Arch. Math. Logic (2013) 52:317–333
DOI 10.1007/s00153-012-0318-3

Mathematical Logic

Herbrand consistency of some finite fragments of bounded arithmetical theories

Saeed Salehi

Received: 11 October 2011 / Accepted: 2 December 2012 / Published online: 21 December 2012

Abstract We formalize the notion of Herbrand Consistency in an appropriate way for bounded arithmetics, and show the existence of a finite fragment of $\mathrm{I}\Delta_0$ whose Herbrand Consistency is not provable in $\mathrm{I}\Delta_0$. We also show the existence of an $\mathrm{I}\Delta_0$-derivable Π_1-sentence such that $\mathrm{I}\Delta_0$ cannot prove its Herbrand Consistency.

Keywords Herbrand consistency · Bounded arithmetic · Gödel's Second Incompleteness Theorem

Mathematics Subject Classification (2000) 03F40 · 03F25 · 03F30

1 Introduction

A consequence of Gödel's Second Incompleteness Theorem is Π_1-separation of some mathematical theories; for example ZFC is not Π_1-conservative over PA since $\mathrm{ZFC} \vdash \mathrm{Con}(\mathrm{PA})$ but (by Gödel's theorem) $\mathrm{PA} \nvdash \mathrm{Con}(\mathrm{PA})$, where Con is the consistency predicate. Inside PA, the hierarchy $\{\mathrm{I}\Sigma_\mathrm{n}\}_{n\geqslant 0}$ is not Π_1-conservative, since $\mathrm{I}\Sigma_{\mathrm{n+1}} \vdash \mathrm{Con}(\mathrm{I}\Sigma_\mathrm{n})$ (but again $\mathrm{I}\Sigma_\mathrm{n} \nvdash \mathrm{Con}(\mathrm{I}\Sigma_\mathrm{n})$). As for the bounded arithmetics, we only know that the elementary arithmetic $\mathrm{I}\Delta_0 + \mathrm{Exp}$ is not Π_1-conservative over $\mathrm{I}\Delta_0 + \bigwedge_\mathrm{j} \Omega_\mathrm{j}$ (see Corollary 5.34 of [6]). One candidate for Π_1-separating $\mathrm{I}\Delta_0 + \mathrm{Exp}$ from $\mathrm{I}\Delta_0$ was the Cut-Free Consistency of $\mathrm{I}\Delta_0$ (see [8]): it was already known

---

S. Salehi (✉)
Department of Mathematics, University of Tabriz,
29 Bahman Blvd., P. O. Box 51666–17766, Tabriz, Iran
e-mail: root@saeedsalehi.ir; salehipour@tabrizu.ac.ir
URL: http://saeedsalehi.ir/

S. Salehi
School of Mathematics, Institute for Research in Fundamental Sciences (IPM),
P. O. Box 19395–5746, Tehran, Iran

that $I\Delta_0 + \mathrm{Exp} \vdash \mathrm{CFCon}(I\Delta_0)$ and it was presumed that $I\Delta_0 \nvdash \mathrm{CFCon}(I\Delta_0)$, where CFCon stands for Cut-Free Consistency. Though this presumption took rather a long to be established (see [14]), it opened a new line of research.

The problem of provability (or unprovability) of the cut-free consistency of weak arithmetics is an interesting (double) generalization of Gödel's Second Incompleteness Theorem: the theory (being restricted to bounded or weak arithmetics) and also the consistency predicate are both weakened. Here, we do not intend to outline the history of this research line, and refer the reader to [11,12]. Nevertheless, we list some prominent results obtained so far, to put our new result in perspective.

Herbrand Consistency is denoted by HCon and (Semantic) Tableau Consistency by TabCon. Adamowicz (with Zbierski in 2001 [2] and) in 2002 [3] showed that $I\Delta_0 + \Omega_m \nvdash \mathrm{HCon}(I\Delta_0 + \Omega_m)$ for $m \geqslant 2$. She had already shown the unprovability $I\Delta_0 + \Omega_1 \nvdash \mathrm{TabCon}(I\Delta_0 + \Omega_1)$ in 1996 (but appeared in 2001 as [1]). Salehi improved the result of [3] in [10] by showing that $I\Delta_0 + \Omega_1 \nvdash \mathrm{HCon}(I\Delta_0 + \Omega_1)$ (see also [12]) and the result of [2] in [9,10] by showing $S \nvdash \mathrm{HCon}(S)$ where S is an $I\Delta_0$-derivable Π_2-sentence. This result also implied that $I\Delta_0 \nvdash \mathrm{HCon}(\overline{I\Delta_0})$ holds for a re-axiomatization $\overline{I\Delta_0}$ of $I\Delta_0$. Willard [13] showed in 2002 that $I\Delta_0 \nvdash \mathrm{TabCon}(I\Delta_0)$ and also $I\Delta_0 \nvdash \mathrm{HCon}(I\Delta_0 + \Omega_0)$, where Ω_0 is the axiom of the totality of the squaring function $\Omega_0 : \forall x \exists y [y = x \cdot x]$. This was improved by the author in [12] by showing $I\Delta_0 \nvdash \mathrm{HCon}(I\Delta_0)$, without using the Ω_0 axiom. It was also proved in [13] that $V \nvdash \mathrm{HCon}(V)$ for an $I\Delta_0$-derivable Π_1-sentence V. Kołodziejczyk [7] showed in 2006 that the unprovability $I\Delta_0 + \bigwedge_j \Omega_j \nvdash \mathrm{HCon}(I\Delta_0 + \Omega_1)$ holds; his result was stronger in a sense that it showed $I\Delta_0 + \bigwedge_j \Omega_j \nvdash \mathrm{HCon}(S + \Omega_1)$ for a finite fragment $S \subseteq I\Delta_0$.

In this paper we use an idea of an anonymous referee of [12] for defining evaluations in a more effective way (Definition 5) suitable for bounded arithmetics; this is a great step forward, noting our mentioning in [12] that "[o]ur definition of Herbrand Consistency is not best suited for $I\Delta_0$". We then partially answer the question proposed by the anonymous referee of [11] (see Conjecture 4.1 in [11]). The author is grateful to both the referees, for suggestions and inspirations.

We show the existence of a finite fragment T of $I\Delta_0$ such that $I\Delta_0 \nvdash \mathrm{HCon}(T)$; this generalizes the result of [12]. We also show the existence of an $I\Delta_0$-derivable Π_1-sentence U such that $I\Delta_0 \nvdash \mathrm{HCon}(U)$; this generalizes the main result of [9,10] and [13]. For keeping the paper short, and to avoid repeating some technical details, we apologetically invite the reader to consult [11,12]. We also assume familiarity with the Bible of this field [6].

2 Herbrand consistency of arithmetical theories

For getting a unique Skolemized formula, it is more convenient to negation normalize and rectify the formula.

Definition 1 (*Rectified Negation Normal Form*) A formula is in negation normal form when no implication symbol $\rightarrow$ appears in it, and the negation symbol $\neg$ appears behind the atomic formulas only.

A formula is rectified when different quantifiers refer to different variables and no variable appears both free and bound in the formula. $\diamond$

Any formula can be uniquely negation normalized by removing the implication connectives (replacing formulas of the form $A \rightarrow B$ with $\neg A \vee B$) and then pushing the negations inside the sub-formulas by de Morgan's Law, until they get to the atomic formulas. Renaming the variables can rectify any formula. Thus one can negation normalize and rectify a formula uniquely, up to a variable renaming.

Definition 2 (*Skolemization*) For any existential formula $\exists x A(x)$ with $m(\geqslant 0)$ free variables, let $\mathsf{f}_{\exists x A(x)}$ be a new m-ary function symbol (which does not occur in A; cf. [5]). For any rectified negation normal formula φ we define φ^S inductively:

- $\varphi^S = \varphi$ for atomic or negated-atomic formula φ
- $(\varphi \wedge \psi)^S = \varphi^S \wedge \psi^S$
- $(\varphi \vee \psi)^S = \varphi^S \vee \psi^S$
- $(\forall x \varphi)^S = \forall x \varphi^S$
- $(\exists x \varphi)^S = \varphi^S[\mathsf{f}_{\exists x \varphi(x)}(\overline{y})/x]$ where $\overline{y}$ are the free variables of $\exists x \varphi(x)$.

