The Review of Symbolic Logic
Volume 18, Number 2, June 2025

SATISFACTION CLASSES WITH APPROXIMATE DISJUNCTIVE CORRECTNESS

ALI ENAYAT
Department of Philosophy, Linguistics, and Theory of Science,
University of Gothenburg

Abstract. The seminal Krajewski–Kotlarski–Lachlan theorem (1981) states that every countable recursively saturated model of PA (Peano arithmetic) carries a full satisfaction class. This result implies that the compositional theory of truth over PA commonly known as $\mathsf{CT}^-[\mathsf{PA}]$ is conservative over PA. In contrast, Pakhomov and Enayat (2019) showed that the addition of the so-called axiom of disjunctive correctness (that asserts that a finite disjunction is true iff one of its disjuncts is true) to $\mathsf{CT}^-[\mathsf{PA}]$ axiomatizes the theory of truth $\mathsf{CT}_0[\mathsf{PA}]$ that was shown by Wcisło and Łełyk (2017) to be nonconservative over PA. The main result of this paper (Theorem 3.12) provides a foil to the Pakhomov–Enayat theorem by constructing full satisfaction classes over arbitrary countable recursively saturated models of PA that satisfy arbitrarily large approximations of disjunctive correctness. This shows that in the Pakhomov–Enayat theorem the assumption of disjunctive correctness cannot be replaced with any of its approximations.

§1. Introduction. Intuitively speaking, a binary relation S on a model $\mathcal{M}$ of PA (Peano arithmetic) is said to be a *satisfaction class* if S 'behaves like' the usual Tarskian satisfaction relation $\mathrm{Sat}_{\mathcal{M}}$ on $\mathcal{M}$, but in sharp contrast to the usual Tarskian satisfaction relation on $\mathcal{M}$, if $\mathcal{M}$ is nonstandard, then S is required to decide the 'truth' of at least some *nonstandard* 'formulae' of $\mathcal{M}$. This notion was brought to prominence in the 1970s and 1980s, thanks to the fine efforts of a number of logicians, including (in alphabetical order) Barwise, Kotlarski, Krajewski, Lachlan, Murawski, Ratajczyk, Kossak, Schlipf, Schmerl, Smith, and Wilmers. A notable precursor is Robinson whose landmark 1963 paper [28] probed the subtle obstacles in the development of a coherent Tarski-style semantic framework for nonstandard formulae.

The flurry of activity in the 1970s and 1980s revealed two distinct 'flavors' of satisfaction classes: *full* satisfaction classes and *inductive* satisfaction classes. A full satisfaction class on a model $\mathcal{M}$ of arithmetic is required to decide the 'truth' of *every* arithmetic 'formula' in $\mathcal{M}$ (including the nonstandard ones, if $\mathcal{M}$ is nonstandard) while obeying the usual Tarskian recursive clauses that relate the truth of a formula to the truth of its components. In contrast, an inductive satisfaction class S on a model $\mathcal{M}$ is required to satisfy the following two properties: (1) S obeys Tarski's recursive clauses for (at least) all standard arithmetical formulae; and (2) the expansion $(\mathcal{M}, S)$

Received: November 22, 2023.
2020 *Mathematics Subject Classification*: Primary 03C62, 03F30, Secondary 03F25.
Key words and phrases: Peano arithmetic, axiomatic theories of truth, conservativity, disjunctive correctness.

© The Author(s), 2024. Published by Cambridge University Press on behalf of The Association for Symbolic Logic.

doi:10.1017/S1755020324000182

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

of $\mathcal{M}$ satisfies the scheme of induction in the language obtained by augmenting the arithmetical language with a new predicate symbol representing S. Kaye's textbook [18] covers the basics of both types of satisfaction classes. It is known that if $\mathcal{M}$ is a model of PA that carries a satisfaction class that is either full or inductive, then $\mathcal{M}$ is recursively saturated. For countable models, the converse also holds, i.e., every countable recursively saturated model $\mathcal{M}$ of PA carries a full satisfaction class S_1, as well as an inductive satisfaction class S_2; but the existence of a satisfaction class on $\mathcal{M}$ that is both full and inductive implies that the formal consistency of PA holds in $\mathcal{M}$ (and much more) and thus by Gödel's second incompleteness theorem not every countable recursively saturated model of PA carries a satisfaction class that is both full and inductive. The monograph [20] by Kossak and Schmerl includes a more advanced treatment of inductive satisfaction classes, but it does not contain material on full satisfaction classes. This can be explained by the fact that inductive satisfaction classes have proved to be indispensable in the model theory of arithmetic, but full satisfaction classes have not played a comparable role. To the best of the author's knowledge, the only known prominent application of full satisfaction classes over nonstandard models to the model theory of PA is due to Smith [29], who employed them to calibrate the logical complexity of the notions of recursive saturation and resplendence in the context of models of PA (as respectively Σ^1_1 and Δ^1_2).

However, full satisfaction classes have captured the imagination of philosophical logicians since they are intimately linked with the grand topic of axiomatic theories of truth, and in particular they shed light on the vibrant debate concerning the deflationist conception of truth, especially in connection with the so-called conservativeness argument. This philosophical interest has galvanized the subject of satisfaction classes and has led to a new wave of technical advances and questions over the past decade. The monographs by Halbach [16] and Cieśliński [4] provide an overview of the philosophical motivations as well as minutiae of the technical preliminaries.

There is also a notable *methodological* asymmetry between inductive satisfaction classes and full satisfaction classes: given an arbitrary model $\mathcal{M}$ of PA it is routine to construct an elementary extension of $\mathcal{M}$ that carries an inductive satisfaction class with the help of the standard tools of the trade (the ingredients are definable partial satisfaction classes, and compactness). However, an air of mystery has surrounded full satisfaction classes ever since the original tour de force Kotlarski–Krajewski–Lachlan construction [23] that was based on the ad hoc technology of $\mathcal{M}$-logic (an infinitary logical system based on a nonstandard model $\mathcal{M}$). The Kotlarski–Krajewski–Lachlan construction implies that the axiomatic theory of truth commonly known as $\mathsf{CT}^-[\mathsf{PA}] = \mathsf{CT}^- + \mathsf{PA}$ is *conservative* over PA, i.e., if an arithmetical sentence φ is provable in $\mathsf{CT}^-[\mathsf{PA}]$, then φ is already provable in PA. Here CT^- is a finitely axiomatized theory formulated in the language of arithmetic augmented with a truth predicate T (the minus superscript indicates that no instances of the induction scheme mentioning T have been added to the theory). Decades later, a new versatile model-theoretic method of constructing full satisfaction classes was presented by Visser and the author [12]; this new method has been refined in various directions, e.g., as in Cieśliński's monograph [4], and in the joint work of Łełyk and Wcisło with the author [9]. The conservativity of $\mathsf{CT}^-[\mathsf{PA}]$ over PA has also been established by proof theoretic methods by Leigh [25], and more recently by Cieśliński [5].

The conservativity of $\mathsf{CT}^-[\mathsf{PA}]$ over PA together with Gödel's second incompleteness theorem implies that the sentence $\mathrm{Con}(\mathsf{PA})$ expressing the consistency of PA is not

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

provable in $\mathsf{CT}^-[\mathsf{PA}]$. However, it is well-known that the formal consistency of PA (and much more) is readily provable in the stronger theory $\mathsf{CT}[\mathsf{PA}]$, which is the result of strengthening $\mathsf{CT}^-[\mathsf{PA}]$ with the scheme of induction over natural numbers for all formulae in extended language (i.e., the language obtained by extending the language of PA with a truth predicate). Indeed, it is straightforward to demonstrate the consistency of PA within the subsystem $\mathsf{CT}_1[\mathsf{PA}]$ of $\mathsf{CT}[\mathsf{PA}]$, where $\mathsf{CT}_n[\mathsf{PA}]$ is the subtheory of $\mathsf{CT}[\mathsf{PA}]$ with the scheme of induction over natural numbers limited to formulae in the extended language that are at most of complexity Σ_n [32, theorem 2.8]. However, the case of $\mathsf{CT}_0[\mathsf{PA}]$ has taken considerable effort to analyze. Kotlarski [22] established that $\mathsf{CT}_0[\mathsf{PA}]$ is a subtheory of $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{Ref}(\mathsf{PA})$, where $+\mathsf{Ref}(\mathsf{PA})$ is the sentence in the extended language stating that "every first-order consequence of PA is true". Łełyk [26] demonstrated that the converse also holds, which immediately implies that $\mathsf{CT}_0[\mathsf{PA}]$ is not conservative over PA since the formal consistency of PA is readily provable in $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{Ref}(\mathsf{PA})$.[1] Kotlarski's aforementioned theorem was refined by Cieśliński [3] who proved that $\mathsf{CT}^-[\mathsf{PA}]+$"T is closed under propositional proofs" and $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{Ref}(\mathsf{PA})$ axiomatize the same theory. Later Cieśliński's result was refined by Pakhomov and the author [10] by demonstrating that $\mathsf{CT}^-[\mathsf{I}\Delta_0 + \mathsf{Exp}] + \mathsf{DC}$ and $\mathsf{CT}_0[\mathsf{PA}]$ axiomatize the same first-order theory, where DC (disjunctive correctness) is the axiom that states that a disjunction is true iff one of its disjuncts is true. This result in particular shows that $\mathsf{CT}^-[\mathsf{I}\Delta_0 + \mathsf{Exp}] + \mathsf{DC}$ and $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{DC}$ axiomatize the same theory.

The recent work of Cieśliński, Łełyk, and Wcisło [7] refined the aforementioned work of Pakhomov and the author by showing that $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{DC}_{\mathrm{out}}$ is an axiomatization of $\mathsf{CT}_0[\mathsf{PA}]$, where $\mathsf{DC}_{\mathrm{out}}$ is the 'half' of DC that says every true disjunction has a true disjunct. In summary, the arithmetical strength of $\mathsf{CT}^-[\mathsf{PA}]$ augmented with seemingly innocuous axioms such as "truth is closed propositional proofs" or even "If a disjunction is true, then it has a true disjunct" goes beyond PA.[2] The philosophical ramifications of the nonconservativity of $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{DC}$ has been explored by Fujimoto [14], whose work shows that the nonconservativity of $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{DC}$ over PA introduces a new twist to the conservativity argument in relation to the deflationist conception of truth.

