Archive for Mathematical Logic
https://doi.org/10.1007/s00153-024-00915-8

Mathematical Logic

# Indiscernibles and satisfaction classes in arithmetic

**Ali Enayat**[1]

Received: 18 December 2022 / Accepted: 22 February 2024


**Abstract**

We investigate the theory Peano Arithmetic with Indiscernibles (PAI). Models of PAI are of the form $(\mathcal{M}, I)$, where $\mathcal{M}$ is a model of PA, $I$ is an unbounded set of order indiscernibles over $\mathcal{M}$, and $(\mathcal{M}, I)$ satisfies the extended induction scheme for formulae mentioning $I$. Our main results are Theorems A and B following. **Theorem A.** *Let* $\mathcal{M}$ *be a nonstandard model of* PA *of any cardinality.* $\mathcal{M}$ *has an expansion to a model of* PAI *iff* $\mathcal{M}$ *has an inductive partial satisfaction class.* Theorem A yields the following corollary, which provides a new characterization of countable recursively saturated models of PA: **Corollary.** *A countable model* $\mathcal{M}$ of PA *is recursively saturated iff* $\mathcal{M}$ *has an expansion to a model of* PAI. **Theorem B.** *There is a sentence* $\alpha$ *in the language obtained by adding a unary predicate* $I(x)$ *to the language of arithmetic such that given any nonstandard model* $\mathcal{M}$ *of* PA *of any cardinality*, $\mathcal{M}$ *has an expansion to a model of* $\mathrm{PAI} + \alpha$ *iff* $\mathcal{M}$ *has a inductive full satisfaction class.*

**Keywords** Peano arithmetic · Indiscernibles · Satisfaction classes.

**Mathematics Subject Classification** Primary 03F30 · 03F25; Secondary 03C62.

## 1 Introduction

We investigate an extension of PA (Peano Arithmetic), denoted PAI, which is equipped with a designated unbounded class of indiscernibles (see Sect. 3 for the precise definition). The motivation to study PAI arose from the study [5] of the set-theoretic counterpart $\mathrm{ZFI}_<$ of PAI, where it is shown that there is an intimate relationship between $\mathrm{ZFI}_<$ and large cardinals, thus indicating that the set-theoretical consequences of $\mathrm{ZFI}_<$ go well beyond ZFC.

✉ Ali Enayat
ali.enayat@gu.se

1 Department of Philosophy, Linguistics, and the Theory of Science University of Gothenburg, Gothenburg, Sweden

In light of the results obtained in [5] it is natural to investigate PAI since it is well-known [10] that PA is bi-interpretable with the theory $\mathrm{ZF}^{-\infty}+\mathrm{TC}$, where $\mathrm{ZF}^{-\infty}$ is the system of set theory obtained from ZF by replacing the axiom of infinity by its negation, and TC is the sentence asserting that every set is contained in a transitive set (which in the presence of the other axioms implies that the transitive closure of every set exists). The aforementioned proof of the bi-interpretability of PA and $\mathrm{ZF}^{-\infty}+\mathrm{TC}$ can be readily extended to show the bi-interpretability of PAI and $\mathrm{ZFI}^{-\infty}+\mathrm{TC}$; here $\mathrm{ZFI}^{-\infty}$ is the result of augmenting the theory $\mathrm{ZF}^{-\infty}$ with a scheme that stipulates that $I$ is an unbounded subset of ordinals whose elements form a class of indiscernibles over the ambient set-theoretic universe.

Our main results are Theorems A and B of the abstract that relate PAI to the well-studied notions of (a) inductive partial satisfaction classes and (b) inductive full satisfaction classes, which are intimately connected (respectively) with the axiomatic theories of truth known as UTB and CT (see Section 2.2). After presenting preliminaries in Sect. 2, we present the basic features of PAI in Sect. 3, and use them in Sect. 4 to establish (refinements of) Theorems A and B of the abstract. In Sect. 5 we examine PAI through the lens of interpretability, and in Sect. 6 we probe the model-theoretic behavior of fragments of PAI. Finally, in Sect. 7 we present a list of open problems that are motivated by the results in the preceding sections.

## 2 Preliminaries

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

### 2.1 Theories and models

**2.1.1 Definition** The language of arithmetic, $\mathcal{L}_{\mathrm{A}}$, 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 (a) through (g) below, $\mathcal{L} \supseteq \mathcal{L}_{\mathrm{A}}$ and $\mathcal{M}$ and $\mathcal{N}$ are $\mathcal{L}$ -structures.

(a) $\Sigma_0(\mathcal{L}) = \Pi_0(\mathcal{L}) = \Delta_0(\mathcal{L})$ = the collection of $\mathcal{L}$-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 $\mathcal{L}$-term not involving $x$. More generally, $\Sigma_{n+1}(\mathcal{L})$ consists of formulae of the form $\exists x_0 \cdots \exists x_{k-1}\, \varphi$, where $\varphi \in \Pi_n(\mathcal{L})$; and $\Pi_{n+1}(\mathcal{L})$ consists of formulae of the form $\forall x_0 \cdots \forall x_{k-1}\, \varphi$, where $\varphi \in \Sigma_n$ (with the convention that $k = 0$ corresponds to an empty block of quantifiers). *We shall omit the reference to $\mathcal{L}$ if $\mathcal{L} = \mathcal{L}_{\mathrm{A}}$, e.g.,* $\Sigma_n := \Sigma_n(\mathcal{L}_{\mathrm{A}})$. Also, we shall write $\Sigma_n(X)$ instead of $\Sigma_n(\mathcal{L}_{\mathrm{A}} \cup \{X\})$, where $X$ is a new predicate symbol. We often conflate formal symbols with their denotations (if there is no risk of confusion).

(b) PA (Peano Arithmetic) is the result of adding the scheme of induction for all $\mathcal{L}_{\mathrm{A}}$-formulae to the finitely axiomatizable theory known as (Robinson's) Q. PA($\mathcal{L}$) is the theory obtained by augmenting PA with the scheme of induction for all

$\mathcal{L}$-formulae. $\mathrm{I}\Sigma_n(\mathcal{L})$ is the fragment of $\mathrm{PA}(\mathcal{L})$ with the induction scheme limited to $\Sigma_n(\mathcal{L})$-formulae. Given a new predicate $X$, we write $\mathrm{PA}(X)$ and $\mathrm{I}\Sigma_n(X)$ (respectively) instead of $\mathrm{PA}(\mathcal{L}_\mathrm{A}\cup\{X\})$ and $\mathrm{I}\Sigma_n(\mathcal{L}_\mathrm{A}\cup\{X\})$.

(c) If $\varphi(x)$ is an $\mathcal{L}$-formula, $\varphi^{\mathcal{M}} := \{m \in M : \mathcal{M} \models \varphi(x)\}$. For $X \subseteq M$, then we say that $X$ is $\mathcal{M}$-*definable* if $X$ is first order definable (parameters allowed) in $\mathcal{M}$.

(d) A subset $X$ of $M$ is $\mathcal{M}$-*finite* ( *or* $\mathcal{M}$-*coded*) if $X = c_E$ for some $c \in M$, where $c_E = \{m \in M : \mathcal{M} \models m \in_{\mathrm{Ack}} c\}$, and $m \in_{\mathrm{Ack}} c$ is shorthand for the formula expressing " the $m$-th bit of the binary expansion of $c$ is 1".

(e) A subset $X$ of $M$ is said to be *piecewise-coded* (in $\mathcal{M}$) if $\{x \in M : x < m$ and $x \in X\}$ is $\mathcal{M}$-finite for each $m \in M$.

(f) $\mathcal{M}$ is *rather classless* if any piecewise-coded subset of $M$ is already $\mathcal{M}$-definable (By a theorem of Kaufmann, every extension of PA has a recursively saturated rather classless model [13, Theorem 10.1.5]).

(g) We identify the longest well-founded initial submodel of models of PA with the ordinal $\omega$. The following result was established by Kossak [11, Proposition 3.2] for models $\mathcal{M}$ of PA; the generalization to models of $\mathrm{I}\Delta_0 + \mathrm{Exp}$ appears in [6, Lemma 4.2] (note that Exp is the axiom stating the totality of the exponential function).

**2.1.1 Theorem** *Let* $\mathcal{M} \models \mathrm{I}\Delta_0 + \mathrm{Exp}$, *and* $X \subseteq M$. *The following are equivalent:*

($i$) $(\mathcal{M}, X) \models \mathrm{I}\Delta_0(X)$.

($ii$) *X is piecewise-coded in* $\mathcal{M}$.

### 2.2 Satisfaction classes, truth theories, and recursive saturation

**2.2.1 Definition** Suppose $\mathcal{M} \models \mathrm{PA}$, and $S \subseteq M$.

(a) $S$ is said to be *inductive*, if $(\mathcal{M}, S) \models \mathrm{PA}(S)$.

(b) $S$ is said to be a *partial satisfaction class* if $S$ satisfies Tarski's recursive conditions for a satisfaction predicate for all standard formulae. Thus a typical member of $S$ is of the form $\langle \varphi, \overline{a}\rangle$ , where $\varphi \in \mathrm{Form}_m^{\mathcal{M}}$ = the set of $\mathcal{L}_\mathrm{A}$ -formulae in $\mathcal{M}$ with $m$ free variables, where $m \in M$ (note that $\varphi$ need not be standard) and $\overline{a} \in M$ is an $m$ -tuple in the sense of $\mathcal{M}$.

(c) $S$ is said to be a *full satisfaction class* if $S$ satisfies Tarski's recursive conditions for a satisfaction predicate for all formulae in $\mathcal{M}$.

- For better readability we will often write $\langle \varphi, \overline{a}\rangle \in S$ or $\varphi(\overline{a}) \in S$ instead of the more official $S(\langle \varphi, \overline{a}\rangle)$. Also, if $\varphi$ is a sentence (i.e., has no free variables), we will write $\varphi \in S$ instead of $\langle \varphi, \varnothing\rangle \in S$ (where $\varnothing$ is the empty tuple).

The theories UTB (Uniform Tarski Biconditionals) and CT (Compositional Truth) described below are well studied in the literature of axiomatic theories of truth (see, e.g., the monographs by Cieśliński [2] and Halbach [8]).

- Note that for the purposes of this paper, satisfaction and truth are interchangeable, but in general there are subtle differences between the two, see [3].

**2.2.2 Definition** In what follows $T(x)$ is a new unary predicate, $c$ is a new constant symbol, $\mathrm{Form}_1$ is the set of (Gödel-numbers of) $\mathcal{L}_\mathrm{A}$-formulae with exactly one free variable, and $\dot{x}$ is the arithmetically definable function that outputs the numeral for $x$ given the input $x$.

**(a)** UTB is $\mathrm{PA}(T) + \{\forall x(\varphi(x) \leftrightarrow T(\ulcorner \varphi(\dot{x}) \urcorner) : \varphi(x) \in \mathrm{Form}_1\}$.
**(b)** $\mathrm{UTB}(c) = \mathrm{UTB} + \{c > n : n \in \omega\}$.
**(c)** $\mathrm{CT} = \mathrm{PA}(T) + \theta$, where $\theta$ is a single sentence that stipulates that $T$ satisfies Tarski's inductive clauses for a truth predicate for arithmetical sentences.
**(d)** $\mathrm{CT}(c) = \mathrm{CT} + \{c > n : n \in \omega\}$.

The following proposition is well-known and easy to prove; the nontrivial direction of part (a) is the right-to-left part, which employs a routine overspill argument; part (b) follows easily from part (a) and the definitions involved. The proofs of (c) and (d) are routine but somewhat laborious.

