcal{N}_1^+ \models \mathsf{Th}^+(\mathcal{N}_0)$. Let S_1 be the binary relation on N_1 defined via
$$S_1(c, \alpha) \Leftrightarrow \alpha \in U_c.$$
It is evident that S_1 is an F_1-satisfaction class, $S_1 \supseteq S_0$ and $(c, \alpha) \in S_0$ whenever $c \in F_0$, $\alpha \in N_0$, and $(c, \alpha) \in S_1$. □

3.2. Theorem. *Let $\mathcal{M}_0$ be a model of* PA *of any cardinality.*

(a) *If S_0 is an F_0-satisfaction class on $\mathcal{M}_0$, then there is an elementary extension $\mathcal{M}$ of $\mathcal{M}_0$ that carries a full satisfaction class that extends S_0.*

(b) *There is an elementary extension $\mathcal{M}$ of $\mathcal{M}_0$ that carries a full satisfaction class.*

Proof: Note that (b) is an immediate consequence of (a) since we may choose F_0 to be the set of *atomic* $\mathcal{M}_0$-formulae and S_0 to be the obvious satisfaction predicate for F_0. To establish (a), we note that by Lemma 3.1 there is an elementary extension $\mathcal{M}_1$ of $\mathcal{M}_0$ that carries an F_1-satisfaction class, where $F_1 := \mathsf{Form}^{\mathcal{M}_0}$. Lemma 3.1 allows this argument to be carried out ω-times to yield two sequences $\langle \mathcal{M}_i : i \in \omega \rangle$ and $\langle S_i : i \in \omega \rangle$ that satisfy the following properties for each $i \in \omega$:

(1) $\mathcal{M}_i \prec \mathcal{M}_{i+1}$;

(2) S_{i+1} is an F_{i+1}-satisfaction class on $\mathcal{M}_{i+1}$ with $F_{i+1} := \mathsf{Form}^{\mathcal{M}_i}$; and

(3) $S_i = S_{i+1} \cap \{(c, \alpha) : c \in F_i,\ \mathcal{M}_i \models \mathsf{Asn}(\alpha, c)\}$.

Let $\mathcal{M} := \bigcup_{i \in \omega} \mathcal{M}_i$, and $S := \bigcup_{i \in \omega} S_i$. Tarski's elementary chain theorem and (1) together imply that $\mathcal{M}$ elementarily extends $\mathcal{M}_0$. It is easy to see, using (2) and (3), that S is a full satisfaction class on $\mathcal{M}$. □

[7] As shown in [2] this feature can be exploited to construct 'pathological' satisfaction classes, such as the one mentioned at the end of Section 6 of this paper.

Theorem 3.2, when coupled with the completeness theorem of first order logic, immediately yields the following conservativity result.

3.3. Corollary. $\mathsf{PA}^{\mathsf{FS}}$ *is a conservative extension of* PA**.**

Proof: Suppose not. Then for some arithmetical sentence φ we have:

(1) $\mathsf{PA}^{\mathsf{FS}} \vdash \varphi$, and

(2) $\mathsf{PA} \nvdash \varphi$.

Since (2) implies that $\mathsf{PA} \cup \{\neg\varphi\}$ is consistent, by the completeness theorem for first order logic, there is a model $\mathcal{M}_0 \models \mathsf{PA} \cup \{\neg\varphi\}$. On the other hand, by part (b) of Theorem 3.2 there is an elementary extension $\mathcal{M}_1$ of $\mathcal{M}_0$ that carries a full satisfaction class, and therefore by (1) $\mathcal{M}_1 \models \varphi$. This contradicts the fact that $\mathcal{M}_1$ elementarily extends $\mathcal{M}_0$. □

3.4. Corollary. *Every resplendent model of* PA *carries a full satisfaction class. In particular, every countable recursively saturated model of* PA *carries a full satisfaction class.*

Proof: The first claim directly follows from the definition of a resplendent model. The second claim follows from the first claim, when coupled with the key result that countable recursively saturated models are resplendent (see [9, Section 15.2] for more detail). □