Finally, the Skolemized form φ^{Sk} of the formula φ is obtained by removing all the (universal) quantifiers of φ^S. The resulted formula is open (quantifier-less), with probably some free variables. If those (free) variables are substituted with some ground (variable-free) terms, we obtain an Skolem instance of that formula. ◇

Summing up, to get an Skolem instance of a given formula φ we first negation normalize and then rectify it to get a formula φ^{RNNF}; then we remove the quantifiers of $(\varphi^{\mathrm{RNNF}})^S$ to get $(\varphi^{\mathrm{RNNF}})^{\mathrm{Sk}}$, and substituting its free variables with some ground terms, gives us an Skolem instance of the formula φ. Let us note that the Skolem instances of a formula are determined uniquely.

Theorem 1 (Herbrand-Skolem-Gödel) *Any theory T is equi-consistent with its Skolemized theory. In other words, T is consistent if and only if every finite set of Skolem instances of T is (propositionally) satisfiable.*

Example 1 In the language of arithmetic $\mathscr{L}_A = \{\mathsf{0}, \mathsf{S}, +, \cdot, \leqslant\}$, let $\mathrm{Ind}_\Box$ be

$$\psi(\mathsf{0}) \wedge \forall x[\psi(x) \rightarrow \psi(\mathsf{S}(x))] \rightarrow \forall x \psi(x)$$
$$\text{where } \psi(x) = \exists y[y \leqslant x \cdot x \wedge y = x \cdot x].$$

This is an axiom of $\mathrm{I}\Delta_0$. Rectified Negation Normal Form $(\mathrm{Ind}_\Box)^{\mathrm{RNNF}}$ of $\mathrm{Ind}_\Box$ is

$$\begin{aligned}
&\forall u[u \not\leqslant \mathsf{0} \cdot \mathsf{0} \vee u \neq \mathsf{0} \cdot \mathsf{0}] \bigvee \\
&\exists w\Big[\exists z[z \leqslant w \cdot w \wedge z = w \cdot w] \wedge \forall v[v \not\leqslant \mathsf{S}(w) \cdot \mathsf{S}(w) \vee v \neq \mathsf{S}(w) \cdot \mathsf{S}(w)]\Big] \bigvee \\
&\forall x \exists y[y \leqslant x \cdot x \wedge y = x \cdot x].
\end{aligned}$$

Then $((\mathrm{Ind}_\Box)^{\mathrm{RNNF}})^S$ can be computed as:

$$\begin{aligned}
&\forall u[u \not\leqslant \mathsf{0} \cdot \mathsf{0} \vee u \neq \mathsf{0} \cdot \mathsf{0}] \bigvee \\
&\big[[\mathfrak{q}(\mathfrak{c}) \leqslant \mathfrak{c} \cdot \mathfrak{c} \wedge \mathfrak{q}(\mathfrak{c}) = \mathfrak{c} \cdot \mathfrak{c}] \wedge \forall v[v \not\leqslant \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \vee v \neq \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c})]\big] \bigvee \\
&\forall x[\mathfrak{q}(x) \leqslant x \cdot x \wedge \mathfrak{q}(x) = x \cdot x],
\end{aligned}$$

where $\mathfrak{q}(x)$ is the Skolem function symbol for the formula $\exists z[z \leqslant x \cdot x \wedge z = x \cdot x]$, and the constant $\mathfrak{c}$ is the Skolem function symbol for the sentence of the second disjunct:

$$\exists w\big[\exists z[z \leqslant w \cdot w \wedge z = w \cdot w] \wedge \forall v[v \nleqslant \mathsf{S}(w) \cdot \mathsf{S}(w) \vee v \neq \mathsf{S}(w) \cdot \mathsf{S}(w)]\big].$$

Finally, the Skolemized form $(\mathrm{Ind}_{\square})^{\mathrm{Sk}}$ of φ is obtained as:

$$\begin{aligned}
&[u \nleqslant \mathsf{0} \cdot \mathsf{0} \vee u \neq \mathsf{0} \cdot \mathsf{0}] \bigvee \\
&\big[[\mathfrak{q}(\mathfrak{c}) \leqslant \mathfrak{c} \cdot \mathfrak{c} \wedge \mathfrak{q}(\mathfrak{c}) = \mathfrak{c} \cdot \mathfrak{c}] \wedge [v \nleqslant \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \vee v \neq \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c})]\big] \bigvee \\
&[\mathfrak{q}(x) \leqslant x \cdot x \wedge \mathfrak{q}(x) = x \cdot x].
\end{aligned}$$

Substituting $u/\mathsf{0}$, $v/\mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c})$, x/t will result in the following Skolem instance of φ:

$$\begin{aligned}
&[\mathsf{0} \nleqslant \mathsf{0} \cdot \mathsf{0} \vee \mathsf{0} \neq \mathsf{0} \cdot \mathsf{0}] \bigvee \\
&\big[[\mathfrak{q}(\mathfrak{c}) \leqslant \mathfrak{c} \cdot \mathfrak{c} \wedge \mathfrak{q}(\mathfrak{c}) = \mathfrak{c} \cdot \mathfrak{c}] \wedge [\mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \nleqslant \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \vee \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \neq \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c})]\big] \bigvee \\
&[\mathfrak{q}(t) \leqslant t \cdot t \wedge \mathfrak{q}(t) = t \cdot t].
\end{aligned}$$

◇

Propositional satisfiability is usually arithmetized from the usual provability, only in propositional logic (see e.g. [6]); but in a series of more recent papers, this notion has been arithmetized differently, according to ones needs ([1–4,7,9–13]). We formalize the notion of propositional satisfiability by means of evaluations (as in the op. cit. papers) on sets of (Skolem) ground terms, but in a more effective way. To get a small evaluation on a given set of terms, we first sort its members, and then require the equality relation to be a congruence.

We will call the ground terms constructed from Skolem function (and constant) symbols, simply *terms*. The $\mathscr{L}_A$-terms, where $\mathscr{L}_A = \{\mathsf{0}, \mathsf{S}, +, \cdot, \leqslant\}$ is the language of arithmetic, will be written by typewriter font (like $\mathtt{r}, \mathtt{t}, \mathtt{s}, \ldots$) and the other (Skolem) terms will be written in italic font (like $r, s, t, \ldots$). For a set A, its cardinality will be denoted by $|A|$, and for a sequence p, its length will be also denoted by $|p|$. For the elements of p, the $(i+1)$th member of p is denoted by $(p)_i$ for any $i < |p|$; so $p = \langle (p)_0, (p)_1, \ldots (p)_{|p|-1} \rangle$. Let $\approx$ and $\prec$ be two new symbols not in $\mathscr{L}_A$.

Definition 3 (*Pre-Evaluation*) For a set of terms Λ (with $|\Lambda| \geqslant 2$), a pre-evaluation on Λ is a sequence p that satisfies the following conditions:

(1) length of p is $|p| = 2|\Lambda| - 1$;
(2) for any $0 \leqslant i \leqslant |\Lambda| - 1$ we have $(p)_{2i} \in \Lambda$;
(3) for any $1 \leqslant i \leqslant |\Lambda| - 1$ we have $(p)_{2i-1} \in \{\prec, \approx\}$;
(4) for any term $t \in \Lambda$ there exists a unique $0 \leqslant j \leqslant |\Lambda| - 1$ such that $(p)_{2j} = t$. ◇

In other words, a pre-evaluation on Λ sorts (organizes) the terms in Λ, starting from the smallest and ending in the largest.

Example 2 A pre-evaluation on $\{\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6\}$ is a sequence like $p = \langle \alpha_4, \prec, \alpha_7, \approx, \alpha_1, \approx, \alpha_5, \prec, \alpha_3, \prec \alpha_6, \approx, \alpha_2 \rangle$. ◇

Here we note that a *sub-string* of a sequence $\alpha_1\alpha_2\ldots\alpha_n$ is a sub-sequence of it in the form $\alpha_{1+i}\alpha_{2+i}\ldots\alpha_{m+i}$ where $0 \leqslant i$ and $m+i \leqslant n$; in other words, a sub-string of a sequence is a prefix of a suffix of that sequence.