In this paper we present two new constructions of extensional satisfaction classes over models of PA, such satisfaction classes are inter-definable with their less famous siblings known as 'truth classes' (see Proposition 2.14). The specific features of the truth classes constructed in this paper can be summarized as follows.

The first construction (at work in the proofs of Theorem 3.3 and 3.4) employs basic results in the model theory of PA to show that given a countable recursively saturated model $\mathcal{M}$ of PA, there are arbitrarily large cuts I in $\mathcal{M}$ with the property that there is a truth class T on $\mathcal{M}$ that is compositional for the collection of sentences that have at most i occurrences of closed terms for some i in I, and it satisfies disjunctive correctness

[1] This result was first claimed by Kotlarski [22], but his proof outline of $\mathsf{Ref}(\mathsf{PA})$ within $\mathsf{CT}_0[\mathsf{PA}]$ was found to contain a serious gap in 2011 by Heck and Visser; this gap cast doubt over the veracity of Kotlarski's claim until the issue was resolved by Łełyk in his doctoral dissertation [26]. Łełyk's work was preceded by the discovery of an elegant proof of the nonconservativity of $\mathsf{CT}_0[\mathsf{PA}]$ over PA by Wcisło and Łełyk [32].

[2] The proof in [7] uses the version of CT^- that includes the so-called regularity axiom; it is not clear if the main result of [7] holds without the inclusion of the regularity axiom. In contrast, the formulation in [10] of CT^- does not include the regularity axiom.

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

for such sentences. Note that the truth classes constructed in Theorem 3.3 and 3.4 are not full, and as indicated in Question 3.7 it is not clear whether they can be extended to full satisfaction classes.

The second construction (at work in the proof of Theorem 4.5) is far more involved than the first one since it uses a vast array of technical results to elaborate the construction of full truth classes in the joint paper [12] of Visser and the author to show that given a countable recursively saturated model $\mathcal{M}$ of PA, there are arbitrarily large cuts I in $\mathcal{M}$ such that there is a *full* truth class T on $\mathcal{M}$ that is disjunctively correct for disjunctions whose number of disjuncts is in I. This shows that in the aforementioned result of Pakhomov and the author the assumption of disjunctive correctness cannot be replaced with any of its approximations. As indicated in Remark 4.7, Theorem 4.5 implies the conservativity of an axiomatic theory of truth over PA that uses a parameterized family of truth predicate T_x (where x varies over the domain of discourse) such that T_x satisfies CT^- and T_x is disjunctively correct for disjunctions whose number of disjuncts is at most x.

We should draw attention to the recent joint work [1] of Athar Abdul-Quader and Mateusz Łełyk who also explore approximations of disjunctive correctness, but their results are complementary to those obtained in this paper since their focus is on a different set of problems (see also Remark 4.6).

The main idea of the first construction was discovered by the author in 2009; the same idea was earlier hit upon in 1980 by Smoryński in a letter to Jim Schmerl; for more detail see Section 6 of [8] (which is a preliminary version of the present paper). The protoform of the second construction appeared in a privately circulated manuscript [11] written in collaboration with Albert Visser, and is included here with his kind permission.

§2. Preliminaries.

In this section we present the relevant notations, conventions, definitions, and results that are needed in the subsequent sections.

2.1. *Models of arithmetic.*

DEFINITION 2.1. *The language of Peano Arithmetic, $\mathcal{L}_{\mathsf{PA}}$, is $\{+, \cdot, \mathrm{S}, 0\}$. We use the convention of writing M, M_0, N, etc. to (respectively) denote the universes of discourse of structures $\mathcal{M}$, $\mathcal{M}_0, \mathcal{N}$, etc. In what follows $\mathcal{M}$ and $\mathcal{N}$ are models of* PA.

(a) PA (Peano arithmetic) is the result of adding the scheme of induction for all $\mathcal{L}_{\mathsf{PA}}$-formulae to the finitely axiomatizable theory known as (Robinson's) Q.
(b) $\Sigma_0 = \Pi_0 = \Delta_0 =$ the collection of $\mathcal{L}_{\mathsf{PA}}$-formulae all of whose quantifiers are bounded by $\mathcal{L}$-terms (i.e., they are of the form $\exists x \leq t$, or of the form $\forall x \leq t$, where t is an $\mathcal{L}_{\mathsf{PA}}$-term not involving x). More generally, Σ_{n+1} consists of formulae of the form $\exists x_0 \ldots \exists x_{k-1}\ \varphi$, where $\varphi \in \Pi_n$; and Π_{n+1} consists of formulae of the form $\forall x_0 \ldots \forall x_{k-1}\ \varphi$, where $\varphi \in \Sigma_n$ (with the convention that $k = 0$ corresponds to an empty block of quantifiers).
(c) If $p \in M$ and $\varphi(\overline{x}.y)$ is an $\mathcal{L}_{\mathsf{PA}}$-formula, where $\overline{x}$ is an n-tuple of variables, $\varphi(\overline{x}, p)^{\mathcal{M}} := \{\overline{m} \in M^n : \mathcal{M} \models \varphi(\overline{m}, p\}$. For $X \subseteq M^n$, X is $\mathcal{M}$-*definable* if $X = \varphi(\overline{x}, p)^{\mathcal{M}}$ for some $\mathcal{L}_{\mathsf{PA}}$-formula $\varphi(\overline{x}.y)$ and for some parameter p in M.

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

(d) A subset X of M is *$\mathcal{M}$-finite* (*or $\mathcal{M}$-coded*) if $X = c_E$ for some $c \in M$, where $c_E = \{m \in M : \mathcal{M} \models m \in_{\text{Ack}} c\}$, and $m \in_{\text{Ack}} c$ is shorthand for the $\mathcal{L}_{\text{PA}}$-formula expressing "the mth bit of the binary expansion of c is 1".
(e) We identify the longest well-founded initial segment of models of PA with the ordinal ω. In this context, when $\mathcal{M}$ is nonstandard, we refer to members of ω as *standard* elements, and we refer to ω as the *standard cut* of $\mathcal{M}$.
(f) A subset I of M is a *cut* of $\mathcal{M}$ if I is an initial segment of $\mathcal{M}$ with no last element.

Both constructions presented in this paper employ known isomorphism results about countable recursively saturated models of arithmetic. The following classical result can be found in Kaye's monograph [18].

THEOREM 2.2. *Suppose $\mathcal{M}$ and $\mathcal{N}$ are countable recursively saturated models of* PA. *$\mathcal{M}$ and $\mathcal{N}$ are isomorphic iff* $\text{SSy}(\mathcal{M}) = \text{SSy}(\mathcal{N})$ *and* $\text{Th}(\mathcal{M}) = \text{Th}(\mathcal{N})$.

In order to state the next isomorphism theorem we need to recall the following definitions:

DEFINITION 2.3. *Suppose I is a proper cut of $\mathcal{M}$. I is* ω-coded from below (above) *in $\mathcal{M}$ iff for some $c \in M$, $\{(c)_n : n \in \omega\}$ is cofinal in I (downward cofinal in $M \setminus I$); here $(c)_n$ is the exponent of the* nth *prime in the prime factorization of* c *within $\mathcal{M}$.*

DEFINITION 2.4. *Suppose I is a cut of a model $\mathcal{M}$ of* PA. $\text{SSy}_I(\mathcal{M})$ *is the collection of subsets of I that are coded in $\mathcal{M}$. More precisely,* $\text{SSy}_I(\mathcal{M})$ *consists of sets of the form $c_E \cap I$, as* c *ranges in $\mathcal{M}$, where $c_E = \{x \in M : \mathcal{M} \models x \in_{\text{Ack}} c\}$ and $\in_{\text{Ack}}$ is the Ackermann membership defined by: $x \in_{\text{Ack}} c$ iff the* xth *digit of the binary expansion of* c *is* 1.

THEOREM 2.5 (Kossak–Kotlarski [19, theorem 2.1, corollary 2.3]). *Suppose $\mathcal{M}$ and $\mathcal{N}$ are countable recursively saturated models of* PA *with* $\text{Th}(\mathcal{M}) = \text{Th}(\mathcal{N})$, *and furthermore suppose that I is a cut shared by $\mathcal{M}$ and $\mathcal{N}$ such that* $\text{SSy}_I(\mathcal{M}) = \text{SSy}_I(\mathcal{N})$ *and I is not ω-coded from above in $\mathcal{M}$ or in $\mathcal{N}$. Then $\mathcal{M}$ and $\mathcal{N}$ are isomorphic over I, i.e., there is an isomorphism f between $\mathcal{M}$ and $\mathcal{N}$ such that $f(i) = i$ for each $i \in I$.*

REMARK 2.6. *Suppose $\mathcal{M}$ is a nonstandard model of* PA.

(a) Thanks to overspill, a cut I of $\mathcal{M}$ cannot be both ω-coded from above and ω-coded from below in $\mathcal{M}$. In particular, if c is a nonstandard element of $\mathcal{M}$, then the cut determined by finite powers of c is not ω-coded from above in $\mathcal{M}$. Thus there are arbitrarily large cuts in $\mathcal{M}$ that are not ω-coded from above in $\mathcal{M}$.
(b) It is easy to see that if I is a *strong* cut of $\mathcal{M}$, then I is not ω-coded from above.
(c) Suppose $\mathcal{M}$ is countable and recursively saturated. There are arbitrarily large strong cuts J of $\mathcal{M}$ such that $J \prec \mathcal{M}$ and $J \cong \mathcal{M}$. One way of seeing this is as follows: by the resplendence property of $\mathcal{M}$, $\mathcal{M}$ carries an inductive satisfaction class S. Next, use the Phillips–Gaifman refinement of the MacDowell–Specker theorem to build a countable conservative elementary end extension $(\mathcal{M}^*, S^*)$ of $(\mathcal{M}, S)$. Note that ACA_0 holds in $\big(\mathcal{M}, \text{SSy}_M(\mathcal{M}^*)\big)$ since $\mathcal{M}^*$ is a conservative extension of $\mathcal{M}$ and therefore by [20, theorem 7.3.2] $\mathcal{M}$ is strong in $\mathcal{M}^*$. Also observe that $\mathcal{M}^*$ is recursively saturated since $\mathcal{M}^*$ carries an inductive satisfaction class. By part (a), given any c in $\mathcal{M}$ there is a cut

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

I of $\mathcal{M}$ that includes c such that I is not ω-coded from above in $\mathcal{M}$. Since $\mathcal{M}^*$ is an end extension of $\mathcal{M}$, I is not ω-coded from above in $\mathcal{M}^*$ and $\mathrm{SSy}_I(\mathcal{M}) = \mathrm{SSy}_I(\mathcal{M}^*)$. Therefore by Theorem 2.5 there is an isomorphism $f : \mathcal{M}^* \to \mathcal{M}$ that is the identity on I. So if we let $J = f(M)$, it is evident that J is a strong cut of $\mathcal{M}$ that includes c, $J \prec \mathcal{M}$, and $J \cong \mathcal{M}$.