**2.2.3 Proposition** *The following holds for every model $\mathcal{M}$ of* PA *of any cardinality.*

*(a) $\mathcal{M}$ has an inductive partial satisfaction class iff $\mathcal{M}$ has an expansion to a model of* UTB.
*(b) $\mathcal{M}$ is nonstandard and has an inductive partial satisfaction class iff $\mathcal{M}$ has an expansion to a model of* UTB($c$).
*(c) $\mathcal{M}$ has an inductive full satisfaction class iff $\mathcal{M}$ has an expansion to a model of* CT.
*(d) $\mathcal{M}$ is nonstandard and has an inductive full satisfaction class iff $\mathcal{M}$ has an expansion to a model of* CT($c$).

The concepts of recursive saturation and satisfaction classes are intimately tied, as witnessed by the following classical result of Barwise and Schlipf whose proof invokes the resplendence property of countable recursively saturated models (for a proof, see Corollary 15.12 of [9]). Note that the right-to-left implication in the Barwise-Schlipf theorem holds for uncountable models $\mathcal{M}$ as well (and is proved by a simple overspill argument). However, by Tarski's undefinability of truth theorem the left-to-right direction fails for 'Kaufmann models' (i.e., recursively saturated rather classless models).

**2.2.4 Theorem** *(Barwise-Schlipf) A countable model $\mathcal{M}$ of* PA *is recursively saturated iff $\mathcal{M}$ has an inductive partial satisfaction class.*

## 2.3 Indiscernibles

**2.3.1 Definition** Given a linear order $(X, <)$, and nonzero $n \in \omega$, we use $[X]^n$ to denote the set of all *increasing* sequences $x_1 < \cdots < x_n$ from $X$.

**2.3.2 Definition** Given a structure $\mathcal{M}$, some linear order $(I, <)$ where $I \subseteq M$, we say that $(I, <)$ is *a set of order indiscernibles in* $\mathcal{M}$ if for any $\mathcal{L}(\mathcal{M})$-formula $\varphi(x_1, \ldots, x_n)$, and any two $n$-tuples $\overline{i}$ and $\overline{j}$ from $[I]^n$, we have:

$$\mathcal{M} \models \varphi(i_1, \ldots, i_n) \leftrightarrow \varphi(j_1, \ldots, j_n).$$

**2.3.3 Definition** Suppose $\mathcal{M}$ has parameter-free definable Skolem functions, and $(I, <_I)$ is a set of order indiscernibles in $\mathcal{M}$, and $I_0 \subseteq I$. We use the notation $\mathcal{M}_{I_0}$ *to denote the elementary submodel of* $\mathcal{M}$ *generated by* $I_0$ (via the parameter-free definable functions of $\mathcal{M}$).

- Note that the universe $M_{I_0}$ of $\mathcal{M}_{I_0}$ consists of the elements of $M$ that are pointwise definable in $(M, i)_{i \in I_0}$.

**2.3.4 Definition** A 1-type $p(x)$ in the language of arithmetic is said to be an *unbounded indiscernible type* if $p(x)$ is a nonprincipal type satisfying: (1) there is no constant Skolem term $c$ such that $x \leq c$ is in $p(x)$, and (2) for any model $\mathcal{M}$ of PA, if $I \subseteq M$ is a set of realizations of $p(x)$, then $I$ is a set of indiscernibles in $\mathcal{M}$.

**2.3.5 Theorem** (Kossak-Schmerl [13, Theorems 3.1.4 and 3.2.10]) *A type $p(x)$ is minimal (in the sense of Gaifman) iff $p(x)$ is an unbounded indiscernible type.*

***2.3.6 Remark*** As indicated in the remark following [13, Theorems 3.1.2], a minimal type $p(x)$ can be arranged to be recursive, and therefore in light of Theorem 2.3.5 every recursively saturated model of PA realizes an unbounded indiscernible type.

## 2.4 Interpretability

**2.4.2 Definition** Suppose $U$ and $V$ are first order theories, and for the sake of notational simplicity, let us assume that $U$ and $V$ are theories that *support a definable pairing function*. We use $\mathcal{L}_U$ and $\mathcal{L}_V$ to respectively designate the languages of $U$ and $V$.

(a) An interpretation $\mathcal{I}$ of $U$ in $V$, written:

$$U \trianglelefteq_{\mathcal{I}} V,$$

is given by a translation $\tau$ of each $\mathcal{L}_U$ -formula $\varphi$ into an $\mathcal{L}_V$-formula $\varphi^\tau$ with the requirement that $V \vdash \varphi^\tau$ for each $\varphi \in U$, where $\tau$ is determined by an $\mathcal{L}_V$-formula $\delta(x)$ (referred to as a *domain formula*), and a mapping $P \mapsto_\tau A_P$ that translates each $n$-ary $\mathcal{L}_U$-predicate $P$ into some $n$-ary $\mathcal{L}_V$-formula $A_P$. The translation is then lifted to the full first order language in the obvious way by making it commute with propositional connectives, and subject to the following clauses:

$$(\forall x \varphi)^\tau = \forall x(\delta(x) \to \varphi^\tau) \text{and } (\exists x \varphi)^\tau = \exists x(\delta(x) \wedge \varphi^\tau).$$

Note that each interpretation $U \trianglelefteq_{\mathcal{I}} V$ gives rise to an *inner model construction* that *uniformly* builds a model $\mathcal{I}(\mathcal{M}) \models U$ for any $\mathcal{M} \models V$.

(b) $U$ is *interpretable* in $V$ (equivalently: $V$ *interprets* $U$), written $U \trianglelefteq V$, iff $U \trianglelefteq_{\mathcal{I}} V$ for some interpretation $\mathcal{I}$.

(c) Given arithmetical theories $U$ and $V$, $U$ is *$\omega$-interpretable in $V$ if* the interpretation of 'numbers' and the arithmetical operations of the interpreted theory $U$ are the same as those of the interpreting theory $V$.

(d) $U$ is *locally interpretable* in $V$, written $U \trianglelefteq_{\mathrm{loc}} V$, if $U_0 \trianglelefteq V$ for every finite subtheory $U_0$ of $U$.
(e) $U$ and $V$ are *mutually interpretable* when $U \trianglelefteq V$ and $V \trianglelefteq U$.
(f) $U$ is a *retract* of $V$ iff there are interpretations $\mathcal{I}$ and $\mathcal{J}$ with $U \trianglelefteq_{\mathcal{I}} V$ and $V \trianglelefteq_{\mathcal{J}} U$, and a binary $U$ -formula $F$ such that $F$ is, $U$-verifiably, an isomorphism between $\mathrm{id}_U$ (the identity interpretation on $U$) and $\mathcal{J} \circ \mathcal{I}$. In model-theoretic terms, this translates to the requirement that the following holds for every $\mathcal{M} \models U$:

$$F^{\mathcal{M}} : \mathcal{M} \xrightarrow{\cong} \mathcal{M}^* := \mathcal{I}(\mathcal{J}(\mathcal{M})).$$

**(g)** $U$ and $V$ are *bi-interpretable* iff there are interpretations $\mathcal{I}$ and $\mathcal{J}$ as above that witness that $U$ is a retract of $V$, and additionally, there is a $V$-formula $G$, such that $G$ is, $V$-verifiably, an isomorphism between the ambient model of $V$ and the model of $V$ given by $\mathcal{I} \circ \mathcal{J}$. In particular, if $U$ and $V$ are bi-interpretable, then given $\mathcal{M} \models U$ and $\mathcal{N} \models V$, we have

$$F^{\mathcal{M}} : \mathcal{M} \xrightarrow{\cong} \mathcal{M}^* := \mathcal{I}(\mathcal{J}(\mathcal{M})) \text{ and } G^{\mathcal{N}} : \mathcal{N} \xrightarrow{\cong} \mathcal{N}^* := \mathcal{J}(\mathcal{I}(\mathcal{N})).$$

**(h)** The above notions can also be localized at a pair of models; in particular suppose $\mathcal{N}$ is an $\mathcal{L}_U$ -structure and $\mathcal{M}$ is an $\mathcal{L}_V$-structure. For example, we say that $\mathcal{N}$ is *interpretable* in $\mathcal{M}$, written $\mathcal{N} \trianglelefteq \mathcal{M}$ (equivalently: $\mathcal{M} \trianglerighteq \mathcal{N}$) iff the universe of discourse of $\mathcal{N}$, as well as all the $\mathcal{N}$-interpretations of $\mathcal{L}_U$ -predicates are $\mathcal{M}$-definable. Similarly, we say that $\mathcal{M}$ and $\mathcal{N}$ are *bi-interpretable* if there are parametric interpretations $\mathcal{I}$ and $\mathcal{J}$ , together with an $\mathcal{M}$-definable $F$ and an $\mathcal{N}$ -definable map $G$ such that:

$$F^{\mathcal{M}} : \mathcal{M} \xrightarrow{\cong} \mathcal{M}^* := \mathcal{I}(\mathcal{J}(\mathcal{M})) \text{ and } G^{\mathcal{N}} : \mathcal{N} \xrightarrow{\cong} \mathcal{N}^* := \mathcal{J}(\mathcal{I}(\mathcal{M})).$$

- Recall that a theory $U$ (with sufficient coding apparatus) is *reflexive* if the formal consistency of each finite fragment of $U$ is provable in $U$.

The following results are classical. See Theorem 2.35 of [7] for an exposition of Mostowski's Reflection Theorem, and Theorem 5 of Chapter VI of [18] for an exposition of Orey's Compactness Theorem.

**2.4.2 Theorem** (Mostowski's Reflection Theorem) *For all $\mathcal{L} \supseteq \mathcal{L}_{\mathrm{A}}$, every extension (in the same language) of PA $(\mathcal{L})$ is reflexive.*

**2.4.3 Theorem** (Orey's Compactness Theorem) *If $U$ is reflexive, and $V \trianglelefteq_{\mathrm{loc}} U$ for some recursively enumerable theory $V$, then $V \trianglelefteq U$.*

## 3 The basics of PAI

**3.1 Definition** PAI is the theory formulated in $\mathcal{L}_{\mathrm{A}}(I)$ whose axioms are (1) through (3) below. Note that we write $x \in I$ instead of $I(x)$ for better readability.

(1) $\mathrm{PA}(I)$.
(2) The sentence $\mathrm{Ubd}(I)$ that expresses: " $I$ is unbounded".
(3) The scheme $\mathrm{Indis}_{\mathcal{L}_{\mathrm{A}}}(I) = \{\mathrm{Indis}_{\varphi}(I) : \varphi$ is an $\mathcal{L}_{\mathrm{A}}$-formula$\}$ stipulating that $I$ forms a class of order indiscernibles for the ambient model of arithmetic. More explicitly, for each $n$ -ary formula $\varphi(v_1, \ldots, v_n)$ in the language $\mathcal{L}_{\mathrm{A}}$, $\mathrm{Indis}_{\varphi}(I)$ is the following sentence:

$$\begin{aligned} &\forall x_1 \in I \cdots \forall x_n \in I \\ &\quad \forall y_1 \in I \cdots \forall y_n \in I \\ &\quad [(x_1 < \cdots < x_n) \wedge (y_1 < \cdots < y_n) \rightarrow (\varphi(x_1, \cdots, x_n) \leftrightarrow \varphi(y_1, \cdots, y_n))]. \end{aligned}$$

$\mathrm{PAI}^{\circ}$ is the weakening of PAI in which the scheme $\mathrm{Indis}_{\mathcal{L}_{\mathrm{A}}}(I)$ is weakened to the scheme $\mathrm{Indis}^{\circ}_{\mathcal{L}_{\mathrm{A}}}(I) = \{\mathrm{Indis}^{\circ}_{\varphi}(I) : \varphi$ is an $\mathcal{L}_{\mathrm{A}}$-formula$\}$, where $\mathrm{Indis}^{\circ}_{\varphi}(I)$ is the following sentence:

$$\begin{aligned} &\forall x_1 \in I \cdots \forall x_n \in I \quad \forall y_1 \in I \cdots \forall y_n \in I \\ &\quad [(x_1 < \cdots < x_n) \wedge (y_1 < \cdots < y_n) \wedge (\ulcorner \varphi \urcorner < x_1 \wedge \ulcorner \varphi \urcorner < y_1) \\ &\quad\quad \rightarrow (\varphi(x_1, \cdots, x_n) \leftrightarrow \varphi(y_1, \cdots, y_n))]. \end{aligned}$$

**3.2 Proposition** *Let* $\mathbb{N}$ *be the standard model of* PA.

*(a)* $\mathbb{N}$ *does not have an expansion to a model of* PAI *(equivalently: Every model of* PAI *is nonstandard).*
*(b)* $\mathbb{N}$ *has an expansion to* $\mathrm{PAI}^{\circ}$.
*(c) If* $(\mathcal{M}, I)$ *is a nonstandard model of* $\mathrm{PAI}^{\circ}$, *and c is any nonstandard element of* $\mathcal{M}$, *then* $(\mathcal{M}, I^{>c}) \models \mathrm{PAI}$, *where* $I^{>c} = \{i \in I : i > c\}$.