4. Truth Classes

With the exception of Krajewski's original paper [11], what we refer to as a 'truth class' here has been dubbed 'satisfaction class' in the model-theoretic literature. More specifically, Krajewski [11] employed the framework of satisfaction classes over base theories formulated in relational languages as in this paper, however, the later series of papers [10], [16], and [17] all used the framework of truth classes over Peano arithmetic formulated in a relational language, augmented with 'domain constants'. Later, Kaye [9] developed the theory of satisfaction classes over models of PA in languages incorporating function symbols; his work was extended by Engström [3] to truth classes over models of PA in functional languages.

As explained in this section, there is a simple canonical correspondence between truth classes over models of PA (in a relational language) and certain types of satisfaction classes, here referred to as 'extensional'. The main aim of this section is to demonstrate that the method of building satisfaction classes in the previous section can be conveniently modified so as to yield full *extensional* satisfaction classes (and thereby: full truth classes) over appropriate models of PA.

Within PA one can easily define an injective function c that yields the code for a constant symbol $\overline{x}$ for each member x of the domain. This enables PA to internally represent the language $\mathcal{L}^{+}_{\mathsf{PA}} = \mathcal{L}_{\mathsf{PA}}+$ 'domain constants'. We can then add a unary predicate $\mathsf{T}(x)$ denoting a *truth class* (instead of a binary

predicate $\mathsf{S}(x, y)$ for a satisfaction class) to $\mathcal{L}_{\mathsf{PA}}$, whose intended interpretation is "x is the code of a true *sentence* σ", where σ is an arithmetical sentence formulated in a language $\mathcal{L}^{+}_{\mathsf{PA}}$. We will make this more precise in the following definition.

4.1. Definition. $\mathsf{PA}^{\mathsf{FT}} := \mathsf{PA} \cup \mathsf{Tarski}(\mathsf{T})$, where $\mathsf{Tarski}(\mathsf{T})$ is the conjunction of the universal generalizations of $\mathsf{tarski}_0(\mathsf{T})$ through $\mathsf{tarski}_4(\mathsf{T})$, all formulated in the language $\mathcal{L}_{\mathsf{PA}} \cup \{\mathsf{T}(\cdot)\}$, as described below.[8] In what follows $\mathsf{Sent}(x)$ is the $\mathcal{L}_{\mathsf{PA}}$-formula that expresses "x is a formula of $\mathcal{L}^{+}_{\mathsf{PA}}$ with no free variables", and R ranges over relations symbols in $\mathcal{L}_{\mathsf{PA}}$.

- $\mathsf{tarski}_0(\mathsf{T}) := (\mathsf{T}(x) \to \mathsf{Sent}(x))$.
- $\mathsf{tarski}_{1,R}(\mathsf{T}) := \left(\ulcorner R(\bar{t}_0, \cdots, \bar{t}_{n-1})\urcorner = x\right) \to (R(t_0, \cdots, t_{n-1}) \leftrightarrow \mathsf{T}(x))$.
- $\mathsf{tarski}_2(\mathsf{T}) := (x = \neg y) \to (\mathsf{T}(x) \leftrightarrow \neg\mathsf{T}(y_1))$.
- $\mathsf{tarski}_3(\mathsf{T}) := (x = y_1 \vee y_2) \to (\mathsf{T}(x) \leftrightarrow (\mathsf{T}(y_1) \vee \mathsf{T}(y_2)))$.
- $\mathsf{tarski}_4(\mathsf{T}) := (x = \exists v_i \varphi) \to (\mathsf{T}(x) \leftrightarrow \exists z\ \mathsf{T}(\varphi(\overline{z})))$.

T is a *full truth class* on $\mathcal{M}$ if $(\mathcal{M}, T) \models \mathsf{PA}^{\mathsf{FT}}$.