2.2. Satisfaction and truth classes. Truth and satisfaction are often used interchangeably in mathematical logic, and this conflation is also at work when it comes to the terms 'truth class' and 'satisfaction class'. In this subsection we provide the precise definitions of each. As we shall see in Proposition 2.14, a truth class is essentially an extensional satisfaction class.[3]

DEFINITION 2.7. *We will use the following abbreviations relating to the arithmetization of syntax*; *note that all the formulae in the list below are* $\mathcal{L}_{\mathsf{PA}}$*-formulae.*

(a) $\mathsf{Form}(x)$ is the formula expressing "x is (the code of) an $\mathcal{L}_{\mathsf{PA}}$-formula".
(b) $\mathsf{Sent}(x)$ is the conjunction of $\mathsf{Form}(x)$ and the formula expressing "x has no free variables".
(c) $\mathsf{Var}(x)$ is the formula expressing "x is (the code of) a variable".
(d) $\mathsf{Asn}(\alpha)$ is the formula expressing "α is the code of an assignment", where an assignment here simply refers to a function whose domain consists of a (finite) set of variables.
(e) $y \in \mathsf{FV}(x)$ is the formula expressing "$\mathsf{Form}(x)$ and y is a free variable of x".
(f) $y \in \mathsf{Dom}(\alpha)$ is the formula expressing "the domain of α includes y".
(g) $\mathsf{Asn}(\alpha, x)$ is the following formula expressing "α is an assignment for x":

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

(h) For assignments α and α', $\alpha' \supseteq \alpha$ expresses "the domain of α' extends the domain of α and $\alpha(v) = \alpha'(v)$ for all v in the domain of α".
(i) $x \lhd y$ 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 $y \in \mathsf{Form}$ and the following disjunction:

$$(y = \neg x) \vee \exists z \left((y = x \vee z) \vee (y = z \vee x)\right) \vee \exists v \in \mathsf{Var}\, (y = \exists v\, x).$$

(j) Given an $\mathcal{L}_{\mathsf{PA}}$-term s and an assignment α whose domain includes the free variables of s, $\mathsf{val}(s, \alpha)$ is the PA-definable function that outputs the value of s when its free variables are replaced by the values specified by α.

DEFINITION 2.8. *The theory* $\mathsf{CS}^{-}(\mathsf{F})$ *defined below is formulated in an* expansion *of* $\mathcal{L}_{\mathsf{PA}}$ *by adding a fresh* binary *predicate* $\mathsf{S}(x, y)$ *(denoting satisfaction) and a fresh unary predicate* F *(denoting a specified collection of formulae). The binary/unary distinction is of course not an essential one since* PA *has access to a definable pairing function.*

[3] For the subtle differences between satisfaction classes and truth classes see [4, 6]. Historically speaking, Krajewski [24] employed the framework of satisfaction classes over various theories formulated in relational languages, however, the later series of papers [23, 29, 30] all used the framework of truth classes over PA formulated in a relational language, augmented with 'domain constants' (which is the approach taken in [12]). Later, Kaye [18] developed the theory of satisfaction classes over models of PA in languages incorporating function symbols; Kaye's work was extended by Engström [13] to truth classes over models of PA in functional languages.

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

However, the binary/unary distinction at the conceptual level *marks the key difference between the concepts of satisfaction and truth*.

(a) $\mathsf{CS}^{-}(\mathsf{F})$ 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 with the proviso that in what follows α and α' range over assignments, and s and t range over $\mathcal{L}_{\mathsf{PA}}$-terms (that might have free variables). It is helpful to bear in mind that the axioms of $\mathsf{CS}^{-}(\mathsf{F})$ collectively express "F is a subset of arithmetical formulae that is closed under immediate subformulae; each member of S is an ordered pair of the form (x,α), where x is in F and α is an assignment for x; and S satisfies Tarski's compositional clauses for a satisfaction predicate".

- $\mathsf{tarski}_0(\mathsf{S},\mathsf{F}) := \big[\mathsf{F}(x) \to \mathsf{Form}(x)\big] \wedge \big[y \lhd x \wedge \mathsf{F}(x) \to \mathsf{F}(y)\big] \wedge \big[\mathsf{S}(x,\alpha) \to \big(\mathsf{F}(x) \wedge \alpha \in \mathsf{Asn}(x)\big)\big]$.
- $\mathsf{tarski}_1(\mathsf{S},\mathsf{F}) := \big[\mathsf{F}(x) \wedge x = {}^{\ulcorner}s = t^{\urcorner} \wedge \alpha \in \mathsf{Asn}(x)\big] \to \big[\mathsf{S}(x,\alpha) \leftrightarrow \mathsf{val}(s,\alpha) = \mathsf{val}(t,\alpha)\big]$.
- $\mathsf{tarski}_2(\mathsf{S},\mathsf{F}) := \big[\mathsf{F}(x) \wedge (x = \neg y) \wedge \alpha \in \mathsf{Asn}(x)\big] \to \big[\mathsf{S}(x,\alpha) \leftrightarrow \neg\mathsf{S}(y,\alpha)\big]$.
- $\mathsf{tarski}_3(\mathsf{S},\mathsf{F}) := \big[\mathsf{F}(x) \wedge (x = y_1 \vee y_2) \wedge \alpha \in \mathsf{Asn}(x)\big] \to \big[\mathsf{S}(x,\alpha) \leftrightarrow \big(\mathsf{S}\,(y_1, \alpha \restriction \mathsf{FV}(y_1)) \vee \mathsf{S}\,(y_2, \alpha \restriction \mathsf{FV}(y_2))\big)\big]$.
- $\mathsf{tarski}_4(\mathsf{S},\mathsf{F}) := \big[\mathsf{F}(x) \wedge (x = \exists v\ y) \wedge \alpha \in \mathsf{Asn}(x))\big] \to \big[\mathsf{S}(x,\alpha) \leftrightarrow \exists \alpha' \supseteq \alpha\ \mathsf{S}(y,\alpha')\big]$.

(b) CS^{-} is the theory whose axioms are obtained by substituting the predicate $\mathsf{F}(x)$ by the $\mathcal{L}_{\mathsf{PA}}$-formula $\mathsf{Form}(x)$ in the axioms of $\mathsf{CS}^{-}(\mathsf{F})$. Thus the axioms on CS^{-} are formulated in the language obtained by adding S to $\mathcal{L}_{\mathsf{PA}}$ (with no mention of F).

DEFINITION 2.9. *Let* $\mathcal{M} \models \mathsf{PA}$, *and suppose* $F \subseteq \mathsf{Form}^{\mathcal{M}} = \{m \in M : \mathcal{M} \models \mathsf{Form}(m)\}$.

(a) A subset S of M^2 is said to be an *F-satisfaction class on $\mathcal{M}$* if $(\mathcal{M}, F, S) \models \mathsf{CS}^{-}(\mathsf{F})$, here the interpretation of F is F and the interpretation of S is S. S is a *satisfaction class* on $\mathcal{M}$ if S is an F-satisfaction class for some F.

(b) 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 $(\varphi_1, \alpha_1) \in S$, where $(\varphi_0, \alpha_0) \sim (\varphi_1, \alpha_1)$ means that after substituting numerals in φ_0 in accordance with α_0 we obtain precisely the same formula obtained by substituting numerals in φ_1 in accordance with α_1.

(c) A subset T of M is said to be a *full satisfaction class on $\mathcal{M}$* if $(\mathcal{M}, S) \models \mathsf{CS}^{-}$ (equivalently: if $(\mathcal{M}, F, S) \models \mathsf{CT}^{-}(\mathsf{F})$) for $F = \mathsf{Form}^{\mathcal{M}}$.

DEFINITION 2.10. *The theory* $\mathsf{CT}^{-}(\mathsf{F})$ *defined below is formulated in an* expansion *of* $\mathcal{L}_{\mathsf{PA}}$ *by adding a fresh* unary *predicate* $\mathsf{T}(x)$ (*denoting satisfaction*) *and a fresh unary predicate* F (*denoting a specified collection of formulae*).

(a) $\mathsf{CT}^{-}(\mathsf{F})$ is the conjunction of the universal generalizations of the formulae $\mathsf{tarski}_0(\mathsf{S},\mathsf{F})$ through $\mathsf{tarski}_4(\mathsf{S},\mathsf{F})$. In what follows we use following conventions: $\mathsf{FSent}(x)$ expresses "x is an $\mathcal{L}_{\mathsf{PA}}$-sentence obtained by substituting closed terms of $\mathcal{L}_{\mathsf{PA}}$ for every free variable of a formula in F"; $y \lhd x$ expresses "y is an immediate subformula of x" as in Definition 2.7(i); s and t range over closed terms of $\mathcal{L}_{\mathsf{PA}}$; s° denotes the value of the closed term s, i.e., $s^{\circ} = \mathsf{val}(\alpha, \varnothing)$

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

(where val is as in Definition 2.7(j) and $\varnothing$ is the empty assignment); σ and its variants $(\sigma', \sigma_1, \sigma_2)$ range over elements of FSent; $\varphi(v)$ ranges over $\mathcal{L}_{\mathsf{PA}}$-formulae; $\mathsf{F}_{\leq 1}(\varphi(v))$ expresses "$\mathsf{F}(\varphi)$ and φ has at most one free variable v"; and $\varphi[\underline{x}/v]$ is (the code of) the formula obtained by substituting all occurrences of the variable v in φ with the numeral representing x.