***Proof*** (a) is an immediate consequence of the fact that the standard model of PA is pointwise definable, and therefore it does not even have a distinct pair of indiscernibles.

To see that (b) holds, fix some enumeration $\langle \varphi_n : n \in \omega \rangle$ of all arithmetical formulae, and use Ramsey's theorem to construct a sequence $\langle H_n : n \in \omega \rangle$ of subsets of $\omega$ such that for each $n \in \omega$ the following three conditions hold:

(1) $H_n$ is infinite.
(2) $H_n \supseteq H_{n+1}$.
(3) $H_n$ is $\varphi_n$-indiscernible (i.e., $\mathrm{Indis}_{\varphi_n}(H_n)$ holds).

Then recursively define $\langle i_n : n \in \omega \rangle$ by: $i_0 = \max\{\min\{H_0\}, \ulcorner \varphi_0 \urcorner\}$, and $i_{n+1}$ is the least $i \in H_{n+1}$ that is greater than both $i_n$ and $\ulcorner \varphi_n \urcorner$. It is easy to see that $(\mathcal{M}, I) \models \mathrm{PAI}^{\circ}$, where $I = \langle i_n : n \in \omega \rangle$.[1].

Since (c) readily follows from the definitions involved, the proof is complete. □

- In light of part (c) of Theorem 3.2, most results in this paper about PAI have a minor variant in which PAI is replaced by $\mathrm{PAI}^{\circ}$.

[1] The construction of $I$ uses similar ideas as the construction of unbounded indiscernible types in [13, Theorem 3.1.2].

**3.3 Theorem** *Each finite subtheory of* PAI *has an* $\omega$*-interpretation in* PA. *Consequently:*

*(a)* PAI *is a conservative extension of* PA.
*(b) PAI is interpretable in PA, hence PA and PAI are mutually interpretable.*
*(c) PAI is interpretable in* $ACA_0$ *( but not vice-versa).*[2]

***Proof*** The $\omega$-interpretability of any finite subtheory of PAI in PA is an immediate consequence of the well-known schematic provability of Ramsey's theorem $\omega \to (\omega)^n_2$ in PA for all metatheoretic $n \geq 2$ [7, Theorem 1.5, Chapter II]. This makes it evident that (a) holds, and together with Orey's Compactness Theorem 2.4.3, yields (b).[3] Finally, (c) follows from (b) since PA is trivially interpretable in $\mathrm{ACA}_0$. The parenthetical clause of (c) is an immediate consequence of (b) and the classical fact that PA is not interpretable in $\mathrm{ACA}_0$ (the ingredients of whose proof are Mostowski's reflection theorem for PA, finite axiomatizability of $\mathrm{ACA}_0$, and Gödel's second incompleteness theorem). □

***3.4 Remark*** As pointed out by the referee, part (a) of Theorem 3.3 can also be established by taking advantage of some classical facts about unbounded indiscernible types (reviewed in Subsection 2.3) and resplendent models. More specifically, using Remark 2.3.6 and [9, Theorem 15.11], one can readily verify that *there is an unbounded indiscernible type* $p(x)$ *such that every resplendent model* $\mathcal{M}$ *of* PA *has an expansion* $(\mathcal{M}, I) \models \mathrm{PAI}$ *with the property that every element of* $I$ *realizes* $p(x)$.

- In what follows $\mathrm{Form}_k$ is the set of $\mathcal{L}_\mathrm{A}$-formulae with precisely $k$ free variables.

**3.5 Theorem** *The following schemes are provable in* PAI*:*

*(a) The **apartness** scheme:*

$$\{\mathrm{Apart}_\varphi : \varphi \in \mathrm{Form}_{n+1}, n \in \omega\},$$

*where* $\mathrm{Apart}_\varphi$ *is the following formula:*

$$\forall i \in I\ \forall j \in I\, [i < j \to \forall x_1, \ldots, x_n < i\ \ (\exists y\ \varphi(\overline{x}, y) \to \exists y < j\ \varphi(\overline{x}, y))]\,.$$

*(b) The **diagonal indiscernibility**[4] scheme:*

$$\{\mathrm{Indis}^+_\varphi : \varphi \in \mathrm{Form}_{n+1+r},\ n, r \in \omega,\ r \geq 1\},$$

[2] $\mathrm{ACA}_0$ is the well-known finitely axiomatizable subsystem of second arithmetic that is conservative over PA.

[3] Standard techniques can be used to show that the proof of Theorem 3.3(a) yields a feasible reduction of PAI in PA. In other words, there is a polynomial-time function $f$ such that, given the (binary code) of a proof $\pi$ of an arithmetical sentence $\varphi$ in PAI, $f(\pi)$ is the (the binary code of) a proof $f(\pi)$ of $\varphi$ in PA. In particular, PAI has at most polynomial speed-up over PA.

[4] This stronger notion of indiscernibility appears often in expositions of the Paris–Harrington independence result; the same notion is dubbed " strong indiscernibility" in [13, Definition 3.2.8].

*where* $\mathrm{Indis}^+_\varphi(I)$ *is the following formula:*

$$\forall i \in I\, \forall \overline{j} \in [I]^r\, \forall \overline{k} \in [I]^r\ [(i < j_1) \wedge (i < k_1)] \\ \longrightarrow [\forall x_1, \ldots, x_n < i\ (\varphi(\overline{x}, i, j_1, \ldots, j_r) \leftrightarrow \varphi(\overline{x}, i, k_1, \ldots, k_r))].$$

***Proof*** Let $(\mathcal{M}, I) \models \mathrm{PAI}$. To verify that the apartness scheme holds in $(\mathcal{M}, I)$, fix some $i_0 \in I$ and some $\varphi(\overline{x}, y) \in \mathrm{Form}_{n+1}$. Then, since the $I$ is unbounded and the collection scheme holds in $(\mathcal{M}, I)$, and $I$ is unbounded in $\mathcal{M}$, there is some $j_0 \in I$ with $i_0 < j_0$ such that:

$$(\mathcal{M}, I) \models \forall \overline{x} \in [i_0]^n\ (\exists y \varphi(\overline{x}, y) \to \exists y < j_0)\, \varphi(\overline{x}, y)).$$

The above, together with the indiscernibility of $I$ in $\mathcal{M}$, makes it evident that $(\mathcal{M}, I) \models \mathrm{Apart}_\varphi$.

To verify that $\mathrm{Indis}^+_\varphi(I)$ holds in $(\mathcal{M}, I)$, we will first establish a weaker form of diagonal indiscernibility of $I$ in which all $j_n < k_1$ (thus all the elements of $\overline{j}$ are less than all the elements of $\overline{k}$). Fix some $\varphi \in \mathrm{Form}_{n+1+r}$ and $i_0 \in I$. Within $\mathcal{M}$ consider the function $f : [M]^r \to \mathcal{P}([i_0]^n)$ defined by:

$$f(\overline{y}) := \{\overline{a} \in [i_0]^n : \varphi(\overline{a}, i_0, \overline{y})\}.$$

Since $(\mathcal{M}, I)$ satisfies the collection scheme and $I$ is unbounded in $\mathcal{M}$, this shows there are $y_1 < \cdots < y_{2r}$ in $I$ such that:

$$f(y_1, \ldots, y_r) = f(y_{r+1}, \ldots, y_{2r}).$$

Thus $(\mathcal{M}, I)$ satisfies:

$$\forall \overline{x} \in [i_0]^n\ \left[\varphi(\overline{x}, i_0, y_1, \ldots, y_r) \leftrightarrow \varphi(\overline{x}, i_0, y_{r+1}, \ldots, y_{2r})\right].$$

By the indiscernibility of $I$ in $\mathcal{M}$, the above implies the following weaker form of $\mathrm{Indis}^+_\varphi(I)$:

$$\forall i \in I\, \forall \overline{j} \in [I]^r\, \forall \overline{k} \in [I]^r\ [(i < j_1) \wedge (j_n < k_1)] \\ \longrightarrow \left[\forall \overline{x} \in [i_0]^n\ (\varphi(\overline{x}, i, j_1, \cdots, j_r) \leftrightarrow \varphi(\overline{x}, i, k_1, \cdots, k_r))\right].$$

We will now show that the above weaker form of $\mathrm{Indis}^+_\varphi(I)$ already implies $\mathrm{Indis}^+_\varphi(I)$. Given $i \in I$, $\overline{a} \in [I]^r$ and $\overline{b} \in [I]^r$, with $i < a_1$ and $i < b_1$, choose $\overline{y} \in [I]^r$ with $y_1 > \max\{a_n, b_n\}$. Then by the above we have:

$$\mathcal{M} \models \left[\forall \overline{x} \in [i_0]^n\ (\varphi(\overline{x}, i, a_1, \ldots, a_r) \leftrightarrow \varphi(\overline{x}, i, y_1, \ldots, y_r))\right],$$

and

$$\mathcal{M} \models \left[\forall \overline{x} \in [i_0]^n \ \ (\varphi(\overline{x}, i, b_1, \ldots, b_r) \leftrightarrow \varphi(\overline{x}, i, y_1, \ldots, y_r))\right],$$

which together imply:

$$\mathcal{M} \models \left[\forall \overline{x} \in [i_0]^n \ \ (\varphi(\overline{x}, i, a_1, \cdots, a_r) \leftrightarrow \varphi(\overline{x}, i, b_1, \cdots, b_r))\right].$$

□

- Note that the diagonal indiscernibility scheme for $\mathcal{L}_{\mathrm{A}}$-formulae ensures that if $(\mathcal{M}, I) \models \mathrm{PAI}$ and $i \in I$, then $I^{\geq i}$ is a set of indiscernibles over the expanded structure $(\mathcal{M}, m)_{m<i}$, where $I^{\geq i} = \{x \in I : x \geq i\}$.

## 4 Main results

In this section we prove refinements of Theorems A and B of the abstract (as in Theorems 4.6 and 4.12).