Definition 4 (*Equality and Order in Pre-Evaluations*) In a pre-evaluation p on Λ define the relations $\approx_p$ and $\prec_p$ on Λ^2 by the following conditions for $s, t \in \Lambda$:

(1) $s \approx_p t$ if there exists a sub-string q of p of length $2l-1$ ($l \geqslant 1$) such that
  (a) either $((q)_0 = s \,\&\, (q)_{2l-2} = t)$ or $((q)_0 = t \,\&\, (q)_{2l-2} = s)$;
  (b) for any $1 \leqslant i \leqslant l-1$, $(q)_{2i-1} = \approx$.

(2) $s \prec_p t$ if there exists a sub-string q of p of length $2l-1$ ($l \geqslant 1$) such that
  (a) $(q)_0 = s$ and $(q)_{2l-2} = t$;
  (b) there exists some $1 \leqslant i \leqslant l-1$ for which $(q)_{2i-1} = \prec$. ◇

Example 2 (Continued) We have $\alpha_1 \approx_p \alpha_5 \approx_p \alpha_7$ and $\alpha_2 \approx_p \alpha_6$. Also, $\alpha_4 \prec_p \alpha_1, \alpha_4 \prec_p \alpha_5, \alpha_4 \prec_p \alpha_7, \alpha_1 \prec_p \alpha_2, \alpha_1 \prec_p \alpha_3$, and $\alpha_1 \prec_p \alpha_6$ hold. ◇

Lemma 1 (Equivalence and Order by Pre-Evaluation) *Let Λ be a set of terms, and p be a pre-evaluation on Λ.*

(1) The relation $\approx_p$ is an equivalence on Λ.
(2) The relation $\prec_p$ is a total order on $\Lambda/\approx_p$.
(3) The relations $\approx_p$ and $\prec_p$ are compatible with each other: if $t \approx_p s$, and $t \prec_p u$ (respectively, $u \prec_p t$), then $s \prec_p u$ (respectively, $u \prec_p s$).

Proof The parts (1) and (2) are immediate. For (3), suppose $t \approx_p s$ and $t \prec_p u$. Then there is a sub-string q of p which starts from t and ends with u and contains at least one special symbol $\prec$. There must also be some other sub-string q' which starts from either t or s and ends with the other one, and all its special symbols are equality $\approx$. If q' starts from s (and so ends with t), then the concatenation of q' and q results in a sub-string which starts from s and ends with u and contains some special symbol $\prec$. Whence $s \prec_p u$. And if q' starts from t, then q cannot be a sub-string of q' because all the special symbols in q' are $\approx$ and q contains at least one special symbol $\prec$. Thus q' has to be a sub-string of q. Then there must exist a sub-string of p which starts from s and ends with u and contains a special symbol $\prec$; whence $s \prec_p u$. The other case ($u \prec_p t$) can be proved very similarly. □

Definition 5 (*Evaluation*) A pre-evaluation p on a set of terms Λ is called an evaluation when for any term $t, s \in \Lambda$ and any term $u(x)$ with the free variable x, if $t \approx_p s$ and $u(t/x), u(s/x) \in \Lambda$ then $u(t/x) \approx_p u(s/x)$. ◇

In other words, an evaluation on Λ is a pre-evaluation p on Λ whose equivalence relation $\approx_p$ is a congruence relation on Λ.

Definition 6 (*Satisfaction in an Evaluation*) Let Λ be a set of terms and p an evaluation on it. For terms $t, s \in \Lambda$ we write $p \models t = s$ when $t \approx_p s$ holds. We also write $p \models t \leqslant s$ when either $t \approx_p s$ or $t \prec_p s$ holds. So, for atomic formulas φ in the language of arithmetic $\mathscr{L}_A$ we have defined the notion of satisfaction $p \models \varphi$. The satisfaction relations can be extended to all open (quantifier-less) formulas as usual:

- $p \models \varphi \wedge \psi \iff p \models \varphi$ and $p \models \psi$
- $p \models \varphi \vee \psi \iff p \models \varphi$ or $p \models \psi$
- $p \models \varphi \rightarrow \psi \iff$ if $p \models \varphi$ then $p \models \psi$
- $p \models \neg\varphi \iff p \not\models \varphi$ ◇

Lemma 2 (Leibniz's Law) *Any evaluation p on any set of terms Λ satisfies all the available Skolem instances of the axioms of equational logic, in particular Leibniz's Law: for any $t, s \in \Lambda$ and any open formula $\varphi(x)$, we have $p \models t = s \wedge \varphi(t) \rightarrow \varphi(s)$.*

Proof Suppose $p \models t = s$. By induction on (the complexity) of (the open formula) φ one can show that $p \models \varphi(t)$ if and only if $p \models \varphi(s)$. For atomic φ it follows from Lemma 1 (on the compatibility of $\prec_p$ and $\approx_p$), and for the more complex formulas it follows from the inductive definition of satisfaction in evaluations. □

Definition 7 (T*-evaluation on* Λ) For a set of terms Λ, an Skolem instance of a formula is called to be available in Λ if all the terms appearing in it belong to Λ. For a theory T and a set of terms Λ and an evaluation p on Λ, we say that p is an T-evaluation on Λ if p satisfies every Skolem instance of every sentence in T which is available in Λ. ◇

So, T-evaluations, for a theory T, are kind of partial models of T.

Example 3 Let T be axiomatized by the following sentences in $\mathscr{L}_A$:

- $\forall x[x \cdot \mathsf{0} = \mathsf{0}]$;
- $\exists y \leqslant \mathsf{0} \cdot \mathsf{0}[y = \mathsf{0} \cdot \mathsf{0}] \wedge \forall x\big[\exists y \leqslant x \cdot x[y = x \cdot x] \rightarrow$
  $\exists y \leqslant \mathsf{S}(x) \cdot \mathsf{S}(x)[y = \mathsf{S}(x) \cdot \mathsf{S}(x)]\big] \rightarrow \forall x \exists y \leqslant x \cdot x[y = x \cdot x]$.

Let $\Lambda = \{\mathsf{0}, \mathsf{0}\cdot\mathsf{0}, \mathfrak{c}, \mathfrak{c}\cdot\mathfrak{c}, \mathfrak{q}(\mathfrak{c}), \mathsf{S}(\mathfrak{c})\cdot\mathsf{S}(\mathfrak{c}), t, t\cdot t, \mathfrak{q}(t)\}$ where $\mathfrak{c}$ and $\mathfrak{q}$ are as in Example 1. As we saw in that example, the following is an instance of the second axiom ($\mathrm{Ind}_{\Box}$), which is also available in Λ:

$$[\mathsf{0} \not\leqslant \mathsf{0} \cdot \mathsf{0} \vee \mathsf{0} \neq \mathsf{0} \cdot \mathsf{0}] \ \bigvee$$
$$\big[[\mathfrak{q}(\mathfrak{c}) \leqslant \mathfrak{c} \cdot \mathfrak{c} \wedge \mathfrak{q}(\mathfrak{c}) = \mathfrak{c} \cdot \mathfrak{c}] \wedge [\mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \not\leqslant \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \vee \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c}) \neq \mathsf{S}(\mathfrak{c}) \cdot \mathsf{S}(\mathfrak{c})]\big] \ \bigvee$$
$$[\mathfrak{q}(t) \leqslant t \cdot t \wedge \mathfrak{q}(t) = t \cdot t].$$

Suppose p is an T-evaluation on Λ. By the first axiom p must satisfy the instance $\mathsf{0} \cdot \mathsf{0} = \mathsf{0}$, so we should have $p \models \mathsf{0} \cdot \mathsf{0} = \mathsf{0}$. Thus, p cannot satisfy the first disjunct of the above instance. Indeed, p cannot satisfy the second disjunct either, because for any term u we have $p \models u \leqslant u \wedge u = u$. Thus, p cannot satisfy the second conjunct of the second disjunct. Whence, p must satisfy the third disjunct of the above instance, and in particular we should have $p \models \mathfrak{q}(t) = t \cdot t$. ◇

For a theory T, if Λ is the set of all (ground) terms (constructed from the language of T and the Skolem function symbols of the axioms of T), then any T-evaluaton on Γ (if exists) is a *Herbrand Model* of T. Now, Herbrand's Theorem can be read as

A theory T is consistent if and only if
for every finite set of (Skolem) terms, there exists an T-evaluation on it.