4.2. Definition. A *substitution* for a formula ψ of $\mathcal{L}_{\mathsf{PA}}$ is a function

$$\sigma : \mathsf{FV}(\psi) \to \mathsf{Var}$$

such that σ respects substitutability in the 'usual way', i.e., if x is a free variable of ψ, then x is not in the scope of any quantifier that binds $\sigma(x)$. Given ψ and σ as above, let $\psi * \sigma$ be the formula obtained from ψ by applying the substitution σ, and A be the set of pairs (φ, α) such that α is an assignment for the formula φ. This allows us to define a key equivalence relation $\sim$ on A by decreeing that $(\varphi_0, \alpha_0) \sim (\varphi_1, \alpha_1)$ iff there is some $(\psi, \beta) \in A$, and there are substitutions σ_0 and σ_1 for ψ, with

$$\varphi_i = \psi * \sigma_i \text{ and } \beta = \alpha_i \circ \sigma_i, \text{ for } i = 0, 1.$$

In the above, $\alpha_i \circ \sigma_i$ is the composition of α_i and σ_i. It is important to bear in mind that, intuitively speaking, $(\varphi_0, \alpha_0) \sim (\varphi_1, \alpha_1)$ means that φ_0 and φ_1 are the same except for their free variables, and for all variables x and y, if x occurs freely in the same position in φ_0 as y does in φ_1, then $\alpha_0(x) = \alpha_1(y)$.

- An F-satisfaction class S is *extensional* if for all φ_0 and φ_1 in F, $\mathcal{M} \models (\varphi_0, \alpha_0) \sim (\varphi_1, \alpha_1)$ implies $(\varphi_0, \alpha_0) \in S$ iff $(\varphi_1, \alpha_1) \in S$.[9]

The following proposition describes a canonical correspondence between extensional satisfaction classes and truth classes. The routine but laborious proof is left to the reader.

[8] $\mathsf{PA}^{\mathsf{FT}}$ is the relational analogue of the theory of $\mathsf{CT}\restriction$ in Halbach's monograph [8]. The base theory of $\mathsf{CT}\restriction$ is PA formulated in a *functional* language. The conservativity of $\mathsf{CT}\restriction$ over the functional language version of PA can also be established using the techniques of this paper (see Section 6).

[9] Note that an extensional satisfaction predicate need not be closed under re-naming of bound variables.

- In what follows c is the $\mathcal{M}$-definable injection $m \mapsto_{\mathsf{c}} \overline{m}$ that designates a constant symbol $\overline{m}$ for each $m \in M$, and $\varphi(\mathsf{c} \circ \alpha)$ is the sentence in the language $\mathcal{L}^+_{\mathsf{PA}}$ obtained by replacing each occurrence of a free variable x of φ with the constant symbol $\overline{m}$, where $\alpha(x) = m$.

4.3. Proposition. *Suppose $\mathcal{M} \models \mathsf{PA}$, T is a full truth class on $\mathcal{M}$, and S is an extensional full satisfaction class on $\mathcal{M}$.*

(a) *$\mathcal{S}(T)$ is an extensional satisfaction class on $\mathcal{M}$, where $\mathcal{S}(T)$ is defined as the collection of ordered pairs (φ, α) such that $\varphi(\mathsf{c} \circ \alpha) \in T$.*

(b) *$\mathcal{T}(S)$ is a truth class on $\mathcal{M}$, where $\mathcal{T}(S)$ is defined as the collection of $\varphi \in \mathcal{L}^+_{\mathsf{PA}}$ such that for some $\psi \in \mathcal{L}^+_{\mathsf{B}}$ and some assignment α for ψ, $\varphi = \psi(\mathsf{c} \circ \alpha)$ and $(\psi, \alpha) \in S$.*

(c) *$\mathcal{S}(\mathcal{T}(S)) = S$, and $\mathcal{T}(\mathcal{S}(T)) = T$.*

Before describing the construction of extensional satisfaction classes we need the preliminaries presented in Definition 4.4 and Lemma 4.5.