- $\mathsf{tarski}_0(\mathsf{T},\mathsf{F}) := \big[\mathsf{T}(x) \to \mathsf{FSent}(x)\big] \wedge \big[y \lhd x \wedge \mathsf{F}(x) \to \mathsf{F}(y)\big]$.
- $\mathsf{tarski}_1(\mathsf{T},\mathsf{F}) := \mathsf{T}(s = t) \leftrightarrow s^\circ = t^\circ$.
- $\mathsf{tarski}_2(\mathsf{T},\mathsf{F}) := \big[\mathsf{FSent}(\sigma) \wedge (\sigma = \neg\sigma')\big] \to \big(\mathsf{T}(\sigma) \leftrightarrow \neg\mathsf{T}(\sigma')\big)$.
- $\mathsf{tarski}_3(\mathsf{T},\mathsf{F}) := \big[\mathsf{FSent}(\sigma) \wedge \sigma = \sigma_1 \vee \sigma_2\big] \to \big[\mathsf{T}(\sigma) \leftrightarrow \big(\mathsf{T}(\sigma_1) \vee \mathsf{T}(\sigma_2)\big)\big]$.
- $\mathsf{tarski}_4(\mathsf{T},\mathsf{F}) := \mathsf{F}_{\leq 1}(\varphi(v)) \to \big[\mathsf{T}(\exists v\, \varphi) \leftrightarrow \exists x\, \mathsf{T}(\varphi[\underline{x}/v])\big]$.

(b) CT^- is the theory whose axioms are obtained by substituting the predicate $\mathsf{F}(x)$ by the $\mathcal{L}_{\mathsf{PA}}$-formula $\mathsf{Form}(x)$ in the axioms of $\mathsf{CT}^-(\mathsf{F})$. Thus the axioms on CT^- are formulated in the language obtained by adding T to $\mathcal{L}_{\mathsf{PA}}$ (with no mention of F).

DEFINITION 2.11. *Let $\mathcal{M} \models \mathsf{PA}$, and suppose $F \subseteq \mathsf{Form}^{\mathcal{M}}$, and F is closed under direct subformulae of $\mathcal{M}$. Recall that $\mathsf{FSent}^{(\mathcal{M},F)}$ consists of $m \in M$ such that $(\mathcal{M}, F)$ satisfies "*m *is an $\mathcal{L}_{\mathsf{PA}}$-sentence obtained by substituting closed terms of $\mathcal{L}_{\mathsf{PA}}$ for the free variables of a formula in* F*".*

(a) A subset T of M is an *F-truth class* on $\mathcal{M}$ if $(\mathcal{M}, F, T) \models \mathsf{CT}^-(\mathsf{F})$, here the interpretation of F is F and the interpretation of T is T. T is a *truth class* on $\mathcal{M}$ if T is an F-truth class for some F.

(b) A subset T of M is a *full truth class* on $\mathcal{M}$ if $(\mathcal{M}, T) \models \mathsf{CT}^-$; equivalently: if $(\mathcal{M}, F, T) \models \mathsf{CT}^-(\mathsf{F})$ for $F = \mathsf{Form}^{\mathcal{M}}$.

(c) An F-truth class T on $\mathcal{M}$ is *F-disjunctively correct*, if whenever $\delta \in \mathsf{FSent}^{\mathcal{M}}$ and $\{\varphi_i : i < m\}$ is an $\mathcal{M}$-coded subset of $\mathsf{Sent}^{\mathcal{M}}$ (for some possibly nonstandard $m \in M$), the following holds:

$$\left(\mathcal{M} \models \delta = \bigvee_{i<m} \varphi_i\right) \Longrightarrow (\delta \in T \Leftrightarrow \exists i < m\ \varphi_i \in T)\,, \qquad (\bigtriangledown)$$

where any grouping of $\{\varphi_i : i < m\}$ can be used in $\mathcal{M}$ for forming the disjunction δ.

(d) Given a cut I of $\mathcal{M}$, an F-truth class T on $\mathcal{M}$ is *I-disjunctively correct* if T is disjunctively correct for disjunctions whose number of disjuncts is a member of I; more precisely, if $(\bigtriangledown)$ holds whenever $\delta \in \mathsf{FSent}^{\mathcal{M}}$ and $m \in I$.

(e) Let $\mathsf{Val}^{\mathcal{M}}$ be the set of theorems of first-order logic as computed in $\mathcal{M}$. Given an F-truth class T on $\mathcal{M}$, T is *F-deductively correct* if whenever $\sigma \in T$, $\sigma' \in \mathsf{FSent}^{\mathcal{M}}$, and $(\sigma \to \sigma') \in \mathsf{Val}^{\mathcal{M}}$, then $\sigma' \in T$.

(f) Given a cut I of $\mathcal{M}$ and a truth class T on $\mathcal{M}$, T is *I-deductively correct* if T is F-deductively correct for $F = \mathsf{Form}^{\mathcal{M}} \cap I$.

REMARK 2.12. *Let $\mathcal{M}$ be a nonstandard model of* PA, *and* $\mathrm{Sat}_{\mathcal{M}}$ *be the usual Tarskian truth predicate on $\mathcal{M}$.* $\mathrm{Sat}_{\mathcal{M}}$ *induces an ω-truth class* $\mathrm{True}_{\mathcal{M}}$ *on $\mathcal{M}$, where ω is the standard cut of $\mathcal{M}$. More specifically, suppose $\varphi(x_1, \dots, x_k)$ is a standard k-ary formula, and $t_1, \dots, t_k$ are (codes of) closed $\mathcal{L}_{\mathsf{PA}}$-terms in the sense of $\mathcal{M}$ (thus $t_1, \dots, t_k$ need not be standard). Since $\mathcal{M}$ is a model of* PA, *the closed-term-evaluation function $t \mapsto t^\circ$ is $\mathcal{M}$-definable, and therefore there are elements $m_1, \dots, m_k$ in M such that $\mathcal{M} \models t_i^\circ = m_i$*

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

for $1 \leq i \leq k$. *Thus* $\mathrm{True}_{\mathcal{M}}$ *can be defined as follows*: $\varphi(t_1/x_1, \ldots, t_k/x_k) \in \mathrm{True}_{\mathcal{M}}$ *iff* $(\varphi, \alpha) \in \mathrm{Sat}_{\mathcal{M}}$, *where* α *is the assignment given by* $x_i \mapsto m_i$. *It is evident that* $\mathrm{True}_{\mathcal{M}}$ *is* ω*-disjunctively and* ω*-deductively correct*.

REMARK 2.13. *It is well-known* [15, *sec. V.*5] *that for each nonzero* $n \in \omega$ *there is a unary* Σ_n*-formula* True_{Σ_n} *that serves as a universal* Σ_n*-predicate within* PA (*indeed* $\mathrm{I}\Delta_0 + \mathrm{Exp}$ *suffices for this purpose*). *Thus, if* $\mathcal{M} \models \mathsf{PA}$, *and* $F_n =$ *the set of* Σ_n*-formulae as computed in* $\mathcal{M}$, *then* $\mathrm{True}_{\Sigma_n}^{\mathcal{M}}$ *is an* F_n*-truth class that is* F_n*-disjunctively and* F_n*-deductively correct. Note that if* $\mathcal{M}$ *is nonstandard,* F_n *includes nonstandard formulae.*

The following proposition codifies the inter-definability of truth classes and extensional satisfaction classes. The close relationship between extensional satisfaction classes and truth classes was first made explicit in [12]; for further elaborations see [4, Chapter 7] and [31]. In what follows num is the PA-definable function $m \mapsto_{\mathsf{num}} \underline{m}$, where $\underline{m}$ is the numeral for $m \in M$, and $\varphi(\mathsf{num} \circ \alpha)$ is the $\mathcal{L}_{\mathsf{PA}}$-sentence obtained by replacing each occurrence of a free variable x of φ with $\underline{m}$, where $\alpha(x) = m$.

PROPOSITION 2.14. *Suppose* $\mathcal{M} \models \mathsf{PA}$, T *is an* F-truth class on $\mathcal{M}$, *and* S *is an extensional* F*-satisfaction class on* $\mathcal{M}$.

(a) $\mathcal{S}(T)$ *is an extensional* F*-satisfaction class on* $\mathcal{M}$, *where* $\mathcal{S}(T)$ *is defined as the collection of ordered pairs*(φ, α) *such that* $\varphi(\mathsf{num} \circ \alpha) \in T$.
(b) $\mathcal{T}(S)$ *is an* F*-truth class on*$\mathcal{M}$, *where* $\mathcal{T}(S)$ *is defined as the collection of* φ *such that* $(\varphi, \varnothing) \in S$ (*where* $\varnothing$ *is the empty assignment*).
(c) $\mathcal{S}(\mathcal{T}(S)) = S$, *and* $\mathcal{T}(\mathcal{S}(T)) = T$.

The following result plays a key role in this paper; for a modern exposition see [15, theorem I.4.33].