**4.1 Theorem** *There is a formula $\sigma(x)$ in the language $\mathcal{L}_{\mathrm{A}}(I)$ such that $S = \sigma^{\mathcal{M}}$ is an inductive partial satisfaction class on $\mathcal{M}$ for all models $(\mathcal{M}, I) \models \mathrm{PAI}$.*

***Proof*** We first define a recursive function that transforms each formula $\varphi(\overline{x}) \in \mathrm{Form}_n$ into a $\Delta_0$-formula $\varphi^*(\overline{x}, z_1, \ldots, z_k)$, where $\{z_i : 1 \leq i \in \omega\}$ is a fresh supply of variables added to the syntax of first order logic. In what follows $x$ and $y$ range over the set of variables before the addition of the fresh stock of $z_i$s. We assume that the only logical constants used in $\varphi$ are $\{\neg, \vee, \exists\}$ and none of the fresh variables $z_i$ occurs in $\varphi$. The definition of $\varphi^*$ below will make it clear that $k$ is the $\exists$-depth of $\varphi$.
(1) If $\varphi$ is an atomic $\mathcal{L}_{\mathrm{A}}$ -formula, then $\varphi^* = \varphi$.
(2) $(\neg\varphi)^* = \neg\varphi^*$.
(3) $(\varphi_1 \vee \varphi_2)^* = \varphi_1^* \vee \varphi_2^*$.
(4) $(\exists y\ \varphi)^* = \exists y < z_1\ \widetilde{\varphi^*}$, where $\varphi^* = \varphi^*(\overline{x}, y, z_1, \ldots, z_k)$, and $\widetilde{\varphi^*}$ is the result of replacing $z_i$ with $z_{i+1}$ in $\varphi^*$ for each $1 \leq i \leq k$.

**Claim** $(\nabla)$. *Suppose $\varphi = \varphi(\overline{x}) \in \mathrm{Form}_n$, and $\varphi^* = \varphi^*(\overline{x}, z_1, \ldots, z_k)$, $(\mathcal{M}, I) \models \mathrm{PAI}$, $\overline{a} \in M^n$, and $(i_1, \ldots, i_k) \in [I]^k$ such that there is some $j \in I$ with $j < i_1$ and $a_s < j$ for each $1 \leq s \leq n$. Then $\mathcal{M}$ satisfies:*

$$\varphi(\overline{a}) \leftrightarrow \varphi^*(\overline{a}, i_1, \ldots, i_k).$$

***Proof*** We use induction of the complexity of $\varphi$. The only case that needs an explanation is the existential case, the others go through trivially. Thus, it suffices to verify that if $(i_1, \ldots, i_{k+1}) \in [I]^k$ and there is some $j \in I$ with $j < i_1$ and $a_s < j$ for each $1 \leq s \leq k$, then:

($\nabla$) $\mathcal{M} \models (\exists y\, \varphi(\overline{a}, y) \leftrightarrow \exists y < i_1\, \varphi^*(\overline{a}, y, i_2, \cdots, i_{k+1}))$,
where $(\varphi(\overline{x}, y))^* = \varphi^*(\overline{x}, y, z_1, \cdots, z_k)$. To establish the left-to-right direction of ($\nabla$), suppose $\mathcal{M} \models \exists y\, \varphi(\overline{a}, y)$. By the veracity of the apartness scheme and the assumption that $a_s < j$ for each $1 \leq s \leq n$, there is $b < i_1$ such that $\mathcal{M} \models \varphi(\overline{a}, b)$. Thus since $b$, as well as $a_1, \ldots, a_n$ are all below $i_1$, $i_1$ can serve as the element "$j$" of the inductive assumption, hence allowing us to conclude that $\mathcal{M} \models \varphi(\overline{a}, b)$ iff $\varphi^*(\overline{a}, b, i_2, \ldots, i_{k+1})$, therefore $\mathcal{M} \models \exists y < i_1\, \varphi^*(\overline{a}, y, i_2, \ldots, i_{k+1})$, as desired. The right-to-left direction of ($\nabla$) is trivial. This concludes the proof of the claim ($\nabla$).

We are now ready to show that there is an $(\mathcal{M}, I)$-definable $S \subseteq M$ such that $S$ is an inductive satisfaction class over $\mathcal{M}$. The following procedure takes place in $(\mathcal{M}, I)$, in particular, the variables $n$ and $k$ in ($\mathbb{P}$) range over $M$ and need not be standard:
($\mathbb{P}$) Given any $\varphi(\overline{x}) \in \mathrm{Form}_n$ and any $n$-tuple $\overline{a}$, calculate $(\varphi(\overline{x}))^* = \varphi^*(\overline{x}, z_1, \ldots, z_k)$, and let $j \in I$ be the first element of $I$ such that $\ulcorner\varphi(\overline{x})\urcorner < j$ and $a_s < j$ for each $1 \leq s \leq n$, and then let and $i_1, \ldots, i_k$ to be the first $k$ elements of $I$ that are above $j$. Then define $S$ by:

$$\langle\varphi, \overline{a}\rangle \in S \text{ iff } [\varphi^*(\overline{a}, i_1, \ldots, i_k) \in \mathrm{Sat}_{\Delta_0}],$$

where $\mathrm{Sat}_{\Delta_0}$ is the canonical $\Sigma_1$-definable satisfaction predicate for $\Delta_0$ formulae of arithmetic.
Thus the desired formula $\sigma$ is given by

$$\sigma(\langle\varphi, \overline{a}\rangle) := [\varphi^*(\overline{a}, i_1, \ldots, i_k) \in \mathrm{Sat}_{\Delta_0}].$$

□

***4.2 Remark*** Three remarks are in order concerning the proof of Theorem 4.1.
(a) If $\varphi(\overline{x})$ is a standard formula, $(\mathcal{M}, I) \models \mathrm{PAI}$, and $j \in I$, then the condition $\ulcorner\varphi(\overline{x})\urcorner < j$ in the procedure ($\mathbb{P}$) is automatically satisfied since every element of $I$ is nonstandard. The role of the condition $\ulcorner\varphi(\overline{x})\urcorner < j$ will become clear in the proof of Lemma 4.10 and Theorem 4.11.
(b) If $I'$ is a cofinal subset of $I$ such that $(\mathcal{M}, I') \models \mathrm{PAI}$ and $S'$ is the partial satisfaction class on $\mathcal{M}$ as defined by $\sigma$ in $(\mathcal{M}, I')$, then thanks to the diagonal indiscernibility property of $I$, $S = S'$. This fact comes handy in the proof of Theorem 5.5.
(c) The transformation $\varphi \mapsto \varphi^*$ given in the proof of Theorem 4.1 can be reformulated in the following more intuitive way: Given $\varphi(\overline{x}) \in \mathrm{Form}_n$, find an equivalent formula $\varphi^{\mathrm{pnf}}(\overline{x})$ in the prenex normal form:

$$\varphi^{\mathrm{pnf}}(\overline{x}) = \forall v_1 \exists w_1 \cdots\ \delta(v_1, w_1 \cdots, v_k, w_k, \overline{x}),$$

and then define $(\varphi(\overline{x}))^*$ to be:

$$\forall v_1 < z_1\, \exists w_1 < z_2 \cdots\ \delta(v_1, w_1, \cdots, v_k, w_k, \overline{x}).$$

A similar transformation is found in the proof of the Paris-Harrington Theorem [19].

**4.3 Corollary** *The following hold for every model $\mathcal{M}$ of* PA *of any cardinality.*

*(a) There is no $\mathcal{M}$- definable subset $I$ of $M$ such that $(\mathcal{M}, I) \models$ PAI ( therefore no rather classless recursively saturated model of* PA *has an expansion to a model of* PAI).

*(b) If $\mathcal{M}$ has an expansion to a model of PAI, then $\mathcal{M}$ is recursively saturated; and the converse holds if $\mathcal{M}$ is countable.*

*(c) If $\mathcal{M}$ has an expansion $(\mathcal{M}, I) \models$ PAI, then $M \neq M_I$, where $M_I$ consists of elements of $M$ that are definable in $(\mathcal{M}, i)_{i \in I}$. Thus $(\mathcal{M}_I, I) \not\models$ PAI, where $\mathcal{M}_I$ is the submodel of $\mathcal{M}$ whose universe is $M_I$.*

***Proof*** (a) follows by putting Theorem 4.1 together with Tarski's theorem on undefinability of truth.[5] Alternatively, as pointed out by the referee, the undefinability of $I$ in $\mathcal{M}$ follows from the unboundedness of $I$ and the apartness scheme. To see this, suppose that $I$ is defined by a formula with a parameter $m$ in $\mathcal{M}$. Together with the unboundedness of $I$, this implies that for any every elementary cut $K$ of $\mathcal{M}$ that contains $m$, $K \cap I$ is cofinal in $K$. Thanks to the definability assumption of $I$, we can choose a pair $i < j$ in $I$ above $m$ such that there is no element of $I$ that is strictly between $i$ and $j$. By the appartness scheme, there is an elementary cut $K$ of $\mathcal{M}$ that contains $i$ (and therefore $m$ as well) such that $K < j$. But the choice of $i$ and $j$ makes it impossible for $K \cap I$ to be cofinal in $K$, contradiction.

(b) follows directly by putting Theorem 4.1 with the Barwise-Schlipf Theorem 2.2.4. As pointed out by the referee, the recursive saturation of $\mathcal{M}$ can also be derived from the unboundedness of $I$ together with $(\mathcal{M}, I) \models$ Apart; this follows from a theorem of Smory ński and Stavi [22]. More specifically, using the apartness terminology, the theorem proved by Smoryński and Stavi says that if $I$ is unbounded, $(\mathcal{M}, I) \models$ Apart, and $I$ is coded in an elementary end extension of $\mathcal{M}$, then $\mathcal{M}$ is recursively saturated.

To verify (c) suppose $M_I = M$ for $(\mathcal{M}, I) \models$ PAI. Recall that $\mathcal{M}$ is nonstandard by Proposition 3.2(a). By Theorem 4.1 there is an inductive partial satisfaction class $S$ on $\mathcal{M}$ that is definable in $(\mathcal{M}, I)$. Consider the function

$$h : M \to M,$$

where $h$ is defined in $(\mathcal{M}, I)$ by $h(m) :=$ the (Gödel number of) the least $\mathcal{L}_A$-formula $\varphi(x, \overline{y})$ such that, as deemed by $S$, $m$ is defined by $\varphi(x, \overline{i})$ for some tuple $\overline{i}$ of parameters from $I$, i.e., $S$ contains the sentences $\varphi(m, \overline{i})$ and $\exists! x \ \varphi(x, \overline{i})$. Note that the set of *standard* elements of $\mathcal{M}$ is definable in $(\mathcal{M}, I)$ as the set of $i$ such that $i < j$ for some $j$ in the range of $h$. Thus $(\mathcal{M}, I)$ is a nonstandard model of PAI, in which the standard cut $\omega$ is definable, which is impossible. This concludes the proof of (c). $\square$

[5] An alternative, more direct proof of (a) invokes diagonal indiscernibility and unboundedness of $I$. Suppose to the contrary that $(\mathcal{M}, I) \models$ PAI and $I$ is definable in $\mathcal{M}$ by a formula $\varphi(x, m)$ for some $m \in M$. Let $i_1 < i_2 < i_3$ be the first three elements of $I$ above $m$. Note that $i_2$ and $i_3$ is each pointwise definable in $(\mathcal{M}, m, I)$. Hence $i_2$ and $i_3$ are discernible in $(\mathcal{M}, m, I)$, and therefore they are also discernible in $(\mathcal{M}, m)$ (since $I$ is definable in $\mathcal{M}$ with parameter $m$ ). On the other hand, by the diagonal indiscernibility property of $I$, for any arithmetical formula $\theta(x, y)$, $\mathcal{M}$ satisfies $\theta(m, i_1) \leftrightarrow \theta(m, i_2)$. We have arrived at a contradiction.