4.4. Definition.

(a) Given formulae φ_0 and φ_1 of $\mathcal{L}_{\mathsf{PA}}$, we write $\varphi_0 \approx \varphi_1$ if there is a formula ψ, and substitutions σ_0 and σ_1 for ψ such that $\varphi_i \approx \psi * \sigma_i$ for $i = 0, 1$.

(b) Given $c \in \mathsf{Form}^{\mathcal{M}}$, let $\mathbf{TC}_{\mathcal{M}}(c)$ be the *externally defined* transitive closure of c with respect to the direct subformula relation, i.e.,

$$\mathbf{TC}_{\mathcal{M}}(c) := \bigcup_{n<\omega} \mathbf{TC}_{\mathcal{M}}(c, n),$$

where $\mathbf{TC}_{\mathcal{M}}(c, 0) := \{c\}$ and

$$\mathbf{TC}_{\mathcal{M}}(c, n+1) := \{x \in M : x \lhd^{\mathcal{M}} d \text{ for some } d \in \mathbf{TC}_{\mathcal{M}}(c, n)\}.$$

The following lemma presents salient features of the two equivalence relations $\sim$ and $\approx$.

4.5. Lemma. *Let $\sim$ be as in Definition 4.2; and $\mathbf{TC}_{\mathcal{M}}(c)$ and $\approx$ be as in Definition 4.4.*

(i) If $d \in \mathbf{TC}_{\mathcal{M}}(c)$ and $d \neq c$, then $\neg\,(c \approx d)$.

(ii) $\approx$ preserves the principal connectives, i.e., it relates negations to negations, disjunctions to disjunctions, and existential formulae to existential formulae with the same bound variable. Moreover, if $\neg c \approx \neg d$, then $c \approx d$; if $c \vee d \approx c' \vee d'$, then $c \approx c'$ and $d \approx d'$; and if $\exists t\ c \approx \exists t'\ c'$, then $t = t'$ and $c \approx c'$.

(iii) If $(\varphi_0, \alpha_0) \sim (\varphi_1, \alpha_1)$, then $\varphi_0 \approx \varphi_1$.

(iv) If $(\neg\varphi_0, \alpha_0) \sim (\neg\varphi_1, \alpha_0)$, then $(\varphi_0, \alpha_0) \sim (\varphi_1, \alpha_1)$.

(v) If $(\varphi_0 \vee \varphi_1, \alpha) \sim (\varphi_0' \vee \varphi_1', \alpha')$, then $(\varphi_0, \alpha \restriction \mathsf{FV}(\varphi_0)) \sim (\varphi_0', \alpha' \restriction \mathsf{FV}(\varphi_0'))$ and $(\varphi_1, \alpha \restriction \mathsf{FV}(\varphi_1)) \sim (\varphi_1', \alpha' \restriction \mathsf{FV}(\varphi_1'))$.

(vi) If $\varphi = \exists t\ \psi$, *and* $\varphi' = \exists t'\ \psi'$, *and* $(\varphi, \alpha) \sim (\varphi', \alpha')$, *then* $t = t'$ *and for some* e

$$(\varphi, \alpha[t : e]) \sim (\varphi', \alpha'[t' : e]).^{10}$$

The next Lemma presents a variant of Lemma 3.1 that is our main tool for constructing extensional satisfaction classes.

4.6. Lemma. *Let* $\mathcal{N}_0 \models \mathsf{PA}$, $F_1 := \mathsf{Form}^{\mathcal{N}_0}$, $F_0 \subseteq F_1$, *and suppose* S_0 *is an extensional* F_0*-satisfaction class. Then there is an elementary extension* $\mathcal{N}_1$ *of* $\mathcal{N}_0$ *that carries an extensional* F_1*-satisfaction class* $S_1 \supseteq S_0$ *and* $(c, \alpha) \in S_0$ *whenever* $c \in F_0$, $\alpha \in N_0$, *and* $(c, \alpha) \in S_1$.