Mostowski's Reflection Theorem 2.15. *For each* $n \in \omega$, *the consistency of the set of sentences in* True_{Σ_n} *is provable in* PA. *In particular,* PA *proves the consistency of each of its finitely axiomatizable subtheories.*

REMARK 2.16. *Mostowski's Reflection Theorem, together with the Arithmetized Completeness Theorem allow us to start with any model* $\mathcal{M}$ *of* PA *and build an end extension* $\mathcal{N}$ *of* $\mathcal{M}$ *that satisfies* PA, *and moreover* $\mathcal{N}$ *is* strongly *interpretable in* $\mathcal{M}$ *in the sense that* $\mathcal{N}$ *is interpretable in* $\mathcal{M}$ *and there is an* $\mathcal{M}$*-definable* F*-truth class* $T_{\mathcal{N}}$ *on* $\mathcal{N}$ *for* $F = \mathsf{Form}^{\mathcal{M}}$. *To see this, consider the arithmetical formula* $\theta(i) := \mathrm{Con}(\mathsf{PA}_i)$, *where* PA_i *is the set of axioms of* PA *whose code is at most* i. *Mostowski's Reflection Theorem implies that* $\mathcal{M} \models \theta(n)$ *for each standard* n. *Hence by overspill* $\mathcal{M} \models \theta(e)$ *for some nonstandard element* e *of* $\mathcal{M}$. *By invoking the Arithmetized Completeness Theorem* (*as in* [18, *theorem* 13.13]) *within* $\mathcal{M}$ *we can conclude that there is a model* $\mathcal{N}$ *of* PA *whose elementary diagram is* $\mathcal{M}$*-definable*; *and moreover, by a straightforward internal recursion within* $\mathcal{M}$ *there is an* $\mathcal{M}$*-definable embedding* j *of* $\mathcal{M}$ *onto an initial segment of* $\mathcal{N}$. *Thus by identifying* $\mathcal{M}$ *with its image under* j, *we can assume without loss of generality that* $\mathcal{N}$ *is an end extension of* $\mathcal{M}$. *Note that the Arithmetized Completeness Theorem applied within* $\mathcal{M}$ *yields an* $\mathcal{M}$*-definable predicate* T_0 *such that* $(\mathcal{N}, T_0)$ *satisfies the slightly weaker form of* $\mathsf{CT}^-(\mathsf{F})$ *for* $F = \mathsf{Form}^{\mathcal{M}}$ *in which the closed terms are limited to those in* $\mathcal{M}$. *However, thanks to the availability of the closed-term-evaluation function in* $\mathcal{N}$, *the simple 'trick' used in Remark* 2.12 *can be used to obtain an* $\mathcal{M}$*-definable predicate* T *such that* $(\mathcal{N}, F, T)$ *satisfies* $\mathsf{CT}^-(\mathsf{F})$. *Note that since* $\mathcal{N}$ *is an end extension of* $\mathcal{M}$, $\mathrm{True}_{\Sigma_1}^{\mathcal{M}} \subseteq \mathrm{True}_{\Sigma_1}^{\mathcal{M}}$, *and*

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

therefore if s *is a closed arithmetical term in the sense of* $\mathcal{M}$, *then* $\mathrm{val}^{\mathcal{M}}(s) = \mathrm{val}^{\mathcal{N}}(s)$. *Since* T_0 *is the Tarskian satisfaction class of the structure* $\mathcal{N}$ *as computed in the model* $\mathcal{M}$ *of* PA, T_0 *as well as* T *have many regularity properties of Tarskian satisfaction classes*; *in particular* T *is both* M*-deductively and* M*-deductively closed. As shown in* (*Stage* 1 *of*) *the proof of Theorem* 3.3, *if* $\mathcal{M}$ *is a countable recursively saturated model of* PA, *we can additionally require that there is an isomorphism* $f : \mathcal{N} \to \mathcal{M}$. *Therefore by defining the cut* I *of* $\mathcal{M}$ *as* $f(M)$, *the image* $f(T)$ *of* T *is under* f *is an* $F \cap I$*-satisfaction class on* $\mathcal{M}$ *such that* $f(T)$ *is both* I*-disjunctively and* I*-deductively correct*.

§3. The first construction. Suppose I is a cut of a countable recursively saturated model $\mathcal{M}$ of PA. In Definition 3.1 we define a subset $\mathsf{Frug}_I^{\mathcal{M}}$ of $\mathsf{Form}^{\mathcal{M}}$, and then in Theorem 3.3 we build an F-truth class T on $\mathcal{M}$ for $F = \mathsf{Frug}_\omega^{\mathcal{M}}$ (i.e., for $I = \omega$, where ω is the standard cut of $\mathcal{M}$) such that T is F-disjunctively correct (in the sense of Definition 2.11(c)). In Theorem 3.4 we extend this result to arbitrarily large cuts I of $\mathcal{M}$. Theorem 3.3 and 3.4 could have been packaged as a single result, but in the interest of offering the reader a clear intuition of the mechanism of the first construction we opted for the current format; as a result we provide the full details of the proof of Theorem 3.3 and only offer a proof outline for Theorem 3.4.

DEFINITION 3.1. *A formula* φ *is* frugal *if* φ *has no occurrence of a closed term. We use* $\mathsf{Frug}(\varphi)$ *for the* $\mathcal{L}_{\mathsf{PA}}$*-formula that expresses "*φ *is a frugal* $\mathcal{L}_{\mathsf{PA}}$*-formula". Given* $\mathcal{M} \models \mathsf{PA}$, *and* $m \in M$, *we define*:

$$\mathsf{Frug}_{\leq m}^{\mathcal{M}} := \{\varphi \in M : \mathcal{M} \models \big(\mathsf{Frug}(\varphi) \wedge \text{“}\varphi \text{ has at most } m \text{ distinct free variables”}\big)\}.$$

$$\text{For a cut } I \text{ of } \mathcal{M},\ \mathsf{Frug}_I^{\mathcal{M}} := \bigcup_{i \in I} \mathsf{Frug}_{\leq i}^{\mathcal{M}}.$$

REMARK 3.2. *The above definitions make it evident that for any cut* I *of* $\mathcal{M}$ *that is closed under addition,* $\mathsf{Frug}_I^{\mathcal{M}}$ *is closed under Boolean connectives, existential quantification, and immediate subformulae* (*here we are assuming the coding of formulae is done in a standard way that ensures that the code of every immediate subformulae of a given formula* φ *is less than the code of* φ). *In particular, if* T *is an* F*-truth class on* $\mathcal{M}$ *for* $F = \mathsf{Frug}_I^{\mathcal{M}}$ *and* σ *is a sentence in* $\mathcal{M}$ *that is obtained from replacing all of the free variables of a formula in* F *with closed terms, i.e.,* $\sigma \in \mathsf{FSent}^{(\mathcal{M},F)}$, *then by* $\mathsf{tarski}_2(\mathsf{T},\mathsf{F})$ *either* $\sigma \in T$ *or* $\neg\sigma \in T$ (*but not both*). *Also note that in this context the sentences* $\sigma \in \mathsf{FSent}^{(\mathcal{M},F)}$ *are precisely those sentences* $\sigma \in \mathsf{Sent}^{\mathcal{M}}$ *such that for some* $i \in I$ *the number of distinct closed terms that occur in* σ *is* i (*as computed in* $\mathcal{M}$).

THEOREM 3.3. *Suppose* $\mathcal{M}$ *is a countable recursively saturated model of* PA, *and let* $F = \mathsf{Frug}_\omega^{\mathcal{M}}$. *There is an* F*-truth class* T *on* $\mathcal{M}$ *that is* F*-disjunctively correct.*

Proof. The proof has two stages.

STAGE 1. In this stage, starting with a countable recursively saturated model $\mathcal{M}$ of PA, we use a variant of the construction outlined in Remark 2.16 to build an appropriate end extension $\mathcal{N}$ of $\mathcal{M}$ with two key properties: $\mathcal{M}$ and $\mathcal{N}$ are isomorphic, and yet there is an F-truth class $T_{\mathcal{N}}$ on $\mathcal{N}$ that is definable in $\mathcal{M}$, where $F = \mathsf{Form}^{\mathcal{M}}$. To build such an $\mathcal{N}$, we first observe that by recursive saturation we can find an element c in $\mathcal{M}$ that codes $\mathrm{Th}(\mathcal{M})$ by realizing the recursive type $\Sigma(v)$, where

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

$$\Sigma(v) := \{\varphi \leftrightarrow (\ulcorner\varphi\urcorner \in_{\mathsf{Ack}} v) : \varphi \text{ is sentence of } \mathcal{L}_{\mathsf{PA}}\},$$

where $\ulcorner\varphi\urcorner$ is the Gödel number for φ, $\in_{\mathsf{Ack}}$ is "Ackermann's $\in$", as in Definition 2.1(d). Next, let $\langle\varphi_n : n \in \omega\rangle$ be a recursive enumeration of $\mathcal{L}_{\mathsf{PA}}$-sentences, and consider the arithmetical formula $\gamma(i)$ defined below:

$$\gamma(i) := \mathrm{Con}\left(\{\varphi_j : j < i \wedge \varphi_j \in_{\mathsf{Ack}} c\}\right),$$

where $\mathrm{Con}(X)$ expresses the formal consistency of X. It is easy to see, using Mostowski's Reflection Theorem and our choice of c that $\mathcal{M} \models \gamma(n)$ for each standard n. Hence by overspill $\mathcal{M} \models \gamma(d)$ for some nonstandard element d of $\mathcal{M}$. By invoking the Arithmetized Completeness Theorem [18, theorem 13.13] within $\mathcal{M}$, with a reasoning similar to that in Remark 2.16 we can conclude that there is a model $\mathcal{N}$ of T with the following properties:

(1) $\mathrm{Th}(\mathcal{M}) = \mathrm{Th}(\mathcal{N})$.
(2) There is an $\mathcal{M}$-definable $T_{\mathcal{N}}$ such that $(\mathcal{N}, \mathsf{Form}^{\mathcal{M}}, T) \models \mathsf{CT}^{-}(\mathsf{F})$.
(3) $\mathcal{N}$ is recursively saturated.
(4) There is an $\mathcal{M}$-definable embedding j of $\mathcal{M}$ as an initial segment of $\mathcal{N}$.
(5) $\mathrm{SSy}(\mathcal{M}) = \mathrm{SSy}(\mathcal{N})$.
(6) There is an isomorphism $f : \mathcal{M} \to \mathcal{N}$.

Note that (3) is a consequence of (2) and a routine overspill argument, as in Proposition 15.4 of [18]; (5) is consequence of (4) since standard systems are preserved by end extension; and (6) follows from (1) and (5) by Theorem 2.2.