Thus, the notion of *Herbrand Consistency* of a theory T is (equivalent to) the existence of an T-evaluation on any (finite) set of terms.

Definition 8 (*Skolem Hull*) Let $\mathscr{L}_A^{\rm Sk}$ be the language expanding $\mathscr{L}_A$ by the Skolem function (and constant) symbols of all the existential formulas in the language $\mathscr{L}_A$. That is $\mathscr{L}_A^{\rm Sk} = \{\mathfrak{f}_{\exists x\varphi(x)} \mid \varphi \text{ is an } \mathscr{L}_A - \text{formula}\}$. For a given set of terms Λ, let $\Lambda^{\langle j\rangle}$ be defined by induction on j:

$$\begin{aligned}\Lambda^{\langle 0\rangle} &= \Lambda;\\ \Lambda^{\langle j+1\rangle} &= \Lambda^{\langle j\rangle} \cup \left\{f(t_1,\ldots,t_m) \mid f\in\mathscr{L} \wedge t_1,\ldots,t_m\in\Lambda^{\langle j\rangle}\right\}\\ &\quad\cup\left\{\mathfrak{f}_{\exists x\varphi(x)}(t_1,\ldots,t_m) \mid \ulcorner\varphi\urcorner\leqslant j \wedge t_1,\ldots,t_m\in\Lambda^{\langle j\rangle}\right\},\end{aligned}$$

where $\ulcorner\varphi\urcorner$ is the Gödel code of φ. ◇

Bounding the Gödel code of φ in the above definition will enable us to have some efficient (upper bound) for the Gödel code of $\Lambda^{\langle j\rangle}$ (see [11,12]).

Herbrand's theorem implies that for any $\exists_1$-formula $\exists x\psi(x)$ (where ψ is an open formula) and any theory T, if $T \vdash \exists x\psi(x)$ then there are some (Skolem) terms $t_1,\ldots,t_n$ such that $T^{\rm Sk} \vdash \psi(t_1)\vee\cdots\vee\psi(t_n)$. Usually this observation is called Herbrand's Theorem. We will need a somehow dual of this fact.

Lemma 3 (Herbrand Proof of Universal Formulas) *For a $\forall_1$-formula $\forall x\psi(x)$ (where ψ is open) and a theory T, suppose $T\vdash\forall x\psi(x)$. There exists a finite (standard) $k\geqslant 0$ such that for any set of terms Λ, any T-evaluation p on $\Lambda^{\langle k\rangle}$ and any $t\in\Lambda$, we have $p\models\psi(t)$.*

Proof By $T\vdash\forall x\psi(x)$ the theory $T^{\rm Sk}\cup\{\neg\psi(\mathfrak{c})\}$, where $\mathfrak{c}$ is the Skolem constant symbol for $\exists x\neg\psi(x)$, is inconsistent. Suppose φ is the rectified negation normal form of $\neg\psi$. Then, by Herbrand's theorem, there exists some finite set of terms Γ such that there can be no $(T^{\rm Sk}\cup\{\varphi(\mathfrak{c})\})$-evaluation on it. Since $\mathfrak{c}$ appears in Γ we write it as $\Gamma(\mathfrak{c})$, and by $\Gamma(u)$, where u is an arbitrary term, we denote the set of terms which result from the terms of $\Gamma(\mathfrak{c})$ by replacing $\mathfrak{c}$ with u everywhere. It can be clearly seen that there exists some $k\in\mathbb{N}$ such that for any set of terms Λ and any $t\in\Lambda$ we have $\Gamma(t)\subseteq\Lambda^{\langle k\rangle}$. Whence, there cannot be any $(T^{\rm Sk}\cup\{\varphi(t)\})$-evaluation on $\Lambda^{\langle k\rangle}$. Thus, any T-evaluation p on $\Lambda^{\langle k\rangle}$ must satisfy $p\not\models\varphi(t)$, or equivalently $p\models\psi(t)$. □

Example 4 Let the theory T, in the language of arithmetic $\mathscr{L}_A$, be axiomatized by

(1) $\forall x[\mathsf{S}(x)\neq\mathsf{0}]$	(2) $\forall x,y[x+\mathsf{S}(y)=\mathsf{S}(x+y)]$
(3) $\forall x\exists z[x\neq\mathsf{0}\rightarrow x=\mathsf{S}(z)]$	(4) $\forall x,y\exists z[x\leqslant y\rightarrow z+x=y]$

For the open formula $\psi(x)=(x\leqslant\mathsf{0}\rightarrow x=\mathsf{0})$ we have $T\vdash\forall x\psi(x)$.

Let $\mathfrak{p}(x)$ be the Skolem function for the formula $\exists z[x=\mathsf{0}\vee x=\mathsf{S}(z)]$, and $\mathfrak{h}(x,y)$ be the Skolem function for the formula $\exists z[x\not\leqslant y\vee z+x=y]$. Then the Skolemized form $T^{\rm Sk}$ of the theory T will be as:

(1′) $\mathsf{S}(x)\neq\mathsf{0}$	(2′) $x+\mathsf{S}(y)=\mathsf{S}(x+y)$
(3′) $x=\mathsf{0}\vee x=\mathsf{S}(\mathfrak{p}(x))$	(4′) $x\not\leqslant y\vee\mathfrak{h}(x,y)+x=y$

For a fixed term t let Γ_t be the following set of terms:

$$\{0, t, \mathfrak{h}(t, 0), \mathfrak{h}(t, 0)+t, \mathfrak{p}(t), \mathsf{S}(\mathfrak{p}(t)), \mathfrak{h}(t, 0)+\mathfrak{p}(t), \mathfrak{h}(t, 0)+\mathsf{S}(\mathfrak{p}(t)), \mathsf{S}\big(\mathfrak{h}(t, 0)+\mathfrak{p}(t)\big)\}.$$

Now we show that any T-evaluation p on Γ_t must satisfy $p \models \psi(t)$ or, equivalently, if $p \models t \leqslant 0$ then $p \models t = 0$. Assume $p \models t \leqslant 0$. Then by the fourth axiom we should have $p \models \mathfrak{h}(t, 0)+t = 0$. If $p \models t = 0$ does not hold, then $p \models t \neq 0$, so by the third axiom we have $p \models t = \mathsf{S}(\mathfrak{p}(t))$. Whence, $p \models \mathfrak{h}(t, 0) + \mathsf{S}(\mathfrak{p}(t)) = 0$. On the other hand, by the second axiom, $p \models \mathfrak{h}(t, 0) + \mathsf{S}(\mathfrak{p}(t)) = \mathsf{S}\big(\mathfrak{h}(t, 0) + \mathfrak{p}(t)\big)$. Now we can infer that $p \models \mathsf{S}\big(\mathfrak{h}(t, 0) + \mathfrak{p}(t)\big) = 0$, which is in contradiction with the first axiom. Thus, $p \models t = 0$ must hold, which shows that $p \models \psi(t)$. $\diamond$

As was mentioned before, for a consistent theory T there must exist some Herbrand Model of T.