Proof. Let Θ and Γ be as in the proof of Lemma 3.1, and let

$$\mathsf{Th}^+(\mathcal{N}_0) := \mathsf{Th}(\mathcal{N}_0, a)_{a \in N_0} \cup \Theta \cup \Gamma \cup \Delta,$$

where $\Delta := \{\delta_{cc'} : c,\ c' \in F_1\}$, and

$$\delta_{cc'} := \forall \alpha\ \forall \alpha' \left((c, \alpha) \sim (c', \alpha') \to (\mathsf{U}_c(\alpha) \leftrightarrow \mathsf{U}_{c'}(\alpha'))\right).$$

The proof of the lemma would be complete once we verify that $\mathsf{Th}^+(\mathcal{N}_0)$ has a model. To this end, we shall demonstrate that every finite subset T_0 of $\mathsf{Th}^+(\mathcal{N}_0)$ is interpretable in $\mathcal{N}$. Let C be the collection of $c \in F_1$ such that c appears in T_0. Also, let $\lhd^*$ and $\mathsf{rank}_C(c)$ be precisely as in the proof of Lemma 3.1.

We can extend C to another finite set $\overline{C}$ so that it satisfies a certain closure property, namely: whenever we have $c \approx c'$ and $d \lhd^* c$, where c, c' and d are all in $\overline{C}$, then there is some $d' \in \overline{C}$ such that $d' \lhd^* c'$ with $d \approx d'$. This can be done simply by adding any missing direct subformulae d' by an appropriate recursion.[11] By replacing C by $\overline{C}$ we may therefore additionally assume:

(#) If c and c' are both in C with $c \approx c'$, then $\mathsf{rank}_C(c) = \mathsf{rank}_C(c')$.

As in the proof of Lemma 3.1, we then recursively construct $\{U_c : c \in C\}$ such that:

(1) $(\mathcal{N}_0, U_c)_{c \in C} \models \{\theta_c : c \in C\}$ and

(2) For $c \in C \cap F_0$, $U_c = \{\alpha \in N_0 : (c, \alpha) \in S_0\}$.

---

[10] Here $\alpha[t : e]$ is the assignment obtained by redefining the value of α at the variable t to be e if $t \in \mathrm{Dom}(\alpha)$; note that $\alpha[t : e] = \alpha$ if $t \notin \mathrm{Dom}(\alpha)$.

[11] More specifically, first define $\lhd^\circ$ on C by $d' \lhd^\circ c$ iff $d' \lhd^* c' \approx c$, for some $c' \in C$. Since $\lhd^\circ$ is cycle-free, C is well-founded, and therefore lends itself to a ranking function $\mathsf{rank}^\circ_C(c)$. Let $n = \max\{\mathsf{rank}^\circ_C(c) : c \in C\}$, and for $0 \leq i \leq n$ define $D_i := \{c \in C : \mathsf{rank}^\circ_C(c) = i\}$. Next use a 'backward' recursion to define $E_n, E_{n-1}, \cdots, E_0$ via:

- $E_n := D_n$;
- $E_{n-(i+1)} := D_{n-(i+1)} \cup \{d : d \lhd^{\mathcal{N}_0} c \text{ for some } c \in E_{n-i}\}$.

Finally, let $\overline{C} := E_n \cup \cdots \cup E_0$. It is easy to see that $\overline{C}$ is finite, extends C, and has the desired closure property.

It remains to show:

(3) $(\mathcal{N}_0, U_c)_{c\in C} \models \{\delta_{cc'} : c,\ c' \in C\}$.

We establish (3) by using induction on $\mathsf{rank}_C(c)$ to show that $\forall c \in C\ P(c)$, where $P(c)$ abbreviates:

$$\forall c' \in C\ \ (\mathcal{N}_0, U_c)_{c\in C} \models \forall\alpha\ \forall\alpha'\ \ ((c,\alpha) \sim (c',\alpha') \to (\mathsf{U}_c(\alpha) \leftrightarrow \mathsf{U}_{c'}(\alpha'))).$$

If $\mathsf{rank}_C(c) = 0$ and $(c,\alpha) \sim (c',\alpha')$, then by part (iii) of Lemma 4.5 we have $c \approx c'$, which in turn by (#) assures us that $\mathsf{rank}_C(c') = 0$. This makes it clear that $P(c)$ holds when $\mathsf{rank}_C(c) = 0$ since S_0 is assumed to be an *extensional* F_0-satisfaction class.