***4.4 Remark*** As shown by Schmerl [20], every countable recursively saturated model $\mathcal{M}$ of PA carries a set of indiscernibles $I$ such that $M_I = M$. Thus, in light of part (c) of Corollary 4.3, such a set of indiscernibles $I$ never has the property that $(\mathcal{M}, I) \models$ PAI. Also, as pointed out by the referee, in part (c) of Corollary 4.3, if all elements of $I$ realize an indiscernible type, then it follows from the properties of indiscernible types that $(\mathcal{M}, M_I)$ is not recursively saturated. Moreover, part (c) of Corollary 4.3 should be contrasted with the fact that in any model $\mathcal{M}$ of PA, the Skolem hull of every unbounded definable set is $\mathcal{M}$ (see [13, Lemma 2.1.10]).

***4.5 Remark*** In contrast to part (c) of Corollary 4.3, the proof technique of the Kossak-Schmerl construction of prime inductive partial satisfaction classes (as in Theorem 10.5.2 of [13]) can be readily adapted to show that every countable recursively saturated model $\mathcal{M}$ of PA has a pointwise definable expansion $(\mathcal{M}, I) \models$ PAI.

**4.6 Theorem** *The following are equivalent for a model $\mathcal{M}$ of* PA *of any cardinality:*
$(i)$ *$\mathcal{M}$ has an expansion to a model of* UTB$(c)$.
$(ii)$ *$\mathcal{M}$ has an expansion to a model of* PAI.
*Consequently, $\mathcal{M}$ has an expansion to a model of* UTB *iff $\mathcal{M}$ has an expansion to a model of* PAI°.

***Proof*** Since $(ii) \Rightarrow (i)$ is justified by Theorem 4.1, it suffices to show that $(i) \Rightarrow (ii)$..[6] By Proposition 2.2.3(b) there is an inductive partial satisfaction class $S$ on $\mathcal{M}$. Consider the $\mathcal{L}_A(S)$-formula $\psi(x)$ that expresses:

"there is a definable (in the sense of $S$) unbounded set of indiscernible for $\mathcal{L}_A$-formulae of Gödel-number at most $x$".

More specifically, $\psi(x)$ is the formula $\exists\theta \in \mathrm{Form}_1\ (U(\theta) \wedge H(\theta, x))$, where $U(\theta)$ is the following $\mathcal{L}_A(S)$-sentence:

$$[\forall x \exists y (x < y \wedge \theta(x))] \in S,$$

and $H(\theta, x)$ is the following $\mathcal{L}_A(S)$ -sentence:

$$\forall \varphi \in \mathrm{Form}(\varphi \leq x \rightarrow \mathrm{Indis}_{\varphi}(\theta) \in S),$$

where $\mathrm{Indis}_{\varphi}(\theta)$ is the following $\mathcal{L}_A$-sentence:

$$\begin{aligned} &\forall x_1 \ldots \forall x_{2n} \\ &\quad [(x_1 < \ldots < x_n) \wedge (x_{n+1} < \cdots < x_{2n}) \wedge \bigwedge_{1 \leq i \leq 2n} \theta(x_i)] \rightarrow \\ &(\varphi(x_1, \ldots, x_n) \leftrightarrow \varphi(x_{n+1}, \ldots, x_{2n}))]. \end{aligned}$$

[6] As pointed out by Roman Kossak, $(i) \Rightarrow (ii)$ was first noted in [12, Proposition 4.5].

By the schematic provability of Ramsey's theorem in PA, $(\mathcal{M}, S) \models \psi(n)$ for each $n \in \omega$, so by overspill, $(\mathcal{M}, S) \models \psi(c)$ holds for some nonstandard $c \in M$. Hence there is some $\theta_0 \in \mathrm{Form}_1^{\mathcal{M}}$ such that $(\mathcal{M}, S) \models H(\theta, c)$, and thus $(\mathcal{M}, I) \models \mathrm{PAI}$, where:

$$I := \{m \in M : (\mathcal{M}, S) \models \theta(m) \in S\}.$$

This concludes the proof of the equivalence of $(i)$ and $(ii)$. The 'consequently' clause readily follows from Proposition 3.2 and the equivalence of $(i)$ and $(ii)$. □

- Going back to Theorem 4.1, one might wonder if it is possible for $\sigma^{\mathcal{M}}$ to be a *full* satisfaction class on $\mathcal{M}$. There are certainly many models $(\mathcal{M}, I)$ of PA for which $\sigma^{\mathcal{M}}$ is not a full satisfaction class since the existence of a full inductive satisfaction class on a model $\mathcal{M}$ implies that Con(PA) holds in $\mathcal{M}$ (and much more, see the remarks following Theorem 4.9). The results of the rest of this section are informed by this question.

**4.7 Definition** $\alpha$ is the $\mathcal{L}_{\mathrm{A}}(I)$-sentence expressing "$\sigma$ defines a full satisfaction class", where $\sigma(x)$ is the formula given in the proof of Theorem 4.1.

**4.8 Definition** Given a recursively axiomatized theory $T$ extending $\mathrm{I}\Delta_0 + \mathrm{Exp}$, the *uniform reflection scheme over* $T$, denoted $\mathrm{RFN}(T)$, is defined via:

$$\mathrm{RFN}(T) := \{\forall x(\mathrm{Prov}_T(\ulcorner \varphi(\dot{x}) \urcorner) \to \varphi(x)) : \varphi(x) \in \mathrm{Form}_1\}.$$

The sequence of schemes $\mathrm{RFN}^{\alpha}(T)$, where $\alpha$ is recursive ordinal $\alpha$, is defined as follows:
$\mathrm{RFN}^0(T) = T$;
$\mathrm{RFN}^{\alpha+1}(T) = \mathrm{RFN}(\mathrm{RFN}^{\alpha+1}(T))$;
$\mathrm{RFN}^{\gamma}(T) = \bigcup_{\alpha<\gamma} \mathrm{RFN}^{\alpha}(T)$.

**4.9 Theorem** *(Folklore) The arithmetical consequences of CT are axiomatized by* $\mathrm{PA} + \mathrm{RFN}^{\varepsilon_0}(\mathrm{PA})$.

***Proof*** It is well-known [8, Section 8.6] that the arithmetical consequences of CT coincide with the arithmetical consequences of ACA (the extension of $\mathrm{ACA}_0$ by the full induction scheme). It has long been known that the arithmetical consequences of ACA can be axiomatized by $\mathrm{PA} + \mathrm{RFN}^{\varepsilon_0}(\mathrm{PA})$, a result which has been recently given a new proof in the work of Beklemishev and Pakhomov [1, Sec. 8.3].[7] □

The following lemma, which will come handy at the end of the proof of Theorem 4.11, shows that if $(\mathcal{M}, I) \models \mathrm{PAI} + \alpha$, and $\varphi \in \mathrm{Form}^{\mathcal{M}}$ (note that $\varphi$ is allowed to be nonstandard), then as viewed by $S$, a tail of $I$ satisfies diagonal $\varphi$-indiscernibility. In what follows $\mathrm{FS}(S)$ is the sentence asserting that S is a full satisfaction class.

[7] It is also known that $\mathrm{PA} + \mathrm{RFN}^{\varepsilon_0}(\mathrm{PA})$ can be axiomatized by $\mathrm{PA} + \mathrm{TI}(\varepsilon_{\varepsilon_0})$, where $\mathrm{TI}(\varepsilon_{\varepsilon_0})$ is the scheme of transfinite induction for ordinals less than $\varepsilon_{\varepsilon_0}$ ($\varepsilon_\alpha$ is the $\alpha$-th $\varepsilon$-number, i.e., the $\alpha$-th fixed point of the map $\gamma \mapsto \omega^{\gamma}$).

**4.10 Lemma** *Suppose* $(\mathcal{M}, S, I) \models \mathrm{PAI} + \mathrm{PA}(S, I) + \mathrm{FS}(S)$. *If* $\varphi \in \mathrm{Form}^{\mathcal{M}}_{n+r+1}$ *(where* $n, r \in M$*), then:*

$$(\mathcal{M}, S, I) \models \forall i \in I \ (\varphi < i \ \longrightarrow \theta(S, i, \varphi)),$$

*where* $\theta(S, i, \varphi)$ *is the following* $\mathcal{L}_{\mathrm{A}}(S, I)$*-formula:*

$$\forall \overline{j} \in [I]^r \ \forall \overline{k} \in [I]^r \ [(i < j_1) \wedge (i < k_1)] \\ \longrightarrow [\forall x_1, \cdots, x_n < i \ (\varphi(\overline{x}, i, j_1, \cdots, j_r) \in S \leftrightarrow \varphi(\overline{x}, i, k_1, \cdots, k_r) \in S)].$$

***Proof*** The strategy of establishing the diagonal indiscernibility of $I$ in the proof of Theorem 3.5(b) can be readily carried out in this context, thanks to the fact that $(\mathcal{M}, S, I) \models \mathrm{PA}(S, I)$. □

**4.11 Theorem** *There is a formula* $\iota(x)$ *in the language* $\mathcal{L}_{\mathrm{A}}(T, c)$ *such that for all models* $(\mathcal{M}, T, c)$ *of* CT($c$), $(\mathcal{M}, I) \models \mathrm{PAI} + \alpha$ *for* $I = \iota^{(\mathcal{M}, T, c)}$.

***Proof*** We will describe the formula $\iota(x)$ by working in an arbitrary model $(\mathcal{M}, T, c) \models \mathrm{CT}(c)$. Since $\mathrm{PA}(S) + \mathrm{FS}(S)$ and CT are well-known to be bi-interpretable, we will do most of our work with the model $(\mathcal{M}, S) \models \mathrm{PA}(S) + \mathrm{FS}(S)$, and at the end will take advantage of the nonstandard element $c$. The basic idea is that PA can verify that the formalized (infinite) Ramsey theorem is provable in PA, so using the inductive full satisfaction class $S$ we can follow the strategy of the proof of part (b) of Theorem 3.2 to define a set $I$ in $(\mathcal{M}, S)$ such that $(\mathcal{M}, I) \models \mathrm{PAI} + \alpha$. More specifically, Ramsey's theorem's can be fine-tuned by asserting that if an arithmetically definable coloring $f$ of $m$-tuples is of complexity $\Sigma_n$, then there is an arithmetical infinite monochromatic subset for $f$ of complexity $\Sigma_{n+m+1}$ [7]. Therefore

$$\mathrm{PA} \vdash \forall r \geq 2 \ \forall \varphi \in \mathrm{Form}_r \ \exists \theta \in \mathrm{Form}_1 \ \mathrm{Prov}_{\mathrm{PA}}(\mathrm{Indisc}_{\varphi}(\theta)),$$

where $\mathrm{Indisc}_{\varphi}(\theta)$ is as in the proof of Theorem 4.6. On the other hand, it is well-known that $\mathrm{PA}(S) + \mathrm{FS}(S)$ proves the global reflection principle:[8]

$$\forall \varphi (\mathrm{Prov}_{\mathrm{PA}}(\varphi) \rightarrow S(\varphi)).$$

Hence
(∗) $(\mathcal{M}, S) \models \forall r \geq 2 \ \forall \varphi \in \mathrm{Form}_r \ \exists \theta \in \mathrm{Form}_1 \ \mathrm{Indisc}_{\varphi}(\theta) \in S$.
Reasoning in $(\mathcal{M}, S)$, fix some enumeration $\langle \varphi_m : m \in M \rangle$ of all arithmetical formulae, and use (∗) to construct a sequence $\langle \theta_m : m \in M \rangle$ of elements of $\mathrm{Form}_1^{\mathcal{M}}$ defined via an internal recursion in $(\mathcal{M}, S)$ such that the following three conditions hold:

(1) $(\mathcal{M}, S) \models \forall m \forall x \ \exists y > x \ \theta_m(\dot{y}) \in S$.
(2) $(\mathcal{M}, S) \models \forall m \forall x (\theta_{m+1}(\dot{x}) \in S \rightarrow \theta_m(\dot{x}) \in S)$.