Definition 9 (*Definable Herbrand Models*) Let Λ be a set of terms, and define its Skolem Hull to be $\Lambda^{\langle\infty\rangle} = \bigcup_{n\in\mathbb{N}} \Lambda^{\langle n\rangle}$ (see Definition 8). Suppose p is an evaluation on $\Lambda^{\langle\infty\rangle}$. Define $\mathfrak{M}(\Lambda, p) = \{t/p \mid t \in \Lambda^{\langle\infty\rangle}\}$, where t/p is the equivalence class of the relation $\approx_p$ containing t (cf. Lemma 1). Put the structure on it by

(1) $f^{\mathfrak{M}(\Lambda,p)}(t_1/p, \ldots, t_m/p) = f(t_1, \ldots, t_m)/p$, and
(2) $R^{\mathfrak{M}(\Lambda,p)} = \{(t_1/p, \ldots, t_m/p) \mid p \models R(t_1, \ldots, t_m)\}$,

for any m-ary function symbol f and any m-ary relation symbol R. $\diamond$

Lemma 4 (Herbrand Models by Evaluations) *The structure on $\mathfrak{M}(\Lambda, p)$ is well-defined, and for a theory T, if p is an T-evaluation on Λ then $\mathfrak{M}(\Lambda, p) \models T$.*

3 Bounded arithmetic and Herbrand consistency

By an efficient Gödel coding (see e.g. Chapter V of [6]) we can code sets, sequences (and so the syntactic concepts like Skolem function symbols, Skolem instances, evaluations, etc.) such that the following [6] hold for any sequences α, β:

- $\ulcorner\alpha * \beta\urcorner \leqslant 64 \cdot (\ulcorner\alpha\urcorner \cdot \ulcorner\beta\urcorner)$, where $*$ denotes concatenation;
- $|\alpha| \leqslant \log(\ulcorner\alpha\urcorner)$.

It follows that for any sets A, B we have $\ulcorner A \cup B\urcorner \leqslant 64 \cdot (\ulcorner A\urcorner \cdot \ulcorner B\urcorner)$ and $|A| \leqslant \log(\ulcorner A\urcorner)$. We write $X \in \mathscr{O}(Y)$ to indicate that $X \leqslant Y \cdot n + n$ for some $n \in \mathbb{N}$; that is X is linearly bounded by Y. The above (efficient) coding has the property that for any sequence $U = \langle u_1, \ldots, u_l\rangle$ we have $\log(\ulcorner U\urcorner) \in \mathscr{O}(\sum_i \log(\ulcorner u_i\urcorner))$. For any evaluation p on a set of terms Λ it can be seen that $\log(\ulcorner p\urcorner) \in \mathscr{O}(\log(\ulcorner \Lambda\urcorner))$.

Let us note that all of the concepts introduced so far can be formalized in the language of arithmetic $\mathscr{L}_A$. Here we make the observation that, having an arithmetically definable set of terms Λ, the sets $\Lambda^{\langle j\rangle}$ are all definable in arithmetic (in terms of Λ and j), but the set $\Lambda^{\langle\infty\rangle}$ is not definable by an arithmetical formula. We will come to this point later. The arithmetical theory we are interested here is denoted by $\mathrm{I}\Delta_0$ which is usually axiomatized by Robinson's arithmetic, in the language $\mathscr{L}_A$, plus the induction axiom for bounded formulas (see e.g. [6]).

In this section we prove our main result: the existence of a finite fragment $T \subseteq \mathrm{I}\Delta_0$ whose Herbrand Consistency is not provable in $\mathrm{I}\Delta_0$. As the exponential function $x \mapsto 2^x$ is not available (provably total) in $\mathrm{I}\Delta_0$, we denote by $\mathfrak{log}$ the set of elements x for which $\exp(x) = 2^x$ exists. Let us note that for a model $\mathscr{M}$, the set $\mathfrak{log}(\mathscr{M})$ is the logarithm of the elements of $\mathscr{M}$. The set $\mathfrak{log}$ is closed under S and $+$, but not under $\times$, in $\mathrm{I}\Delta_0$. We will use the term *cut* for any definable and downward closed set (not necessarily closed under S) in the arithmetical models. The formula "$y = \exp(x)$" is expressible in $\mathscr{L}_A$ by a bounded formula, and $\mathrm{I}\Delta_0$ can prove some of the basic properties of exp (cf. [6]), though cannot prove its totality: $\mathrm{I}\Delta_0 \nvdash \forall x \exists y [y = \exp(x)]$. By $\mathfrak{log}^2$ we denote the set of elements x for which $\exp^2(x) = 2^{2^x}$ exists; the superscripts on top of the functions denote the iteration. Similarly, $\mathfrak{log}^n = \{x \mid \exists y [y = \exp^n(x)]\}$, where $\exp^n$ denotes the n time iteration of the exponential function exp.

We use a deep theorem in bounded arithmetic, which happens to be the very last theorem of [6]. It reads, in our terminology, as:

For any $k \geqslant 0$ *there exists a bounded formula* $\varphi(x)$ *such that*
$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x \in \mathfrak{log}^{k+1} \varphi(x)$, *but* $\mathrm{I}\Delta_0 + \Omega_1 \nvdash \forall x \in \mathfrak{log}^{k} \varphi(x)$.

It can be clearly seen that the theorem also holds for $\mathrm{I}\Delta_0$ instead of $\mathrm{I}\Delta_0 + \Omega_1$, and for any cut I (and its logarithm $\log I = \{x \mid \exists y \in I [y = \exp(x)]\}$) instead of $\mathfrak{log}^k$ (and its logarithm $\mathfrak{log}^{k+1}$); see also [3] and (Theorem 3.6 of) [11].

Theorem 2 (Π_1-Separation of Logarithmic Cuts) *For any cut* I *there exists a bounded formula* $\varphi(x)$ *such that the theory* $\mathrm{I}\Delta_0 \cup \{\exists x \in I\ \varphi(x)\}$ *is consistent, but the theory* $\mathrm{I}\Delta_0 \cup \{\exists x \in \log I\ \varphi(x)\}$ *is not consistent.*

We will find the desired finite fragment of $\mathrm{I}\Delta_0$ (whose Herbrand Consistency is not provable in $\mathrm{I}\Delta_0$) in three steps (the following subsections) before proving the main result (in the last subsection). For doing so, we will show that for sufficiently strong finite fragments of $\mathrm{I}\Delta_0$, like T, if $\mathrm{I}\Delta_0 \vdash \mathrm{HCon}(T)$ then the consistency of the theory $\mathrm{I}\Delta_0 \cup \{\exists x \in I\ \theta(x)\}$, for some suitable cut I and a suitable bounded formula θ, implies the consistency of the theory $T \cup \{\exists x \in \log I\ \theta(x)\}$. And this, as we will see, contradicts Theorem 2.

3.1 The first finite fragment

Assuming the consistency of $\mathrm{I}\Delta_0 \cup \{\exists x \in I\ \varphi(x),\ \mathrm{HCon}(T)\}$, and inconsistency of the theory $T \cup \{\exists x \in \log I\ \varphi(x)\}$, we can construct a model $\mathfrak{M}$, from a given model $\mathscr{M} \models \mathrm{I}\Delta_0 \cup \{\exists x \in I\ \varphi(x),\ \mathrm{HCon}(T)\}$, such that $\mathfrak{M} \models T \cup \{\exists x \in \log I\ \varphi(x)\}$; which is in contradiction with the assumptions. For that, let us take a (hypothetical) model $\mathscr{M} \models \mathrm{I}\Delta_0 \cup \{a \in I \wedge \varphi(a)\} \cup \{\mathrm{HCon}(T)\}$ for some $a \in \mathscr{M}$. Then we form the set $\Gamma = \{\underline{0}, \underline{1}, \underline{2}, \ldots \underline{\omega_1(a)}\}$ where $\underline{i}$ is a term in $\mathscr{L}_A$ representing the number i, defined inductively as $\underline{0} = \mathsf{0}$ and $\underline{i+1} = \mathsf{S}(\underline{i})$. Let us note that for sufficiently small elements $i \in \mathscr{M}$ (a code for) the term $\underline{i}$ may exist; in fact for the i's in the cut $\mathscr{I}$ (Definition 10 below) always $\underline{i}$ exists in $\mathscr{M}$. From the assumption $\mathscr{M} \models \mathrm{HCon}(T)$ we find an T-evaluation p on $\Lambda^{\langle j \rangle}$, for a suitable j and a suitable Λ which contains the above set Γ. Then we can form the model $\mathfrak{M}(\Lambda, p)$ and, by some technical details, show that $\mathfrak{M}(\Lambda, p) \models T + \exists x \in \log I\varphi(x)$. The bound $\omega_1(a)$ assures us that the set Γ contains the range of (the bounded) quantifiers in the (bounded) formula $\varphi(a)$. For the Gödel code of $\underline{i}$ we have $\log(\ulcorner \underline{i} \urcorner) \in \mathscr{O}(\log(2^i))$ and so $\log(\ulcorner \Gamma \urcorner) \in \mathscr{O}(\log(2^{(\omega_1(a))^2}))$ whence $\log(\ulcorner \Gamma \urcorner) \in \mathscr{O}\big(\log\big(\exp^2(2(\log a)^2)\big)\big)$. We need the closure of Γ under the Skolem function symbols of (a finite fragment of) $\mathrm{I}\Delta_0$, that is $\Gamma^{\langle\infty\rangle}$ (see Definitions 9 and 8). Since, unfortunately, that set is not definable, we consider the set $\Gamma^{\langle j\rangle}$ for a non-standard j, which makes sense if $\ulcorner \Gamma \urcorner$ (and so a) is non-standard. In case a is standard, then the proof becomes trivial (see below). For some non-standard j with $j \leqslant \log^4(\ulcorner \Gamma \urcorner)$ we can form the set $\Gamma^{\langle j\rangle}$, in case $\omega_2(\ulcorner \Gamma \urcorner)$ exists (see [11,12]). And finally we have $\log\big(\omega_2(\ulcorner \Gamma \urcorner)\big) \in \mathscr{O}\big(\log\big(\exp^2(4(\log a)^4)\big)\big)$.