To verify the inductive step, suppose:

(4) $P(x)$ holds for all $x \in C$ with $\mathsf{rank}_C(x) = i$.

Let $c \in C$ with $\mathsf{rank}_C(c) = i+1$, and suppose $(c,\alpha) \sim (c',\alpha')$, where $c = \exists t\ d$. Then $c' = \exists t'\ d'$, and $d \approx d'$ by part (ii) of Lemma 4.5. Observe that thanks to (#) we have:

(5) $\mathsf{rank}_C(c') = i+1$ and $\mathsf{rank}_C(d) = \mathsf{rank}_C(d') = i$.

Now if $\alpha \in U_c$, then $\alpha[t:e] \in U_d$ for some e by (1), and therefore by part (vi) of Lemma 4.5, we obtain:

(6) $(d, \alpha[t:e]) \sim (d', \alpha'[t':e])$.

Using (4), (5), and (6) we may now conclude that $\alpha'[t:e] \in U_{d'}$, which by (1) yields $\alpha' \in U_d$, thus completing the verification of the quantificational case (by symmetry). A similar reasoning can be carried out for propositional cases. This concludes the proof of consistency of T_0.

The rest is precisely as before: the consistency of $\mathsf{Th}^+(\mathcal{N}_0)$ implies that there is an elementary extension $\mathcal{N}_1$ of $\mathcal{N}_0$ that has an expansion $\mathcal{N}_1^+ := (\mathcal{N}_1, S) \models \mathsf{Th}^+(\mathcal{N}_0)$, and the binary relation S_1 on $\mathcal{N}_1$ defined via

$$S_1(c,\alpha) \Leftrightarrow \alpha \in U_c$$

has the property that S_1 is an extensional F_1-satisfaction, $S_1 \supseteq S_0$, and $(c,\alpha) \in S_0$ whenever $c \in F_0$, $\alpha \in N_0$, and $(c,\alpha) \in S_1$. □

4.7. Theorem. *Let $\mathcal{M}_0 \models \mathsf{PA}$. There is an elementary extension $\mathcal{M}$ of $\mathcal{M}_0$ that carries a full extensional satisfaction class.*

Proof: Since the satisfaction class S_0 on the collection F_0 of atomic formulae of $\mathcal{M}_0$ is extensional, we may use Lemma 4.6 instead of Lemma 3.1 in order to carry out the elementary chain argument of Theorem 3.2. □

By coupling Theorem 4.7 with part (b) of Proposition 4.3 we obtain:

4.8. Corollary. *Every model of* PA *has an elementary extension that carries a full truth class.*

Finally, the line of reasoning employed in the proof of Corollary 3.3 shows, using Corollary 4.8, that:

4.9. Corollary. $\mathsf{PA}^{\mathsf{FT}}$ *is a conservative extension of* PA.

5. Arithmetization, Interpretability, and Conservativity

Here we briefly discuss the *arithmetization* of the constructions of the previous two sections, with an eye towards issues connected with interpretability and conservativity. As explained in [2, Section 4] the compactness and elementary chain argument employed in the proofs of Theorems 3.2 and 4.7 can be implemented in the fragment $\mathsf{I}\Sigma_2$ of PA with the help of the 'Low Basis Theorem' of Recursion Theory. Coupled with Orey's Compactness Theorem, this can used to establish the following:

5.1. Theorem. [2] $\mathsf{PA}^{\mathsf{FT}}$ *is interpretable in* PA.[12]

On the other hand, the technology of LL_1-sets[13] of [6, Theorem 4.2.7.1, p. 104] can be used to show that the proofs of both theorems 3.2 and 4.7 can even be implemented in the fragment $\mathsf{I}\Sigma_1$ of PA. In light of the fact that the statement "$\mathsf{PA}^{\mathsf{FT}}$ is conservative over PA" is a Π_2-statement, and $\mathsf{I}\Sigma_1$ is well known[14] to be Π_2-conservative over PRA, we obtain the following:

5.2. Theorem. [2] *The conservativity of* $\mathsf{PA}^{\mathsf{FT}}$ *over* PA *can be verified in* PRA.[15]

5.3. Remark. The verification of the conservativity of $\mathsf{PA}^{\mathsf{FT}}$ over PA within PRA was first claimed by Halbach in [7], using cut-elimination.[16] Later, Fischer [4] gave a proof, based on the cut-elimination argument in [7], to show that $\mathsf{PA}^{\mathsf{FT}}$ is interpretable in PA. Unfortunately, a gap was discovered recently (by Fujimoto) in the cut-elimination argument in [7], which in turn impaired Fischer's interpretability claim. Happily, Leigh [12] has succeeded in developing a proof-theoretic demonstration of the conservativity of $\mathsf{PA}^{\mathsf{FT}}$ over PA that is implementable in PRA. Moreover, [12, Theorem 1] can be

[12]Indeed B^{FS} turns out to be interpretable in B for all base theories B that have access to the full scheme of induction over their ambient 'numbers'. In particular, $\mathsf{ACA}^{\mathsf{FS}}$ is interpretable in ACA. On the other hand, as shown in [2, Section 8], $\mathsf{ACA}_0^{\mathsf{FS}}$ is *not* interpretable in ACA_0 (more generally, B^{FS} is shown to be *not* interpretable in B, if B is finitely axiomatizable).

[13]LL_1-sets are a special type of 'low' sets.

[14]This classical result was independently established by Mints, Parsons, and Takeuti, using proof-theoretic methods. The work of Paris and Kirby (described in [14, IX.3]), and more recently Avigad [1] has also provided model-theoretic demonstrations of this conservativity result.

[15]Indeed, by using the technique of Friedman [5], this conservativity result is already verifiable in the fragment SEFA (Superexponential Arithmetic) of PRA.

[16]Halbach's base theory in his work is the usual version of PA that is formulated in a functional language.

used to verify the interpretability of $\mathsf{PA}^{\mathsf{FT}}$ over PA, by using Fischer's strategy in [4].[17] Therefore, Theorems 5.1 and 5.2 can be arrived at via two completely different routes.

6. Further Results

In Section 4 we saw that the core methodology of Section 3 can be fine-tuned to build full extensional satisfaction classes. Indeed, as shown in [2] one can strengthen Theorem 4.7 by imposing further desirable conditions on the satisfaction class S. For example, every model $\mathcal{M}_0$ of PA has an elementary extension $\mathcal{M}$ that carries a full extensional satisfaction class S that satisfies all of the following additional properties:

(1) $\mathsf{Sat}_n^{\mathcal{M}} \subseteq S$ for all $n \in \omega$ (see Theorem 2.5 for Sat_n).

(2) If $c \in \mathsf{Form}^{\mathcal{M}}$ and $\mathcal{M} \models$ "c is an axiom of PA", then S deems c to be 'true'.[18]

(3) If c and c' are $\mathcal{M}$-formulae such that $\mathcal{M} \models$ "c' is an alphabetic variant[19] of c", then $(c, \alpha) \in S$ iff $(c', \alpha) \in S$.

Furthermore, the third condition above can be strengthened by accomodating a combination of extensional equivalence and alphabetic equivalence, thereby yielding truth classes that are closed under alphabetic equivalence. We have also shown that a small dose of condition (1) can be used to build full truth classes over models of arithmetical theories formulated in *functional* languages (this result will appear in the projected sequel to [2]).