[8] Indeed, as shown in Łełyk's dissertation [16], the global reflection principle can be proved in the fragment $\mathrm{CT}_0$ of CT.

(3) $(\mathcal{M}, S) \models \forall m\ \mathrm{Indisc}_{\varphi_m}(\theta_m) \in S$.

Let $H_m = \{x \in M : (\mathcal{M}, S) \models \theta_m(\dot{x}) \in S\}$, and within $(\mathcal{M}, S)$ recursively define $\langle i_m : m \in M\rangle$ by: $i_0 = \max\{\min\{H_0\}, \ulcorner\varphi_0\urcorner\}$, and $i_{m+1}$ is the least $i \in H_{m+1}$ that is greater than both $i_m$ and $\ulcorner\varphi_m\urcorner$. It is easy to see that $(\mathcal{M}, I) \models \mathrm{PAI}^\circ$, where $I = \langle i_m : m \in M\rangle$ , and therefore as noted in Proposition 3.2(c) $(\mathcal{M}, I^{>c}) \models \mathrm{PAI}$. The procedure described for constructing $I$ makes it clear that $I$ is definable by an $\mathcal{L}_\mathrm{A}(T, c)$-formula $\iota(x)$.

It remains to show that $(\mathcal{M}, I) \models \mathrm{PAI} + \alpha$. Note that $(\mathcal{M}, S, I) \models \mathrm{PA}(S, I)$, this is precisely where Lemma 4.10 comes to the rescue, since together with the veracity of $\mathrm{PAI} + \mathrm{PA}(S, I) + \mathrm{FS}(S)$ in $(\mathcal{M}, S, I)$ it allows us verify the following nonstandard analogue $(\nabla^*)$ of $(\nabla)$ from the proof of Theorem 4.1 (in what follows the map $\varphi \mapsto \varphi^*$ is defined as in the proof of Theorem 4.1 within $\mathcal{M}$).
$(\nabla^*)$**.** Suppose $\varphi = \varphi(\overline{x}) \in \mathrm{Form}_r^{\mathcal{M}}$ for some $r \in M$ (NB: $r$ need not be standard), $\varphi^* = \varphi^*(\overline{x}, z_1, \ldots, z_{k-1})$, where $k \in M$, $\overline{a} \in M^r$, and $(i_1, \ldots, i_k) \in [I]^k$ such that there is some $j \in I$ *with* $j < i_1$ and $a_s < j$ for each $1 \le s \le r$. Then $(\mathcal{M}, S, I)$ satisfies the following:

$$\varphi(\overline{a}) \in S \leftrightarrow [\varphi^*(\overline{a}, i_1, \ldots, i_k) \in \mathrm{Sat}_{\Delta_0}].$$

Recall that $\sigma(\langle\varphi, \overline{a}\rangle) := [\varphi^*(\overline{a}, i_1, \ldots, i_k) \in \mathrm{Sat}_{\Delta_0}]$. $(\nabla^*)$ assures us that $\sigma^{\mathcal{M}}$ coincides with $S$, and thus $(\mathcal{M}, I) \models \mathrm{PAI} + \alpha$. □

**4.12 Theorem** *The following hold for any model* $\mathcal{M}$ *of* PA *of any cardinality:*

***(a)*** $\mathcal{M}$ *has an expansion to* $\mathrm{CT}(c)$ *iff* $\mathcal{M}$ *has an expansion to* $\mathrm{PAI} + \alpha$.
***(b)*** $\mathcal{M}$ *has an expansion to* CT *iff* $\mathcal{M}$ *has an expansion to* $\mathrm{PAI}^\circ + \alpha$.

***Proof*** We only verify (a) since the argument for (b) is similar. The left-to-right direction of (a) is justified by Theorem 4.11. The other direction is evident thanks to the axiom $\alpha$. □

**4.13 Corollary** *The arithmetical consequences of* $\mathrm{PAI}^\circ + \alpha$ *and* $\mathrm{PAI} + \alpha$ *are axiomatized by* $\mathrm{PA} + \mathrm{RFN}^{\varepsilon_0}(\mathrm{PA})$.

***Proof*** This follows from putting the completeness theorem of first order logic together with Theorems 4.9 and 4.12. □

***4.14 Remark*** In contrast to Theorem 4.6, in Theorem 4.12 $\mathrm{CT}(c)$ cannot be weakened to $\mathrm{CT}_0(c)$, where $\mathrm{CT}_0(c)$ is the fragment of $\mathrm{CT}(c)$ in which the extended induction scheme is limited to $\Delta_0(T)$-formulae. This is because the arithmetical consequences of $\mathrm{CT}_0$ form a tiny fragment of the arithmetical consequences of CT. More explicitly, it has been long known by the cognoscenti that by combining results of Kotlarski [14] and Smoryński [21], the arithmetical consequences of $\mathrm{CT}_0$ can be shown to be axiomatized by $\mathrm{RFN}^\omega(\mathrm{PA})$. The recent work of Łełyk [17] provides a model-theoretic proof of this axiomatizability result, and culminates earlier results obtained in Kotlarski's aforementioned paper, Wcisło and Łełyk's [24], and Łełyk's doctoral dissertation [16].

**4.15 Corollary** *The following are equivalent for every countable model $\mathcal{M}$ of* PA*:*

$(i)$ $\mathcal{M}$ *has an expansion to a model of* $\mathrm{PAI} + \alpha$.
$(ii)$ $\mathcal{M}$ *is recursively saturated and satisfies* $\mathrm{RFN}^{\varepsilon_0}(\mathrm{PA})$.

***Proof*** This is an immediate consequence of Theorems 4.9 and 4.12, together with the resplendence property of countable recursively saturated models. □

***4.16 Remark*** The proof of Corollary 4.15 makes it clear that the countability restriction on $\mathcal{M}$ can be dropped for $(i) \Rightarrow (ii)$. However, the existence of Kaufmann models together with Tarski's undefinability of truth theorem makes it evident that the countability restriction cannot be dropped for $(iii) \Rightarrow (i)$.

## 5 Interpretability analysis of PAI

In this section we examine PAI through the lens of interpretability theory, a lens that brings both the semantic and syntactic features of the theories under its scope into a finer focus. Recall that $\mathrm{UTB}(c)$ was introduced in Definition 2.2.2(b), and $\omega$-interpretability was introduced in Definition 2.4.1(c). In what follows $\mathrm{UTB}_0(c)$ is the obtained from $\mathrm{UTB}(c)$ by weakening $\mathrm{PA}(T)$ to $\mathrm{PA} + \mathrm{I}\Delta_0(T)$ in the definition of $\mathrm{UTB}(c)$.

**5.1 Theorem** *(mutual $\omega$-interpretability results).*

***(a)*** PAI, $\mathrm{UTB}_0(c)$, *and* $\mathrm{UTB}(c)$ *are pairwise mutually $\omega$-interpretable.*
***(b)*** $\mathrm{PAI} + \alpha$ *and* $\mathrm{CT}(c)$ *are pairwise mutually $\omega$-interpretable.*
***(c)*** $\mathrm{PAI}^{\circ} + \alpha$ *and* CT *are pairwise mutually $\omega$-interpretable.*

***Proof*** (a) is easy to show using the proofs of Theorem 4.1 and the $(ii) \Rightarrow (iii)$ direction of Theorem 4.6 (note that the 'infinite constant' $c$ in $\mathrm{UTB}_0(c)$ and in $\mathrm{UTB}(c)$ can be readily defined in a model of PAI as " the least element of $I$" ). Clearly (b) follows from Theorem 4.12. The proof of (c) is similar to the proof of (b). □

We need some general definitions and results before presenting other results about PAI. The following definition is motivated by the work of Albert Visser [23]; it was introduced in [4].

**5.2 Definition** Suppose $U$ is a first order theory.

**(a)** $U$ is *solid* iff the following property ($\mathbb{S}$) holds for all models $\mathcal{M}$, $\mathcal{M}^*$, and $\mathcal{N}$ of $U$:
($\mathbb{S}$) If $\mathcal{M} \trianglerighteq \mathcal{N} \trianglerighteq \mathcal{M}^*$ and there is an $\mathcal{M}$-definable isomorphism $i_0 : \mathcal{M} \to \mathcal{M}^*$, then there is an $\mathcal{M}$-definable isomorphism $i : \mathcal{M} \to \mathcal{N}$.
**(b)** $U$ is *nowhere solid* if ($\mathbb{S}$) is false at every model $\mathcal{M}$ of $U$, i.e., for every model $\mathcal{M}$ of $U$ there exist models $\mathcal{M}^*$, and $\mathcal{N}$ of $U$ such that $\mathcal{M} \trianglerighteq \mathcal{N} \trianglerighteq \mathcal{M}^*$ and there is an $\mathcal{M}$-definable isomorphism $i_0 : \mathcal{M} \to \mathcal{M}^*$, but there is no $\mathcal{M}$-definable isomorphism $i : \mathcal{M} \to \mathcal{N}$.

Visser showed that PA is a solid theory. The following proposition can be readily established using the definitions involved (the proof is straightforward, but notationally complicated). Recall that the notion of a retract was defined in Definition 2.4.1(f).

**5.3 Proposition** *If $U$ is a solid theory and $V$ is a retract of $U$, then $V$ is also solid. In particular, solidity is preserved by bi-interpretations.*
*The proof of solidity of PA shows the more general theorem below that will be useful; the proof of Theorem 5.4 is a slight variant of the proof of solidity of PA presented in* [4]*; we present it for the convenience of the reader.*

**5.4 Theorem** *Suppose $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are models of* PA*, and that $\mathcal{M}_i^+$ is an $\mathcal{L}_i$-structure that is an expansion of $\mathcal{M}_i$ and $\mathcal{M}_i^+ \models \mathrm{PA}(\mathcal{L}_i)$ (for $i = 1, 2, 3$). Then $(\mathbb{S}^+)$ below holds:*
$(\mathbb{S}^+)$ *If $\mathcal{M}_1^+ \trianglerighteq \mathcal{M}_2^+ \trianglerighteq \mathcal{M}_3^+$ and there is an $\mathcal{M}_1^+$-definable isomorphism $i_0 : \mathcal{M}_1 \to \mathcal{M}_3$, then there is an $\mathcal{M}_1^+$-definable isomorphism $i : \mathcal{M}_1 \to \mathcal{M}_2$.*[9]
*Consequently, an isomorphic copy of $\mathcal{M}_2^+$ is $\omega$-interpretable in $\mathcal{M}_1^+$ (moreover, the isomorphism at work is $\mathcal{M}_1^+$- definable).*

***Proof*** Suppose $\mathcal{M}_1^+$, $\mathcal{M}_2^+$, and $\mathcal{M}_3^+$ are as in the assumption of the theorem. Further, assume that:

$$\mathcal{M}_1^+ \trianglerighteq \mathcal{M}_2^+ \trianglerighteq \mathcal{M}_3^+,$$

and suppose there is an $\mathcal{M}_1^+$-definable isomorphism $i_0 : \mathcal{M}_1 \to \mathcal{M}_3$. A key property of $\mathrm{PA}(\mathcal{L})$ is that if $\mathcal{M}^+ \models \mathrm{PA}(\mathcal{L})$ and $\mathcal{N}$ is a model of the fragment Q (Robinson's arithmetic) of PA, then as soon as $\mathcal{M}^+ \trianglerighteq \mathcal{N}$, there is an $\mathcal{M}^+$ -definable initial embedding $j : \mathcal{M} \to \mathcal{N}$, i.e., an embedding $j$ such that the image $j(\mathcal{M})$ of $\mathcal{M}$ is an initial submodel of $\mathcal{N}$ (where $\mathcal{M}$ is the $\mathcal{L}_A$-reduct of $\mathcal{M}^+$). Hence there is an $\mathcal{M}_1^+$-definable initial embedding $j_0 : \mathcal{M}_1 \to \mathcal{M}_3$ and an $\mathcal{M}_2^+$ -definable initial embedding $j_1 : \mathcal{M}_1 \to \mathcal{M}_3$.