STAGE 2. In this stage we construct the desired F-truth class T on $\mathcal{M}$ for $F = \mathsf{Frug}_{\omega}^{\mathcal{M}}$. Suppose $\varphi(x_1, x_2, \ldots, x_k) \in F$, where $k \in \omega$, and let $t_1, t_2, \ldots, t_k$ be elements of $\mathsf{ClTerm}^{\mathcal{M}}$ (closed terms in the sense of $\mathcal{M}$). Note that φ as well as $t_1, t_2, \ldots, t_k$ are allowed to be nonstandard. Using the truth class $T_{\mathcal{N}}$ as in (2) of Stage 1, together with the isomorphism f as (6) of Stage 1, we define T to consist of $\varphi(t_1, t_2, \ldots, t_k) \in \mathsf{FSent}^{(\mathcal{M},F)}$ such that

$$\varphi(\underline{f(m_1)}, \underline{f(m_2)}, \ldots, \underline{f(m_k)}) \in T_{\mathcal{N}},$$

where $m_1, m_2, \ldots, m_k$ are elements of M such that $\mathcal{M} \models m_i = t_i^{\circ}$ for $1 \leq i \leq k$. In other words,

$$\varphi(t_1, t_2, \ldots, t_k) \in T \text{ iff } \varphi(\underline{f(t_1^{\circ})}, \underline{f(t_2^{\circ})}, \ldots, \underline{f(t_k^{\circ})}) \in T_{\mathcal{N}}.$$

In the above we simply write t° for the element of $\mathcal{M}$ such that $\mathcal{M} \models m = t^{\circ}$. As noted in the last sentence of Remark 2.16, since $t \in M$, t° as-computed-in-$\mathcal{M}$ coincides with t° as-computed-in $\mathcal{N}$.

We first verify that T is an F-truth class on $\mathcal{M}$. By the definition of T, the axiom $\mathrm{tarski}_0(\mathsf{F}) = \forall x \ \left[\mathsf{T}(x) \to x \in \mathsf{FSent}\right]$ clearly holds in $(\mathcal{M}, F, T)$. Recall that $\mathrm{tarski}_1(\mathsf{F}, \mathsf{T})$ asserts the equivalence of $\mathsf{T}(s = t)$ and $s^{\circ} = t^{\circ}$ for all closed terms s and t. To see that $(\mathcal{M}, F, T) \models \mathrm{tarski}_1(\mathsf{F})$ we argue as follows:

$$(s = t) \in T \text{ iff } \left(\underline{f(s^{\circ})} = \underline{f(t^{\circ})}\right) \in T_{\mathcal{N}},$$

[by the definition of T]

$$\left(\underline{f(s^{\circ})} = \underline{f(t^{\circ})}\right) \in T_{\mathcal{N}} \text{ iff } f(s^{\circ}) = f(t^{\circ}),$$

[since $(\underline{x})^{\circ} = x$ and $(\mathcal{N}, F, T_{\mathcal{N}}) \models \mathrm{tarski}_0(\mathsf{F})$]

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

$$f(s^\circ) = f(t^\circ) \text{ iff } s^\circ = t^\circ.$$
[since f is injective].

Next we verify that $(\mathcal{M}, F, T)$ satisfies $\text{tarski}_2(\mathsf{F}, \mathsf{T})$, which stipulates that T commutes with negation. Suppose $\varphi(t_1, t_2, \dots, t_k) = \neg\psi(t_1, t_2, \dots, t_k) \in \mathsf{FSent}^{(\mathcal{M},F)}$. Then we have:

$$\neg\psi(t_1, t_2, \dots, t_k) \in T \text{ iff } \neg\psi(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}},$$
[by the definition of T]
$$\neg\psi(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}} \text{ iff } \psi(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \notin T_{\mathcal{N}},$$
[since $(\mathcal{N}, F, T_{\mathcal{N}}) \models \text{tarski}_2(\mathsf{F})$]
$$\psi(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \notin T_{\mathcal{N}} \text{ iff } \psi(t_1, t_2, \dots, t_k) \notin T.$$
[by the definition of T]

An argument similar to the above shows that T commutes with disjunction, thus $\text{tarski}_3(\mathsf{F}, \mathsf{T})$ holds in $(\mathcal{M}, F, T)$; we leave the proof for the reader.

The axiom $\text{tarski}_4(\mathsf{F})$ stipulates that T commutes with existential quantification. For this purpose suppose $\varphi(t_1, t_2, \dots, t_k) = \exists v\ \psi(x, t_1, \dots, t_k) \in \mathsf{FSent}^{(\mathcal{M},F)}$. Then we have:

$$\exists v\ \psi(x, t_1, \dots, t_k) \in T \text{ iff } \exists v\ \psi(v, \underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}},$$
[by the definition of T]
$$\exists v\ \psi(v, \underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}} \text{ iff } \exists b \in N\psi(\underline{b}, \underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}},$$
[since $(\mathcal{N}, F_0, T_{\mathcal{N}}) \models \text{tarski}_4(\mathsf{F})$]
$$\exists b \in N\psi(\underline{b}, \underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}} \text{ iff } \exists c \in M\ f(c) = b, \psi(\underline{c}, t_1, \dots, t_k) \in T.$$
[by the definition of T, together with surjectivity of f]

This concludes our verification that T is an F-satisfaction class on $\mathcal{M}$.

To see that T is F-disjunctively correct suppose $\delta = \bigvee_{i<m} \varphi_i \in \mathsf{FSent}^{(\mathcal{M},F)} = \mathsf{Sent}_\omega^{\mathcal{M}}$ (where m is a possibly nonstandard element of $\mathcal{M}$). Note that each φ_i can be written as $\varphi_i(t_1, \dots, t_k)$ for some $k \in \omega$ with the understanding that the closed terms occurring in φ_i are among $t_1, \dots, t_k$. Then we have:

$$\left(\bigvee_{i<m} \varphi_i(t_1, \dots, t_k)\right) \in T \text{ iff } \left(\bigvee_{i<m} \varphi_i(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)})\right) \in T_{\mathcal{N}},$$
[by the definition of T]
$$\left(\bigvee_{i<m} \varphi_i(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)})\right) \in T_{\mathcal{N}} \text{ iff } \exists i < m \varphi_i(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}},$$
[since $T_{\mathcal{N}}$ is F-disjunctively correct and $\delta \in \mathsf{FSent}^{(\mathcal{M},F)}$],
$$\exists i < m \varphi_i(\underline{f(t_1^\circ)}, \underline{f(t_2^\circ)}, \dots, \underline{f(t_k^\circ)}) \in T_{\mathcal{N}} \text{ iff } \exists i < m \bigvee_{i<m} \varphi_i(t_1, \dots, t_k) \in T.$$
[by the definition of T] □

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

THEOREM 3.4. *Suppose $\mathcal{M}$ is a countable recursively saturated model of* PA. *There are arbitrarily large cuts $I \prec \mathcal{M}$ such that $I \cong \mathcal{M}$ with the property that there is an F-truth class T on $\mathcal{M}$ for $F = \mathsf{Frug}_I^{\mathcal{M}}$ such that T is F-disjunctively correct.*

Proof outline. Proceed as in Stage 1 of the proof of Theorem 3.3 to get hold of the model $\mathcal{N}$, but in Stage 2 invoke Remark 2.6(c) and use Theorem 2.5 instead of Theorem 2.2 to get hold of an isomorphism f between $\mathcal{M}$ and $\mathcal{N}$ that pointwise fixes a prescribed cut of $\mathcal{M}$. Note that by Remark 2.6(c) there are arbitrarily large strong cuts I in $\mathcal{M}$ such that $I \prec \mathcal{M}$ and $I \cong \mathcal{M}$, and by Remark 2.6(b) such cuts are not ω-coded from above in $\mathcal{M}$. □

REMARK 3.5. *The proofs of Theorem 3.3 and 3.4 show that these two results can be strengthened by requiring the truth predicate T to satisfy further desirable properties such as* alphabetic correctness, *which stipulates that T is invariant under the renaming of bound variables, and* generalized term-extensionality, *which stipulates that T is invariant under replacing of terms with the same value. As shown by Łełyk and Wcisło* [27, *theorem* 23] *the theory obtained by the addition of axioms stipulating alphabetic correctness and generalized term extensionality to* $\mathsf{CT}^-[\mathsf{PA}]$ *remains conservative over* PA.

REMARK 3.6. *In answer to a question of the author, Lawrence Wong noted that Wcisło's proof of Lachlan's theorem (as presented in* [21]*) shows that if $\mathcal{M}$ is a model of* PA *that has an expansion to a $\mathsf{Form}_\omega^{\mathcal{M}}$-truth class, then $\mathcal{M}$ is recursively saturated.*

QUESTION 3.7. *Does every countable recursively saturated model $\mathcal{M}$ of* PA *carry a* full *truth class T that is $\mathsf{Form}_\omega^{\mathcal{M}}$-disjunctively correct? More specifically, are the truth classes constructed in Theorem 3.3 and 3.4 extendable to full truth classes?*

§4. The second construction. The *second* construction in this paper employs an arithmetized form of the main construction in [12] as in Theorem 4.1. Our method of proof of Theorem 4.1 is through an arithmetization of the construction of a truth class satisfying of $\mathsf{CT}^-[\mathsf{PA}]$, presented both in [9, 21], which refine the model-theoretic construction given in [12] for PA formulated in a relational language. As we shall see, the compactness and elementary chain argument used in the model-theoretic conservativity proof of $\mathsf{CT}^-[\mathsf{PA}]$ over PA can be proved in the fragment $\mathrm{I}\Sigma_2$ of PA with the help of the so-called Low Basis Theorem of Recursion Theory.[4] Note that the results of this section that appear before Theorem 4.5 can be also established by the arithmetization method presented in [9], or by taking advantage of the main proof-theoretic result of [25].

THEOREM 4.1 (Joint with Albert Visser). *Suppose* $\mathsf{PA} \vdash \mathrm{Con}(\mathsf{B})$, *where* B *is some recursively axiomatizable theory extending* $\mathrm{I}\Delta_0 + \mathrm{Exp}$. *Then* PA *strongly interprets* $\mathsf{CT}^-[\mathsf{B}]$, *i.e., there are definitions δ_1 and δ_2 within* PA *such that for every model $\mathcal{M}$ of* PA, *$\delta_1^{\mathcal{M}} = (\mathcal{M}^*, T) \models \mathsf{CT}^-[\mathsf{B}]$ and $\delta_2^{\mathcal{M}}$ is the elementary diagram of $(\mathcal{M}^*, T)$ as viewed from $\mathcal{M}$.*

Proof. Before starting the proof, we need to review some key definitions and ideas from Recursion Theory. In what follows X and Y are subsets of ω.

[4] Indeed, using the technology of LL_1-sets of [15, theorem 4.2.7.1, p. 104], the main argument presented in this section can be carried out in $\mathrm{I}\Sigma_1$.