One can also use the method of Section 3 to build bizarre satisfaction classes. For example, as shown in [2, Section 5], every model $\mathcal{M}_0$ of PA has an elementary extension $\mathcal{M}$ that has a full satisfaction class S that exhibits the following pathology:

$$\{a \in M : (\sigma_a, \alpha_{\mathsf{Null}}) \in S\} = \omega_{\mathcal{M}},$$

where $\omega_{\mathcal{M}}$ is the well-founded initial segment of $\mathcal{M}$ that is isomorphic to ω, and σ_a is defined for all $a \in M$ by a recursion within $\mathcal{M}$ via the following clauses:

- $\sigma_0 := \exists v_0 \, (v_0 = v_0)$ (or $\sigma_0 =$ any other logically valid sentence);
- $\sigma_{n+1} := (\sigma_n \vee \sigma_n)$.

[17] We are grateful to Graham Leigh for his kind permission to quote his unpublished work here.

[18] As remarked in the last sentence of [10], this condition can also be arranged using the machinery of $\mathcal{M}$-logic. Note that 'axioms of PA' in the sense used here do not include the logical axioms.

[19] c' is an alphabetic variant of c if c' is obtainable from c by the usual rules of re-naming the bound variables of c.

References

[1] J. Avigad, *Saturated models of universal theories*, **Annals of Pure and Applied Logic** vol. 118 (2002), pp. 219-234.

[2] A. Enayat and A. Visser, *Full satisfaction classes in a general setting* (Part I), to appear in **Logic Group Preprint Series**, available at `http://www.phil.uu.nl/preprints/lgps/Utrecht Preprint Series`.

[3] F. Engström, *Satisfaction classes in nonstandard models of first-order arithmetic*, **arXiv.org.math**, available at: `http://arxiv4.library.cornell.edu/abs/math/0209408v1`

[4] M. Fischer, *Minimal truth and interpretability*, **Review of Symbolic Logic,** vol. 2 (2009), pp. 799–815.

[5] H. Friedman, *Finitist proofs of conservation*, **FOM Archives**, available at `http://cs.nyu.edu/pipermail/fom/1999-September/003405.html.`

[6] P. Hájek and P. Pudlák, **Metamathematics of First-Order Arithmetic**, Springer, 1993.

[7] V. Halbach, *Conservative theories of classical truth*, **Studia Logica**, vol. 62 (1999), pp. 353-370.

[8] ----------, **Axiomatic Theories of Truth,** Cambridge University Press, Cambridge, 2011.

[9] R. Kaye, **Models of Peano Arithmetic**, Oxford Logic Guides, Oxford University Press, Oxford, 1991.

[10] H. Kotlarski, S. Krajewski, and A.H. Lachlan, *Construction of satisfaction classes for nonstandard models*, **Canadian Mathematical Bulletin**. vol. 24 (1981), pp. 283–293.

[11] S. Krajewski, *Nonstandard satisfaction classes*, in **Set Theory and Hierarchy Theory: A Memorial Tribute to Andrzej Mostowski** (ed. W. Marek et al.) Lecture Notes in Mathematics, vol. 537, Springer-Verlag, Berlin, 1976, pp. 121-144.

[12] G. E. Leigh, *Deflating truth,* manuscript (November 2012).

[13] V. McGee, *In praise of the free lunch: why disquotationalists should embrace compositional semantics*, in **Self-reference**, CSLI Lecture Notes, 178, CSLI Publ., Stanford, pp. 95-120.

[14] S. Simpson, **Subsystems of second order arithmetic**, Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1999.

[15] S. T. Smith, **Non-standard Syntax and Semantics and Full Satisfaction Classes**, Ph.D. thesis, Yale University, New Haven, Connecticut, 1984.

[16] ----------, *Nonstandard characterizations of recursive saturation and resplendency,* **Journal of Symbolic Logic,** vol. 52 (1987), pp. 842-863.

[17] ----------, *Nonstandard definability,* **Annals of Pure and Applied Logic,** vol. 42 (1989), pp. 21–43.

Ali Enayat
Department of Mathematics and Statistics, American University, 4400 Mass. Ave. NW, Washington, DC 20016-8050, USA.
`enayat@american.edu`

Albert Visser
Department of Philosophy, Bestuursgebouw, Heidelberglaan 6, 584 CS Utrecht, The Netherlands.
`albert.visser@phil.uu.nl`