We claim that both $j_0$ and $j_1$ are surjective. To see this, suppose not. Then $j(\mathcal{M}_1)$ is a proper initial segment of $\mathcal{M}_3$, where $j$ is the $\mathcal{M}_1^+$-definable embedding $j : \mathcal{M}_1 \to \mathcal{M}_3$ given by $j := j_1 \circ j_0$. But then $i_0^{-1}(j(M_1))$ is a proper $\mathcal{M}_1^+$ -definable initial segment of $\mathcal{M}$ with no last element. This is a contradiction since $\mathcal{M}_1^+$ is a model of $\mathrm{PA}(\mathcal{L}_1)$, and therefore no proper initial segment of $M_1$ is $\mathcal{M}_1^+$-definable. Hence $j_0$ and $j_1$ are both surjective; in particular $j_0$ serves as the desired $\mathcal{M}_1^+$ -definable isomorphism between $\mathcal{M}_1$ and $\mathcal{M}_2$.

Since by assumption $\mathcal{M}_1^+ \trianglerighteq \mathcal{M}_2^+$, the $\mathcal{M}_1^+$-definable isomorphism $j_0$ allows us to construct, definably in $\mathcal{M}_1^+$, an isomorphic copy of $\mathcal{M}_2^+$ whose $\mathcal{L}_A$-reduct is $\mathcal{M}_1$. Consequently, an isomorphic copy of $\mathcal{M}_2^+$ is $\omega$ -interpretable in $\mathcal{M}_1^+$. More specifically, the $\omega$-interpretation $\mathcal{I}$ at work has the same universe and arithmetical operations as $\mathcal{M}_1$, and for each $n$-ary relation symbol $R \in \mathcal{L}_2 \backslash \mathcal{L}_1$, the $\mathcal{I}$-interpretation $R^{\mathcal{I}}$ of $R$ is given by $(x_1, \cdots, x_n) \in R^{\mathcal{I}}$ iff $\mathcal{M}_2^+ \models R(j_0(x_1), \cdots, j_0(x_n))$ (the $\mathcal{I}$-interpretation of function symbols is analogously defined). □

**5.5 Theorem** *If $U \in \{\mathrm{PAI}^\circ, \mathrm{PAI}, \mathrm{PAI} + \alpha\}$, then $U$ is nowhere solid (a fortiori: $U$ is not solid).*

[9] Note that the conclusion $i : \mathcal{M}_1 \to \mathcal{M}_2$ cannot in general be strengthened to $i : \mathcal{M}_1^+ \to \mathcal{M}_2^+$, e.g., let $\mathcal{M}_1^+ = (\mathcal{M}, D_1)$ and $\mathcal{M}_1^+ = (\mathcal{M}, D_2)$, where $D_1$ and $D_2$ are distinct $\mathcal{M}$-definable subsets of $M$.

***Proof*** We present the argument for $U = \mathrm{PAI}$, similar arguments work for $\mathrm{PAI}^{\circ}$, and $\mathrm{PAI} + \alpha$. Suppose $(\mathcal{M}, I) \models \mathrm{PAI}$, and let $2I = \{2i : i \in I\}$. The diagonal indiscernibility property of $I$ makes it evident that $(\mathcal{M}, 2I)$ is also a model of PAI. Note that:

$$(\mathcal{M}, I) \trianglerighteq (\mathcal{M}, 2I) \trianglerighteq (\mathcal{M}, I).$$

Clearly the identity map serves as an $(\mathcal{M}, I_1)$ -definable isomorphism $i_0 : (\mathcal{M}, I_1) \to (\mathcal{M}, I_1)$. However, there is no $(\mathcal{M}, I)$-definable isomorphism $f$ between $(\mathcal{M}, I)$ and $(\mathcal{M}, 2I)$ since any model of $\mathrm{PA}(I)$ is definably rigid, i.e., no automorphism of $(\mathcal{M}, I)$ is $(\mathcal{M}, I)$-definable. Thus any such purported $f$ is the identity function, which makes it impossible for $f$ to be an isomorphism between $(\mathcal{M}, I)$ and $(\mathcal{M}, 2I)$. This shows that PAI is not solid. □

**5.6 Theorem** *If* $U \in \{\mathrm{PAI}^{\circ}, \mathrm{PAI}$ *and* $\mathrm{PAI} + \alpha\}$, *then* PA *and* $U$ *are not retracts of each other (a fortiori:* PA *and* $U$ *are not bi-interpretable).*

***Proof*** Again, we present the argument for $U = \mathrm{PAI}$, similar arguments work for $\mathrm{PAI}^{\circ}$ and $\mathrm{PAI} + \alpha$. Proposition 5.4 together with Theorem 5.5 show that PAI is not a retract of PA. To see that PA is not a retract of PAI, it suffices to observe that if some model $\mathcal{M}$ of PA is a retract of a model of PAI, then by Theorem 5.4, $\mathcal{M}$ can parametrically define a class $I$ of indiscernibles for itself. This contradicts Corollary 4.3(a). □

**5.6 Theorem** CT *is solid.*

***Proof*** The solidity of CT can be established with the help of Theorem 5.4 and the fact that if $(\mathcal{M}, T_1, T_2) \models \mathrm{PA}(T_1, T_2)$, where $(\mathcal{M}, T_1)$ and $(\mathcal{M}, T_2)$ are both models of CT, then $T_1 = T_2$ (the proof is based on a simple induction, taking advantage of the assumption that both $T_1$ and $T_2$ satisfy Tarski's recursive clauses for all arithmetical formulae in $\mathcal{M}$). □

***5.8 Remark*** It is not hard to see that the none of the theories $\mathrm{CT}(c)$, UTB, and $\mathrm{UTB}(c)$ are solid. However, $\mathrm{CT}(c)$ has consistent solid extensions. For example, consider the extension of $\mathrm{CT}(c)$ given by $\mathrm{CT}(c)$ + " CT is inconsistent" + " $c$ is the length of the shortest proof of inconsistency of CT". By Gödel's second incompleteness this theory is consistent, and by a reasoning very similar to the proof of Theorem 5.6 it is also solid.

**5.9 Theorem** CT *is a retract of* $\mathrm{PAI}^{\circ} + \alpha$*, and* $\mathrm{CT}(c)$ *is a retract of* $\mathrm{PAI} + \alpha$.

***Proof*** By Theorem 4.14, there is a (uniform) $\omega$ -interpretation $\mathcal{I}_c$ of a model $(\mathcal{M}, I^{>c})$ of $\mathrm{PAI} + \alpha$ within any model $(\mathcal{M}, T, c)$ of $\mathrm{CT}(c)$. On the other hand, the definition of $\mathrm{PAI} + \alpha$ makes it clear that there is a (uniform) $\omega$-interpretation $\mathcal{J}$ of $(\mathcal{M}, T)$ within $(\mathcal{M}, I^{>c})$. A slight variant of this argument (without the use of the constant $c$) shows that CT is a retract of $\mathrm{PAI}^{\circ} + \alpha$, but we need a variation of the interpretation $\mathcal{I}_c$ in order to show that $\mathrm{CT}(c)$ is a retract of $\mathrm{PAI} + \alpha$ because it not clear that the element $c$ is definable in $(M, I^{>c})$. We can get around this problem by modifying the interpretation $\mathcal{I}$ as follows: Given any model $(\mathcal{M}, T, c)$ of CT$(c)$, we first define $I^{>c}$ as in the interpretation $\mathcal{I}$, and then we define the modified interpretation $\mathcal{I}_c$ given by:

$$\mathcal{I}_c(\mathcal{M}, T, c) = (\mathcal{M}, I_c, J), \text{ where } I_c = \{\langle c, i\rangle : i \in I^{>c}\}.$$

Thanks to the diagonal indiscernibility property of $I^{>c}$, $(\mathcal{M}, I_c)$ is a model of PAI. Moreover, $(\mathcal{M}, I_c)$ can be shown to be a model of PAI $+ \alpha$ with the same argument used in the proof of Theorem 4.14 (relying on Lemma 4.11). We can now readily define an interpretation $\mathcal{J}'$ that inverts $\mathcal{I}_c$ by letting $\mathcal{J}'(\mathcal{M}, I_c) = (\mathcal{M}, T, c)$, where $T$ is the unique truth predicate corresponding to the partial satisfaction class given by the formula $\sigma$ (of Theorem 4.1), and $c$ is defined as " the first coordinate of the ordered pair canonically coded by any member of $I_c$". Thus CT$(c)$ is a retract of PAI $+ \alpha$.

**5.10 Theorem** PAI$^\circ + \alpha$ *is not a retract of* CT.

***Proof*** We begin with observing that Proposition 5.3 and Theorem 5.5 show that the $\omega$-interpretation $\mathcal{I}_\sigma$ of CT in PAI$^\circ$ given by the formula $\sigma$ of Theorem 4.1 is not 'invertible', in the sense that there is no interpretation of PAI in CT such that $\mathcal{I}_\sigma$ and $\mathcal{J}$ witness that PAI is retract of CT. We next note that if there are interpretations $\mathcal{I}$ and $\mathcal{J}$ that witness that PAI $+ \alpha$ is a retract of CT, then Theorem 5.4 assures us that verifiably in PAI$^\circ$, the interpretation $\mathcal{I}$ is the same as $\mathcal{I}_\sigma$ up to a definable permutation of universe. This shows that $\mathcal{I}$ is not invertible either, thus concluding the proof. □

## 6 Fragments of PAI

In this section we briefly examine the model-theoretic behavior of subsystems PAI$_n$ $(n \in \omega)$ and PAI$^-$ of PAI.

**5.3 Definition** For $n \in \omega$, PAI$_n$ is the subsystem of PAI in which the extended induction scheme involving $I$ is limited to $\Sigma_n(I)$-formulae, i.e., the axioms of PAI$_n$ consist of PA plus the fragment I$\Sigma_n(I)$ of PA$(I)$, plus axioms (2) and (3) of Definition 3.1 asserting the unboundedness and indiscernibility of $I$. PAI$^-$ is the subsystem of PAI$_0$ with no extended induction scheme involving $I$, so the axioms of PAI$^-$ consist of PA plus axioms (2) and (3) of Definition 3.1.

- Given $\mathcal{M} \models$ PA, it is evident that $(\mathcal{M}, I) \models$ PAI$^-$ iff $I$ is an unbounded set of indiscernibles in $\mathcal{M}$; and by Theorem 2.1.2, $(\mathcal{M}, I) \models$ PAI$_0$ iff $I$ is a piecewise-coded unbounded set of indiscernibles in $\mathcal{M}$.

**6.2 Theorem** *Every model of* PA *has an elementary end extension that has an expansion to a model of* PAI$_0$, *but not to a model of* PAI.