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

- X is *low*-Δ_2, if X is Δ_2, and $X' \leq_T 0'$ (where $\leq_T$ denotes Turing reducibility, and Z' is the Turing-jump of Z).
- More generally, Y is *low*-Δ_2 *in the oracle* X, if Y is Δ_2 in the oracle X, and $Y' \leq_T X'$.

By classical recursion theory, we have:

(1) $X \leq_T Y'$ iff X is Δ_2 in the oracle Y. Next, observe that if $X' \leq_T 0'$ and $Y' \leq_T X'$, then $Y' \leq_T 0'$, hence Y is low-Δ_2 by (1). Therefore:
(2) If X is low-Δ_2, and Y is low-Δ_2 in the oracle X, then Y is low-Δ_2.

The classical 'Low Basis Theorem' of Jockusch and Soare [17] asserts that every infinite finitely branching recursive tree has an infinite branch B such that B is low-Δ_2. Moreover, it is known that the Low Basis Theorem is provable in $\mathrm{I}\Delta_0 + \mathrm{B}\Sigma_2$ (this is due to Clote, whose argument is presented in [15, chap. I, sec. 3(c)]). In particular, $\mathrm{I}\Delta_0 + \mathrm{B}\Sigma_2$ can prove that every countable consistent Δ_1-set of first-order sentences has a low-Δ_2 completion. Since the Henkin construction of a model of a prescribed complete theory can already be performed in $\mathrm{I}\Sigma_1$, this shows that

(3) $\mathrm{I}\Delta_0 + \mathrm{B}\Sigma_2$ proves that every consistent Δ_1 set of first-order sentences has a model $\mathcal{M}$ such that the satisfaction predicate $\mathsf{Sat}_{\mathcal{M}}$ of $\mathcal{M}$ is low-Δ_2.

We are now ready to establish Theorem 4.1. We argue model-theoretically for ease of exposition, and we will work with satisfaction classes instead of truth classes.

- From this point on, we assume the reader is familiar with [12] and follow the notation therein.

Let $\mathcal{M} \models \mathsf{PA}$, and B be as in the assumption of the theorem. By (3) there is a model $\mathcal{N}$ of B that is strongly interpretable in $\mathcal{M}$ such that $\mathsf{Sat}_{\mathcal{N}}$ is low-Δ_2. Using a truth-predicate for Δ_2-predicates we can execute the construction of Lemma 3.1 of [12] (the same idea can be applied to the counterpart of Lemma 3.1 in the proof of the conservativity of [9, 21]); the key idea is that if $\mathsf{Sat}_{(\mathcal{N},S)}$ is low-Δ_2, then so is $\mathsf{Th}^+(\mathcal{N})$, where $\mathsf{Th}^+(\mathcal{N})$ is the theory defined as in [12, Lemma 3.1] involving extra predicates added to the language of arithmetic whose consistency is established by building a partial satisfaction predicate 'by hand' by examining an arbitrarily chosen finite configuration of $\mathcal{N}$-formulae. Therefore there is some elementary extension $\mathcal{N}_1$ of $\mathcal{N}$, and some $\mathsf{Form}^{\mathcal{N}}$-satisfaction class $S_1 \supseteq S$ such that $\mathsf{Sat}_{(\mathcal{N}_1,S_1)}$ is also low-Δ_2, thanks to (2) and the Low Basis Theorem. This shows that the construction

$$(\mathcal{N}, S) \mapsto (\mathcal{N}_1, S_1)$$

can be uniformly described in $\mathrm{I}\Sigma_2 + \mathrm{Con}(\mathsf{B})$. This allows one to obtain an $\mathcal{M}$-definable increasing sequence of structures $\langle (\mathcal{M}_i, S_i) : i \in M \rangle$ (as in the proof of Theorem 3.3 of [12]) with the additional feature that each $(\mathcal{M}_i, S_i)$ is strongly interpretable in $\mathcal{M}$. Consider the limit model:

$$(\mathcal{N}, S) = \bigcup_{i \in M} (\mathcal{M}_i, S_i)\,.$$

$(\mathcal{N}, S)$ is clearly interpretable in $\mathcal{M}$, and additionally (thanks to the elementary chains theorem applied within $\mathcal{M}$) the reduct $\mathcal{N}$ of $(\mathcal{N}, S)$ is strongly interpretable in $\mathcal{M}$, thus $\mathcal{M}$ "knows" that $\mathcal{N}$ is a model of B.

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

However, there is no reason to expect that the expansion $(\mathcal{N}, S)$ is strongly interpretable in $\mathcal{M}$. To circumvent this problem, we appeal to a trick found by Łełyk (first introduced in [9]) to take advantage of $(\mathcal{N}, S)$ in order to show that $\mathcal{M}$ satisfies $\mathrm{Con}(\mathsf{CT}^-[\mathsf{B}])$, thereby concluding the existence of a model of $\mathsf{CT}^-[\mathsf{B}]$ that is strongly interpretable in $\mathcal{M}$. The key idea is to resort to (1) the classical fact of proof theory that any deduction in first order logic can be replaced by another deduction with same conclusion but which has the subformula property,[5] (2) the fact that $\mathsf{CT}^-[\mathsf{B}]$ is the result of augmenting B with only finitely many truth axioms, and (3) the fact that $\mathcal{M}$ has full access to the elementary diagram of $\mathcal{N}$. The veracity of $\mathrm{Con}(\mathsf{CT}^-[\mathsf{B}])$ within $\mathcal{M}$ then allows us to invoke the completeness theorem of first-order logic within $\mathcal{M}$ to get hold of a model of $\mathsf{CT}^-[\mathsf{B}]$ that is strongly interpretable in $\mathcal{M}$. □

COROLLARY 4.2. PA *strongly interprets* $\mathsf{CT}^-[\mathsf{PA}]$.

Proof. Let $\mathcal{M}$ be an arbitrary model of PA. Within $\mathcal{M}$, if PA is inconsistent let B be $\mathrm{I}\Sigma_k$, where k is the first n such that $\mathrm{I}\Sigma_{n+1}$ is inconsistent, otherwise let $\mathsf{B} = \mathsf{PA}$. Thus $\mathrm{Con}(\mathsf{B})$ holds in $\mathcal{M}$. By Mostowski's Reflection Theorem, B extends $\mathrm{I}\Sigma_n$ for each $n \in \omega$, and therefore B extends PA (from an external point of view). Thus Theorem 4.1 applies. □

COROLLARY 4.3. $\mathsf{PA} \vdash \mathrm{Con}(\mathsf{CT}^-[\mathsf{B}])$ *for every finitely axiomatized subtheory* B *of* PA *that extends* $\mathrm{I}\Delta_0 + \mathsf{Exp}$.

Proof. Let B be a finitely axiomatizable subtheory of PA. The formal consistency of B_0 is verifiable in PA by Mostowski's Reflection Theorem. Therefore by Theorem 4.1 PA can produce an internal model of $\mathsf{CT}^-[\mathsf{B}]$ for which it has a satisfaction predicate; which in turn immediately implies that $\mathrm{Con}(\mathsf{CT}^-[\mathsf{B}])$ is provable in PA. □

COROLLARY 4.4. $\mathsf{CT}^-[\mathsf{PA}]$ *is not finitely axiomatizable.*

Proof. Put Corollary 4.3 together with Gödel's second incompleteness theorem. □

We are now ready to present the main result of this section. Recall that the notions of I-disjunctive correctness and I-deductive correctness were defined in Definition 2.11(d,f).

THEOREM 4.5 (Joint with Albert Visser). *Let $\mathcal{M}$ be a countable recursively saturated model of* PA. *There are arbitrarily large cuts I of $\mathcal{M}$ for which $I \prec \mathcal{M}$ and $I \cong \mathcal{M}$, and for each such cut I there is a full truth class T on $\mathcal{M}$ that is I-disjunctively and I-deductively correct.*

Proof. Let $\mathcal{M}$ be a countable recursively saturated model of PA. Let $\mathcal{N}$ be as in the proof of Stage 1 of the proof of Theorem 3.4, thus $\mathcal{N}$ is an elementary extension of $\mathcal{M}$ that is strongly interpreted in $\mathcal{M}$ and $\mathrm{SSy}(\mathcal{M}) = \mathrm{SSy}(\mathcal{N})$. Thanks to Corollary 4.3 there is an elementary extension $\mathcal{N}^*$ of $\mathcal{N}$ that carries a truth class $T_{\mathcal{N}^*}$ such that $(\mathcal{N}^*, T_{\mathcal{N}^*}) \models \mathsf{CT}^-[\mathsf{PA}]$, and $(\mathcal{N}^*, T_{\mathcal{N}^*})$ is strongly interpretable in $\mathcal{M}$, which in turn will allow us to conclude that $(\mathcal{N}^*, T_{\mathcal{N}^*})$ satisfies the following four properties:

[5] This fact is an immediate consequence of the cut-elimination theorem, which is provable in PA; indeed it is well-known that cut-elimination is already provable in the fragment $\mathrm{I}\Delta_0 + \mathsf{Supexp}$ of Primitive Recursive Arithmetic, where Supexp expresses the totality of superexponentiation function (also known as tetration). See [2] for more detail.

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

(1) $T_{\mathcal{N}^*}$ is M-disjunctively and M-deductively correct. This is handled the by usual argument by induction that shows that the classical Tarskian truth predicate respects arbitrary disjunctions and is closed under deductions, as noted in Remark 2.12.
(2) There is an $\mathcal{M}$-definable embedding of $\mathcal{M}$ as an initial segment of $\mathcal{N}^*$.
(3) $\mathrm{SSy}(\mathcal{M}) = \mathrm{SSy}(\mathcal{N}^*)$.
(4) $\mathcal{N}^*$ is recursively saturated.