Definition 10 (*The Cut $\mathscr{I}$*) The cut $\mathscr{I}$ is defined to be $\{x \mid \exists y[y = \exp^2(4(\log a)^4)]\}$, and its logarithm is $\log \mathscr{I} = \{x \mid \exists y[y = \exp^2(4a^4)]\}$. ◇

Applying Theorem 2 to the cut $\mathscr{I}$ defined above, we find a (fixed) bounded formula θ and a finite fragment $T_0 \subseteq \mathrm{I}\Delta_0$ such that the theory the theory $\mathrm{I}\Delta_0 \cup \{\exists x \in \mathscr{I}\theta(x)\}$ is consistent, but the theory $T_0 \cup \{\exists x \in \log \mathscr{I}\theta(x)\}$ is not consistent.

Definition 11 (*The First Fragment T_0*) Let T_0 be a finite fragment of $\mathrm{I}\Delta_0$ for which there exists a (fixed) bounded formula θ such that the theory $\mathrm{I}\Delta_0 \cup \{\exists x \in \mathscr{I}\theta(x)\}$ is consistent, but the theory $T_0 \cup \{\exists x \in \log \mathscr{I}\theta(x)\}$ is not consistent. Let $\mathscr{M}$ be a (fixed) model such that $\mathscr{M} \models \mathrm{I}\Delta_0 \cup \{\exists x \in \mathscr{I}\theta(x)\}$. ◇

In the rest of the paper we will show that for a finite fragment T of $\mathrm{I}\Delta_0$ extending T_0 we have $\mathscr{M} \not\models \mathrm{HCon}(T)$, where HCon is the predicate of Herbrand Consistency.

3.2 The second finite fragment

The proof of the main result goes roughly as follows: if $\mathscr{M} \models \mathrm{HCon}(T)$, for a finite fragment $T \subseteq \mathrm{I}\Delta_0$ to be specified later, then there exists (in $\mathscr{M}$) some T-evaluation p on some $\Lambda^{\langle j\rangle}$, where $\Lambda \supseteq \Gamma$ is to be specified later and Γ and j are as in the previous subsection. Whence we can form the model $\mathfrak{M}(\Lambda, p)$, for which we already have $\mathfrak{M}(\Lambda, p) \models T$ by Lemma 4. Our second finite fragment T_1 will have the property that if $T \supseteq T_1$ then $\mathfrak{M}(\Lambda, p) \models \theta_0(\underline{a}/p)$. The third finite fragment T_2 will have the property that if $T \supseteq T_2$ then we have $\mathfrak{M}(\Lambda, p) \models \underline{a}/p \in \log \mathscr{I}$. So, finally we will get the model $\mathfrak{M}(\Lambda, p)$ which satisfies $\mathfrak{M}(\Lambda, p) \models T + [\underline{a}/p \in \log \mathscr{I} \wedge \theta_0(\underline{a}/p)]$, or, in the other words, $\mathfrak{M}(\Lambda, p) \models T \cup \{\exists x \in \log \mathscr{I}\theta_0(x)\}$ which is in contradiction with (the choice of the first finite fragment) $T_0 \subseteq T$.

Definition 12 (*The Second Fragment T_1*) Let T_1 be a finite fragment of $\mathrm{I}\Delta_0$ which can prove the following ($\mathrm{I}\Delta_0$-provable $\forall^*$-)sentences:

- $x + \mathsf{0} = x$
- $x + \mathsf{S}(y) = \mathsf{S}(x + y)$
- $x \cdot \mathsf{0} = \mathsf{0}$
- $x \cdot \mathsf{S}(y) = x \cdot y + x$
- $x \leqslant \mathsf{0} \leftrightarrow x = \mathsf{0}$
- $x \leqslant \mathsf{S}(y) \leftrightarrow x = \mathsf{S}(y) \vee x \leqslant y$
- $x \leqslant y \vee y \leqslant x$
- $x \leqslant y \leqslant z \rightarrow x \leqslant z$
- $x \leqslant z + x$
- $x \leqslant x + z$
- $x + z \leqslant y + z \rightarrow x \leqslant y$
- $z \neq \mathsf{0} \wedge x \cdot z \leqslant y \cdot z \rightarrow x \leqslant y$
- $x \neq y \leftrightarrow \mathsf{S}(x) \leqslant y \vee \mathsf{S}(y) \leqslant x$
- $x \not\leqslant y \leftrightarrow \mathsf{S}(y) \leqslant x$

and also can prove the following ($\mathrm{I}\Delta_0$-provable $\forall^*\exists^*$-)sentences:

- $x \leqslant y \to \exists z[z + x = y]$
- $y \neq \mathsf{0} \to \exists q, r[x = r + q \cdot y \wedge r \leqslant y]$ ◇

Remark 1 It can be seen that T_1 can prove the following arithmetical sentences:

- $\mathsf{S}(x) \neq \mathsf{0}$
- $\mathsf{S}(x) = \mathsf{S}(y) \to x = y$
- $\mathsf{S}(x) \nleqslant x$
- $x \neq \mathsf{0} \to \exists y[x = \mathsf{S}(y)]$

For a proof, first note that by $x \leqslant y \vee y \leqslant x$ we have $\forall u[u \leqslant u]$, and also from $x \leqslant z + x$ and $x + \mathsf{0} = x$ we get $\forall u[\mathsf{0} \leqslant u]$. Now, if $\mathsf{S}(u) = \mathsf{0}$, then $\mathsf{S}(u) \leqslant \mathsf{0}$, and so by the axiom $x \nleqslant y \leftrightarrow \mathsf{S}(y) \leqslant x$ we get $\mathsf{0} \nleqslant u$, contradiction! Also from the same axiom it follows that $u \nleqslant u \leftrightarrow \mathsf{S}(u) \leqslant u$, and thus $\mathsf{S}(u) \nleqslant u$. If $\mathsf{S}(u) = \mathsf{S}(v)$ and $u \neq v$ then by $x \neq y \leftrightarrow \mathsf{S}(x) \leqslant y \vee \mathsf{S}(y) \leqslant x$ we have either $\mathsf{S}(u) \leqslant v$ or $\mathsf{S}(v) \leqslant u$. If $\mathsf{S}(u) \leqslant v$ then $\mathsf{S}(v) \leqslant v$, contradiction! The other case is similar. Finally, assume $u \neq \mathsf{0}$. Then by $x \leqslant \mathsf{0} \leftrightarrow x = \mathsf{0}$ we have $u \nleqslant \mathsf{0}$ and so the axiom $x \nleqslant y \leftrightarrow \mathsf{S}(y) \leqslant x$ implies that $\mathsf{S}(\mathsf{0}) \leqslant u$. Thus, by $x \leqslant y \to \exists z[z + x = y]$ we have $v + \mathsf{S}(\mathsf{0}) = u$ for some v. Then from $x + \mathsf{S}(y) = \mathsf{S}(x + y)$ and $x + \mathsf{0} = x$ we conclude that $\mathsf{S}(v) = u$. □

The main property of T_1 is the following:

Theorem 3 (The Main Property of T_1) *Suppose $\mathscr{M}$ is a non-standard model such that $\mathscr{M} \models \mathrm{I}\Delta_0 + [a \in \mathscr{I} \wedge \theta(a)] + \mathrm{HCon}(T)$ where θ is a bounded formula and $a \in \mathscr{M}$ is non-standard and $T \vdash T_1$. If $p \in \mathscr{M}$ is an T-evaluation on $\Lambda^{\langle j \rangle}$ where Λ is a set of terms such that $\Lambda \supseteq \Gamma = \{\underline{i} \mid i \leqslant \omega_1(a)\}$ and j is a non-standard element of $\mathscr{M}$, then for any bounded formula $\varphi(x_1, \ldots, x_n)$ and any elements $i_1, \ldots, i_n \leqslant a$ in $\mathscr{M}$, $\mathscr{M} \models \varphi(i_1, \ldots, i_n) \iff \mathfrak{M}(\Lambda, p) \models \varphi(\underline{i_1}/p, \ldots, \underline{i_n}/p)$.*

We prove the theorem by induction on (the complexity) of φ (see also [11,12]).