***Proof*** Fix a minimal type $p(x)$ and any model $\mathcal{M}_0$ of PA, and let $\mathcal{M}$ be an $\omega$-canonical extension of $\mathcal{M}_0$ using $p(x)$ as in section 3.3 of [13]. Thus $\mathcal{M}$ is obtained by an $\omega$-iteration of the process of adjoining an element satisfying $p(x)$. By Theorem 2.3.5 $p(x)$ is an unbounded indiscernible type, which makes it clear that $\mathcal{M}$ carries an unbounded indiscernible subset $I$, and additionally the order-type of $I$ is $\omega$. The latter feature makes it clear that $I$ is piecewise-coded in $\mathcal{M}$, and thus $(\mathcal{M}, I) \models \mathrm{PAI}_0$ in light of Theorem 2.1.1.

It remains to show that $\mathcal{M}$ does not have an expansion to PAI. Suppose not, and let $(\mathcal{M}, I) \models \mathrm{PAI}$. It is easy to see that from the point of view of $(\mathcal{M}, I)$, the order-type of $(I, <)$ is the same as the order-type of $(M, <)$ (where $<$ is the ordering on $M$ given by $\mathcal{M}$). Recall that $\mathcal{M}$ can be written as the union of elementary initial submodels $\mathcal{M}_n$ of $\mathcal{M}$ (as $n$ ranges in $\omega$), where $\mathcal{M}_n$ is obtained by $n$-repetitions of the process of adjoining an element satisfying $p(x)$. By minimality of $p(x)$ this assures us that:
$(*)$ For each $n \in \omega$, and each choice of $c_n \in M_{n+1}\backslash M_n$, $(\mathcal{M}_n, c_1, \ldots, c_n)$ is pointwise definable.
The existence of the above isomorphism $f$ makes it clear that there is some $k \geq 1$ such that $I \cap (M_k\backslash M_{k-1})$ is infinite. In particular we can pick distinct $i_1$ and $i_2$ in $I \cap (M_k\backslash M_{k-1})$, together with elements $\{c_s : 1 \leq s \leq k\}$ such that $c_s \in M_{s+1}\backslash M_s$ for each $s$, and moreover $c_k$ is below both $i_1$ and $i_2$ (since $M_k\backslash M_{k-1}$ has no least element). By the diagonal indiscernibility property of $I$, this implies that $i_1$ and $i_2$ are indiscernible in the structure $(\mathcal{M}_k, c_1, \ldots, c_k, m)_{m\in M_0}$. But this indiscernibility contradicts $(*)$, and thereby completes the proof. □

**6.3 Theorem** *If* $\mathcal{M}$ *is a model of* PA *of countable cofinality that is expandable to a model of* $\mathrm{PAI}^-$*, then* $\mathcal{M}$ *is expandable to a model of* $\mathrm{PAI}_0$*. However, every countable model of* PA *has an uncountable elementary end extension that is expandable to a model of* $\mathrm{PAI}^-$*, but not to* $\mathrm{PAI}_0$.

***Proof*** Suppose $(\mathcal{M}, I) \models \mathrm{PAI}^-$, where $\mathcal{M}$ has countable cofinality. The countable cofinality of $\mathcal{M}$ allows us to construct an unbounded subset $I_0$ of $I$ of order type $\omega$. Since every subset of $\mathcal{M}$ of order-type $\omega$ is piecewise-coded, by Theorem 2.1.1, $(\mathcal{M}, I_0) \models \mathrm{PAI}_0$. To demonstrate the second assertion of the theorem, let $\mathcal{M}_0$ be a countable model of PA and $\mathcal{M}$ be an $\omega_1$-canonical extension of $\mathcal{M}_0$ using some minimal type $p(x)$, i.e., $\mathcal{M}$ is obtained by an $\omega_1$-iteration of the process of adjoining an element satisfying $p(x)$ (as in [13, Section 3.3]). By Theorem 2.2.14 of [13], $\mathcal{M}$ is rather classless, i.e., every piecewise-coded subset of $M$ is definable in $\mathcal{M}$. Thus if $(\mathcal{M}, I) \models \mathrm{PAI}_0$, then $(\mathcal{M}, I) \models \mathrm{PAI}$, and $I$ is $\mathcal{M}$-definable, which contradicts Corollary 4.3(a). □

***6.4 Remark*** As pointed out by the referee, *recursively saturated models of* PA *of countable cofinality are expandable to models of* PAI. To see this, suppose $\mathcal{M}$ is a recursively saturated model of PA of countable cofinality. Then $\mathcal{M}$ has a countable recursively saturated cofinal submodel $\mathcal{K}$. By Theorems 2.2.4 and 4.6. $\mathcal{K}$ has an expansion $(\mathcal{K}, J)$ to a model of PAI. By a theorem of Kotlarski and Schmerl [13, Theorem 3.1.7], there is a unique $I$ such that $(\mathcal{M}, I)$ is an elementary extension of $(\mathcal{K}, J)$, and the result follows.

## 7 Open questions

The following question is inspired by Theorems 4.6 and 4.12.

**7.1 Question** *Is there a set of sentences* $\Sigma$ *in the language of* PAI *that has the property: A nonstandard model* $\mathcal{M}$ *of* PA *(of any cardinality) has an expansion to a model of* PAI+$\Sigma$ *iff* $\mathcal{M}$ *has an expansion to a full satisfaction class?*

The following question was suggested by the referee; it is motivated by Remark 3.4.

**7.2 Question** *Are there models* $(\mathcal{M}, I) \models$ PAI *such that the elements of* $I$ *do not realize an unbounded indiscernible type?*

Questions 7.3 through 7.4 below arise from the results obtained in Section 5. We have not succeeded in ruling out that PAI and UTB$(c)$ are not bi-interpretable; ditto for PAI$+\alpha$ and CT$(c)$. We conjecture that the questions below all have negative answers. As partial evidence for our conjecture, let us observe that Theorem 5.2 and 5.5 show that the $\omega$-interpretation $\mathcal{I}_\sigma$ of UTB$(c)$ in PAI, and CT$(c)$ in PAI $+\,\alpha$, given by the formula $\sigma$ of Theorem 4.1 is not 'invertible', in the sense that there is no interpretation of PAI in UTB$(c)$ such that $\mathcal{I}_\sigma$ and $\mathcal{J}$ witness that PAI is retract of UTB$(c)$.

**7.3 Question** *Is* PAI $+\,\alpha$ *is a retract of* CT$(c)$*?*

**7.4 Question** *Is either of the pair of theories* {PAI, UTB$(c)$} *a retract of the other one?*

**7.5 Question** *Is either of the pair of theories* {PAI $+\,\alpha$, CT$(c)$} *a retract of the other one?*

Questions 7.3 through 7.4 below are motivated by the results in Section 6.

**7.6 Question** *Does Theorem 6.2 lend itself to a hierarchical generalization? In other words, is it true that for every* $n \in \omega$*, every model of* PA *has an elementary end extension that has an expansion to a model of* PAI$_n$*, but not to a model of* PAI$_{n+1}$*? (It is not even clear how to build a model of* PAI$_n$ *for* $n \in \omega$ *that is not a model of* PAI$_{n+1}$*.)*

**7.7 Question** *Is there a model* $\mathcal{M}$ *of* PA *such that* $\mathcal{M}$ *has an expansion to a model of* PAI$_n$ *for each* $n \in \omega$*, but* $\mathcal{M}$ *has no expansion to a model of* PAI*?*

**Acknowledgements** I am grateful to Lawrence Wong, Bartosz Wcisło, Mateusz Łełyk, Jim Schmerl, Roman Kossak, Kentaro Fujimoto, Cezary Cieśliński, Lev Beklemishev, and Athar Abdul-Quader (in reverse alphabetical order) for their priceless feedback and keen interest in this work. Also hats off to the anonymous referee for an enlightening report, as noted throughout the paper. The research presented in this paper was partially supported by the National Science Centre, Poland (NCN), Grant Number 2019/34/A/HS1/00399.

**Funding** Open access funding provided by University of Gothenburg.

**Competing interests** The authors declare no competing interests.

## References


1. Beklemishev, L.D., Pakhomov, F.N.: Reflection algebras and conservation results for theories of iterated truth. Ann. Pure Appl. Logic **173**(5), 103093 (2022)
2. Cieśliński, C.: The epistemic lightness of truth. Deflationism and its logic. Cambridge University Press, Cambridge (2017)
3. Cieśliński, C.: On some problems with truth and satisfaction. In: Philosophical approaches to the foundations of logic and mathematics, pp. 175–192. Brill (2021)
4. Enayat, A.: Variations on a Visserian theme. In: van Eijk, J., Iemhoff, R., Joosten, J. (eds.) Liber Amicorum Alberti a Tribute to Albert Visser, pp. 99–110. College Publications, London (2016)
5. Enayat, A.: Set theory with a proper class of indiscernibles. Fundam. Math. **259**, 33–76 (2022)
6. Enayat, A., Pakhomov, F.: Truth, disjunction, and induction. Arch. Math. Logic **58**, 753–766 (2019)
7. Hájek, P., Pudlák, P.: Metamathematics of first-order arithmetic, Perspectives in Mathematical Logic. Springer, Berlin (1998)
8. Halbach, V.: Axiomatic theories of truth, 2 nd ed. Cambridge University Press, Cambridge (2015)
9. Kaye, R.: Models of peano arithmetic. Oxford University Press, Oxford (1991)
10. Kaye, R., Wong, T.: On interpretations of arithmetic and set theory. Notre Dame J. Form. Logic **48**, 497–510 (2007)
11. Kossak, R.: A certain class of models of Peano arithmetic. J. Symb. Log. **48**, 311–320 (1983)
12. Kossak, R.: Models with the $\omega$-property. J. Symb. Log. **54**(1), 177–189 (1989)
13. Kossak, R., Schmerl, J.: The structure of models of arithmetic. Oxford University Press, Oxford (2006)
14. Kotlarski, H.: Bounded induction and satisfaction classes. Zeitschrift für matematische Logik und Grundlagen der Mathematik **32**, 531–544 (1986)
15. Kotlarski, H., Ratajczyk, Z.: Inductive full satisfaction classes. Ann. Pure Appl. Logic **47**, 199–223 (1990)
16. Łełyk, M.: Axiomatic theories of truth, bounded induction and reflection principles, doctoral dissertation University of Warsaw, (2017)
17. Łełyk, M.: Model theory and proof theory of the global reflection principle. J. Symb. Log. **88**, 738–779 (2023)
18. Lindström, P.: Aspects of incompleteness. In: Hamilton, W. (ed.) Lecture notes in logic, vol. 10. Association for Symbolic Logic, Storrs (1997)
19. Paris, J., Harrington, L.: A mathematical incompleteness in Peano arithmetic. In: Barwise, J. (ed.) Handbook of mathematical logic, pp. 1133–1142. North-Holland, Amsterdam (1977)
20. Schmerl, J.: Recursively saturated models generated by indiscernibles. Notre Dame J. Form. Logic **26**, 99–105 (1985)
21. Smoryński, C.: $\omega$-consistency and reflection. In: Colloque International de Logique, Clermont-Ferrand, 1975, Colloq. Internat. CNRS, 249, CNRS, Paris, pp. 167–181 (1977)
22. Smoryński, C., Stavi, J.: Cofinal extension preserves recursive saturation. In: Model theory of algebra and arithmetic (Proc. Conf., Karpacz), Lecture Notes in Math., vol. 834, Springer, Berlin, pp. 338–345(1980)
23. Visser, A.: Categories of theories and interpretations, Logic in Tehran. Lecture Notes in Logic, vol. 26. Association for Symbolic Logic, La Jolla, CA (2006)
24. Wcisło, B., Łełyk, M.: Notes on bounded induction for the compositional truth predicate. Rev. Symb. Logic **10**, 1–26 (2017)