Note that $\mathrm{Th}(\mathcal{M}) = \mathrm{Th}(\mathcal{N}^*)$ since $\mathcal{N}^*$ elementarily extends $\mathcal{N}$, and $\mathrm{Th}(\mathcal{M}) = \mathrm{Th}(\mathcal{N})$. Together with (3), (4) and Theorem 2.2, this allows us to conclude that $\mathcal{M} \cong \mathcal{N}^*$, which implies that we 'copy over' $T_{\mathcal{N}^*}$ on $\mathcal{M}$, thus $\mathcal{M}$ carries a full truth T class that is I-disjunctively and I-deductively correct for some cut I of $\mathcal{M}$. By using Theorem 2.5 and Remark 2.6 we can conclude that such cuts I can be arranged to be arbitrarily large in $\mathcal{M}$. More specifically, by Remark 2.6(c) there are arbitrarily large strong cuts I in $\mathcal{M}$ such that $I \prec \mathcal{M}$ and $I \cong \mathcal{M}$, and by Remark 2.6(b) such cuts are not ω-coded from above in $\mathcal{M}$. Therefore by Theorem 2.5 for such a given cut I there is an isomorphism $f : \mathcal{N}^* \to \mathcal{M}$ with $f(i) = i$ for all $i \in I$. This makes it clear that $T := f(T_{\mathcal{N}^*})$ is the desired full truth class on $\mathcal{M}$. □

REMARK 4.6. *Suppose* $\mathcal{M} \models \mathsf{PA} + \neg\mathrm{Con}(\mathsf{PA})$ *and* $\mathcal{M}$ *is countable and recursively saturated. Then* $\mathcal{M}$ *carries an inductive satisfaction class* S [18, *corollary* 15.12]. *For such an* S, *there is a topped nonstandard initial segment* I *of* $\mathcal{M}$ *such that* S *satisfies the Tarski conditions for all formulas in* I. *If* $\mathcal{M}$, *and* $\pi \in M$ *codes up a proof of inconsistency of* PA *in* $\mathcal{M}$, *then* $\pi \notin I$. *As shown by Abdul-Quader and Łełyk* [1, *proposition* 43], *the methodology employed in* [7] *can be used to show that we cannot expect to build a full satisfaction class* S *that is disjunctively correct on a cut* I *that contains* π *if* S *satisfies both the regularity axiom* (*also known as generalized term extensionality, as in Remark* 3.5) *and also internal induction. This is in contrast to Theorem* 4.5 *that shows that a full satisfaction class that is* I*-disjunctively correct can always be arranged for arbitrarily high cuts* I *in* $\mathcal{M}$. *Thus, even though the full satisfaction classes constructed in Theorem* 4.5 *can be arranged to satisfy the regularity axiom, in general they do not satisfy the internal induction axiom.*

REMARK 4.7. *Suppose* $\mathsf{T}(x, y)$ *is a binary predicate; we will write it as* $\mathsf{T}_x(y)$ *to indicate that for a fixed* x, T_x *is a truth predicate. Let* CT^* *be the sentence in the language of* PA *augmented with* $\mathsf{T}_x(y)$ *that says that for all* x *the axioms of* CT^- *hold for* T_x. *Then by putting Theorem* 4.5 *together with the completeness theorem of first-order logic, and the fact that every countable model of* PA *has a countable elementary extension that is recursively saturated we can conclude that the theory* $\mathsf{PA} + \mathsf{CT}^* +$ "$\forall x \mathsf{T}_x$ *is disjunctively correct for disjunctions whose number of disjuncts is at most* x" *is conservative over* PA.

REMARK 4.8. *As pointed by the referee, one can give a "soft" proof of Theorem* 4.5 *for the special case when* $\mathcal{M}$ *is a model of True Arithmetic* (*the theory of the standard model of* PA). *The argument goes as follows. It is known that if* I *is a cut of a countable recursively saturated model of* PA *that properly contains the Skolem closure of* 0 *in* $\mathcal{M}$, *then for every* $a \in I$ *and every* $c \in M$, *there is a* $b \in M$ *such that* $c < b$ *and the type of* a *in* $\mathcal{M}$ *coincides with the type of* b *in* $\mathcal{M}$. *This observation allows one to build an automorphism* f *of* $\mathcal{M}$ *such that* $f(a) = b$, *which in turn shows that* $(\mathcal{M}, I) \cong (\mathcal{M}, J)$ *for* $J = f(I)$. *It follows that if* $\mathcal{M}$ *has a full satisfaction class that is* I*-disjunctively correct, then* $\mathcal{M}$ *also has a satisfaction class that is* J*-disjunctively correct* (*and the same for deductive*

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

correctness). *Since the theory* $T(I, S)$ *that says that* I *is a cut and* S *is an* I*-disjunctively correct full satisfaction class is consistent with* $\mathrm{Th}(\mathcal{M})$, *by chronic resplendence* $\mathcal{M}$ *has a cut* I *and a full satisfaction class* S *such that* $(\mathcal{M}, I, S)$ *is a resplendent model of* $T(I, S)$. *In particular* I *is above the Skolem closure of* 0 *and thus the aforementioned observation applies.*

Acknowledgments. I have benefitted from priceless feedback concerning the work reported here from Lawrence Wong, Bartosz Wcisło, Albert Visser, Jim Schmerl, Mateusz Łełyk, Roman Kossak, Cezary Cieśliński, and Athar Abdul-Quader (in reverse alphabetical order). Also many thanks to the anonymous referee whose detailed comments were instrumental in reshaping the preliminary draft of this paper into its current format.

Funding. The research presented in this paper was supported by the National Science Centre, Poland (NCN), grant number 2019/34/A/HS1/00399.

BIBLIOGRAPHY

[1] Abdul-Quader, A., & Łełyk, M. (2024). Pathologies in satisfaction classes. *Annals of Pure and Applied Logic*, **175**(2), Article no. 103387, 32 pp.

[2] Buss, S. (1998). *An introduction to proof theory*. In *Handbook of Proof Theory*. S. R. Buss (ed), Amsterdam: North-Holland, pp. 1–78.

[3] Cieśliński, C. (2010). Deflationary truth and pathologies. *The Journal of Philosophical Logic*, **39**, 325–337.

[4] ———. (2017). *The Epistemic Lightness of Truth: Deflationism and its Logic*. Cambridge: Cambridge University Press.

[5] ———. (2021). Interpreting the compositional truth predicate in models of arithmetic. *Archive for Mathematical Logic*, **60**, 749–770.

[6] ———. (2021). On some problems with truth and satisfaction. In Trepczyński, M., editor. *Philosophical Approaches to the Foundations of Logic and Mathematics*. Leiden: Brill, pp. 175–192.

[7] Cieśliński, C., Łełyk, M., & Wcisło, B. (2023). The two halves of disjunctive correctness. ***Journal of Mathematical Logic***, **23**(2), Article no. 2250026, 28 pp.

[8] Enayat, A. (2023). *Curious satisfaction classes*. Preprint, arXiv:2308.07463 [math.LO].

[9] Enayat, A., Łełyk, M., & Wcisło, B. (2020). Truth and feasible reducibility. *Journal of Symbolic Logic*, **85**, 367–421.

[10] Enayat, A., & Pakhomov, F. (2019). Truth, disjunction, and induction. *Archive for Mathematical Logic*, **58**, 753–766.

[11] Enayat, A., & Visser, A. (2012). Full satisfaction classes in a general setting, privately circulated manuscript.

[12] ———. (2015). New constructions of full satisfaction classes. In Achourioti, T., Galinon, H., Fujimoto, K., & Martínez-Fernández, J., editors. *Unifying the Philosophy of Truth*. New York: Springer, pp. 321–325.

[13] Engström, F. (2002). Satisfaction classes in nonstandard models of first-order arithmetic. Preprint, arXiv:math/0209408 [math.LO].

[14] Fujimoto, K. (2022). The function of truth and the conservativeness argument. *Mind*, **131**, 129–157.

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press

[15] Hájek, P., & Pudlák, P. (1993). *Metamathematics of First-Order Arithmetic*. Berlin: Springer.

[16] Halbach, V. (2011). *Axiomatic Theories of Truth*. Cambridge: Cambridge University Press.

[17] Jockusch, C. G., & Soare, R. I. (1972). Π^0_1 classes and degrees of theories. ***Transactions of the American Mathematical Society***, **173**, 33–56.

[18] Kaye, R. (1991). *Models of Peano Arithmetic*. Oxford Logic Guides. Oxford: Oxford University Press.

[19] Kossak, R., & Kotlarski, H. (1988). Results on automorphisms of recursively saturated models of PA. *Fundamenta Mathematicae*, **129**, 9–15.

[20] Kossak, R., & Schmerl, J. (2006). *The Structure of Models of Arithmetic*. Oxford Logic Guides. Oxford: Oxford University Press.

[21] Kossak, R., & Wcisło, B. (2021). Disjunctions with stopping conditions. *Bulletin of Symbolic Logic*, **27**, 231–253.

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

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

[24] Krajewski, S. (1976). Nonstandard satisfaction classes. In Marek, W., et al., editors. *Set Theory and Hierarchy Theory: A Memorial Tribute to Andrzej Mostowski*, 537. Berlin: Springer, pp. 121–144.

[25] Leigh, G. (2015). Conservativity for theories of compositional truth via cut elimination. *Journal of Symbolic Logic*, **80**, 845–865.

[26] Łełyk, M. (2023). Model theory and proof theory of the global reflection principle. *Journal of Symbolic Logic*, **88**, 738–779.

[27] Łełyk, M., & Wcisło, B. (2021). Local collection and end-extensions of models of compositional truth. *Annals of Pure and Applied Logic*, **172**, Article no. 102941, 22 pp.

[28] Robinson, A. (1963). On languages which are based on non-standard arithmetic. *Nagoya Mathematical Journal*, **22**, 83–117.

[29] Smith, S. (1987). Nonstandard characterizations of recursive saturation and resplendency. *Journal of Symbolic Logic*, **52**, 842–863.

[30] ———. (1989). Nonstandard definability. *Annals of Pure and Applied Logic*, **42**, 21–43.

[31] Wcisło, B. (2024). Truth and collection. Preprint, arXiv:2403.19367 [math.LO].

[32] Wcisło, B., & Łełyk, M. (2017). Notes on bounded induction for the compositional truth predicate. *The Review of Symbolic Logic*, **10**, 455–480.

DEPARTMENT OF PHILOSOPHY, LINGUISTICS, AND THEORY OF SCIENCE
UNIVERSITY OF GOTHENBURG
GOTHENBURG, SWEDEN
E-mail: ali.enayat@gu.se

https://doi.org/10.1017/S1755020324000182 Published online by Cambridge University Press