Lemma 5 (Another Property of T_1) *Suppose $\mathscr{K} \models T_1$ and $a \in \mathscr{K}$, and let $\mathtt{t}$ be an $\mathscr{L}_A$-term. For any $i_1, \ldots, i_n \leqslant a$ in $\mathscr{K}$ and $b \in \mathscr{K}$, if $\mathscr{K} \models b \leqslant \mathtt{t}(i_1, \ldots, i_n)$ then there exist an $\mathscr{L}_A$-term $\mathtt{s}$ and some $j_1, \ldots, j_m \leqslant a$ in $\mathscr{K}$ such that $\mathscr{K} \models b = \mathtt{s}(j_1, \ldots, j_m)$.*

Proof By induction on $\mathtt{t}$ (for simplicity we omit $(i_1, \ldots, i_n)$ from $\mathtt{t}(i_1, \ldots, i_n)$):

– $\mathtt{t} = \mathsf{0}$: if $\mathscr{K} \models b \leqslant \mathsf{0}$ then by the T_1-axiom $x \leqslant \mathsf{0} \leftrightarrow x = \mathsf{0}$ we have $\mathscr{K} \models b = \mathsf{0}$.
– $\mathtt{t} = \mathsf{S}(\mathtt{t}_1)$: if $\mathscr{K} \models b \leqslant \mathsf{S}(\mathtt{t}_1)$ then by $x \leqslant \mathsf{S}(y) \leftrightarrow x = \mathsf{S}(y) \vee x \leqslant y$ which is a T_1-axiom, we have $\mathscr{K} \models b = \mathsf{S}(\mathtt{t}_1) \vee b \leqslant \mathtt{t}_1$, and the result follows from the induction hypothesis.
– $\mathtt{t} = \mathtt{t}_1 + \mathtt{t}_2$: if $\mathscr{K} \models b \leqslant \mathtt{t}_1 + \mathtt{t}_2$ then by the T_1-axiom $x \leqslant y \vee y \leqslant x$ we have that $\mathscr{K} \models b \leqslant \mathtt{t}_2 \vee \mathtt{t}_2 \leqslant b$. If $\mathscr{K} \models b \leqslant \mathtt{t}_2$ then the conclusion follows from the induction hypothesis. Otherwise if $\mathscr{K} \models \mathtt{t}_2 \leqslant b$ then by $x \leqslant y \to \exists z[z + x = y]$ (another T_1-axiom) there exists some $d \in \mathscr{K}$ such that $\mathscr{K} \models d + \mathtt{t}_2 = b$. Thus $\mathscr{K} \models d + \mathtt{t}_2 \leqslant \mathtt{t}_1 + \mathtt{t}_2$, whence by the T_1-axiom $x + z \leqslant y + z \to x \leqslant y$ we have $\mathscr{K} \models d \leqslant \mathtt{t}_1$, and the desired result follows from the induction hypothesis and the fact that $\mathscr{K} \models b = d + \mathtt{t}_2$.

– $\mathtt{t} = \mathtt{t}_1 \cdot \mathtt{t}_2$: assume $\mathscr{K} \models b \leqslant \mathtt{t}_1 \cdot \mathtt{t}_2$. If $\mathscr{K} \models \mathtt{t}_2 = \mathsf{0}$ then $\mathscr{K} \models \mathtt{t}_1 \cdot \mathtt{t}_2 = \mathsf{0}$ by the T_1-axiom $x \cdot \mathsf{0} = \mathsf{0}$. And so $\mathscr{K} \models b \leqslant \mathsf{0}$ is reduced to the first case above. Suppose $\mathscr{K} \models \mathtt{t}_2 \neq \mathsf{0}$. Then by the T_1-axiom $y \neq \mathsf{0} \to \exists q, r[x = r + q \cdot y \wedge r \leqslant y]$ we have $\mathscr{K} \models b = r + q \cdot \mathtt{t}_2 \wedge r \leqslant \mathtt{t}_2$ for some $q, r \in \mathscr{K}$. By the T_1-axiom $x \leqslant z + x$ we have $\mathscr{K} \models q \cdot \mathtt{t}_2 \leqslant r + q \cdot \mathtt{t}_2 = b \leqslant \mathtt{t}_1 \cdot \mathtt{t}_2$, and then using the T_1-axiom $x \leqslant y \leqslant z \to x \leqslant z$ one can infer that $\mathscr{K} \models q \cdot \mathtt{t}_2 \leqslant \mathtt{t}_1 \cdot \mathtt{t}_2$, and finally the T_1-axiom $z \neq \mathsf{0} \wedge x \cdot z \leqslant y \cdot z \to x \leqslant y$ implies that $\mathscr{K} \models q \leqslant \mathtt{t}_1$ (since $\mathscr{K} \models \mathtt{t}_2 \neq \mathsf{0}$). Now, the desired conclusion follows from the induction hypothesis and the fact that $\mathscr{K} \models b = r + q \cdot \mathtt{t}_2 \wedge r \leqslant \mathtt{t}_2 \wedge q \leqslant \mathtt{t}_1$. □

Lemma 6 (Preservation of Atomic Formulas) *With the assumptions of Theorem 3, for any atomic formula $\varphi(x_1, \ldots, x_n)$ and any $i_1, \ldots, i_n \leqslant a$ we have that*

$$\mathscr{M} \models \varphi(i_1, \ldots, i_n) \iff \mathfrak{M}(\Lambda, p) \models \varphi(\underline{i_1}/p, \ldots, \underline{i_n}/p).$$

Proof By the T_1-axioms $x \neq y \leftrightarrow \mathsf{S}(x) \leqslant y \vee \mathsf{S}(y) \leqslant x$ and $x \nleqslant y \leftrightarrow \mathsf{S}(y) \leqslant x$ it suffices to prove the one direction only:

$$\mathscr{M} \models \varphi(i_1, \ldots, i_n) \Longrightarrow \mathfrak{M}(\Lambda, p) \models \varphi(\underline{i_1}/p, \ldots, \underline{i_n}/p).$$

If $\varphi(i_1, \ldots, i_n) =$ "$\mathtt{t}(i_1, \ldots, i_n) \leqslant \mathtt{s}(i_1, \ldots, i_n)$" for some $\mathscr{L}_A$-terms $\mathtt{t}$ and $\mathtt{s}$, then $\mathscr{M} \models \mathtt{t}(i_1, \ldots, i_n) \leqslant \mathtt{s}(i_1, \ldots, i_n)$ implies the existence of some $b \in \mathscr{M}$ such that $\mathscr{M} \models b + \mathtt{t}(i_1, \ldots, i_n) = \mathtt{s}(i_1, \ldots, i_n)$. By the T_1-axiom $x \leqslant x + z$ we have $\mathscr{M} \models b \leqslant \mathtt{s}$, so by Lemma 5 there exist an $\mathscr{L}_A$-term $\mathtt{r}$ and some elements $j_1, \ldots, j_m \leqslant a$ such that $\mathscr{K} \models b = \mathtt{r}(j_1, \ldots, j_m)$. Whence,

$$\mathscr{M} \models \mathtt{r}(j_1, \ldots, j_m) + \mathtt{t}(i_1, \ldots, i_n) = \mathtt{s}(i_1, \ldots, i_n).$$

So, noting that $\mathscr{M}, \mathfrak{M}(\Lambda, p) \models T_1$, it suffices to prove the lemma for the atomic formula $\varphi(i_1, \ldots, i_n)$ of the form $\varphi(i_1, \ldots, i_n) =$ "$\mathtt{t}(i_1, \ldots, i_n) = \mathtt{s}(i_1, \ldots, i_n)$".

For that we note that if $i_1, \ldots, i_n \leqslant a$ then $\mathtt{t}(i_1, \ldots, i_n), \mathtt{s}(i_1, \ldots, i_n) \leqslant \omega_1(a)$. Suppose $\mathscr{M} \models \mathtt{t}(i_1, \ldots, i_n) = \mathtt{s}(i_1, \ldots, i_n) = i$. We show by induction on (the complexity of) $\mathtt{t}$ that

$$\mathscr{M} \models \mathtt{t}(i_1, \ldots, i_n) = i \Longrightarrow \mathfrak{M}(\Lambda, p) \models \mathtt{t}(\underline{i_1}/p, \ldots, \underline{i_n}/p) = \underline{i}/p.$$

Note that the condition $\mathfrak{M}(\Lambda, p) \models \mathtt{t}(\underline{i_1}/p, \ldots, \underline{i_n}/p) = \underline{i}/p$ is equivalent to the condition $\mathscr{M} \models$ "$p \models \mathtt{t}(\underline{i_1}, \ldots, \underline{i_n}) = \underline{i}$". So, it suffices to show the following equivalence by induction on $\mathtt{t}$:

$$\mathscr{M} \models \mathtt{t}(i_1, \ldots, i_n) = i \longleftrightarrow \text{“}p \models \mathtt{t}(\underline{i_1}, \ldots, \underline{i_n}) = \underline{i}\text{”}.$$

For $\mathtt{t} = \mathsf{0}$ and $\mathtt{t} = \mathsf{S}(\mathtt{t}_1)$ the result follows from the definition $\underline{0} = \mathsf{0}$ and $\underline{j+1} = \mathsf{S}(\underline{j})$. And for $\mathtt{t} = \mathtt{t}_1 + \mathtt{t}_2$ and $\mathtt{t} = \mathtt{t}_1 \cdot \mathtt{t}_2$ the result follows from the T_1-axioms $x + \mathsf{0} = x$, $x + \mathsf{S}(y) = \mathsf{S}(x + y)$, $x \cdot \mathsf{0} = \mathsf{0}$, and $x \cdot \mathsf{S}(y) = x \cdot y + x$. □

Hence, the lemma also holds for open formulas φ as well. For bounded formulas we note that the range of quantifiers of $\varphi(i_1, \dots, i_n)$ for $i_1, \dots, i_n \leqslant a$ is contained in the set $\{j \mid j \leqslant \omega_1(a)\}$. This is formally expressed in the following lemma.

Lemma 7 (End-Extension Property) *With the assumptions of Theorem 3, if for some $i \leqslant a$ and some term u we have $(\mathscr{M} \models) p \models u \leqslant \underline{i}$, then there exists some $j \leqslant i$ such that $(\mathscr{M} \models) p \models u = \underline{j}$.*

Proof By induction on the term $\underline{i}$. For $i = 0$, if $p \models u \leqslant 0$ then by Lemma 3, and the T_1-axiom $x \leqslant 0 \leftrightarrow x = 0$, we have $p \models u = 0 = \underline{0}$. For $\underline{i} = \mathsf{S}(\underline{j})$, if $p \models u \leqslant \mathsf{S}(\underline{j})$ then by Lemma 3, and the T_1-axiom $x \leqslant \mathsf{S}(y) \leftrightarrow x = \mathsf{S}(y) \vee x \leqslant y$, we must have that $p \models u = \mathsf{S}(\underline{j}) \vee u \leqslant \underline{j}$. Now the conclusion follows from the induction hypothesis. □

Now we can prove Theorem 3.

Proof (of Theorem 3) By induction on (the complexity of the bounded formula) φ. As the theorem has been proved for open formulas (Lemma 6), it suffices to show that if the theorem holds for the (bounded) formula φ then it also holds for the (bounded) formula $\exists x \leqslant \mathtt{t}(i_1, \dots, i_n)\varphi(x, i_1, \dots, i_n)$ where $\mathtt{t}$ is an $\mathscr{L}_A$-term; in the other words:

$$\mathscr{M} \models \exists x \leqslant \mathtt{t}(i_1, \dots, i_n)\varphi(x, i_1, \dots, i_n) \iff \mathfrak{M}(\Lambda, p) \models \exists x \leqslant \mathtt{t}(\underline{i_1}/p, \dots, \underline{i_n}/p)\varphi(\underline{i_1}/p, \dots, \underline{i_n}/p).$$

If $\mathscr{M} \models b \leqslant \mathtt{t}(i_1, \dots, i_n) \wedge \varphi(b, i_1, \dots, i_n)$, for some $b \in \mathscr{M}$, then by Lemma 5 there are some $\mathscr{L}_A$-term $\mathtt{s}$ and some elements $j_1, \dots, j_m \leqslant a$ in $\mathscr{M}$ such that $\mathscr{M} \models b = \mathtt{s}(j_1, \dots, j_m)$. So, we have $\mathscr{M} \models \varphi(\mathtt{s}(j_1, \dots, j_m), i_1, \dots, i_n)$. Whence, by the induction hypothesis we also have

$$\mathfrak{M}(\Lambda, p) \models \varphi(\mathtt{s}(\underline{j_1}/p, \dots, \underline{j_m}/p), \underline{i_1}/p, \dots, \underline{i_n}/p),$$

thus, noting that we already have

$$\mathfrak{M}(\Lambda, p) \models \mathtt{s}(\underline{j_1}/p, \dots, \underline{j_m}/p) \leqslant \mathtt{t}(\underline{i_1}/p, \dots, \underline{i_n}/p),$$

the desired conclusion holds:

$$\mathfrak{M}(\Lambda, p) \models \exists x \leqslant \mathtt{t}(\underline{i_1}/p, \dots, \underline{i_n}/p)\varphi(\underline{i_1}/p, \dots, \underline{i_n}/p).$$

Conversely, if $\mathfrak{M}(\Lambda, p) \models d \leqslant \mathtt{t}(\underline{i_1}/p, \dots, \underline{i_n}/p) \wedge \varphi(d, \underline{i_1}/p, \dots, \underline{i_n}/p)$ holds for some $d \in \mathfrak{M}(\Lambda, p)$, then by Lemma 5 there exist an $\mathscr{L}_A$-term $\mathtt{s}$ and some elements $l_1, \dots, l_m \leqslant \underline{a}/p$ such that $\mathfrak{M}(\Lambda, p) \models d = \mathtt{s}(l_1, \dots, l_m)$. For each $\alpha \leqslant m$ there is some term $\ell_\alpha \in \Lambda^{\langle\infty\rangle}$ such that $l_\alpha = \ell_\alpha/p$. For each such α we have $\mathfrak{M}(\Lambda, p) \models \ell_\alpha/p \leqslant \underline{a}/p$ or equivalently $\mathscr{M} \models$ "$p \models \ell_\alpha \leqslant \underline{a}$". So, by Lemma 7 there exists some $j_\alpha \leqslant a$ for which we have $\mathscr{M} \models \ell_\alpha = \underline{j_\alpha}$. Whence,

$\mathfrak{M}(\Lambda, p) \models d = \mathtt{s}(\underline{j_1}/p, \ldots, \underline{j_m}/p)$ and so

$$\mathfrak{M}(\Lambda, p) \models \mathtt{s}(\underline{j_1}/p, \ldots, \underline{j_m}/p) \leqslant \mathtt{t}(\underline{i_1}/p, \ldots, \underline{i_n}/p), \text{ and}$$
$$\mathfrak{M}(\Lambda, p) \models \varphi(\mathtt{s}(\underline{j_1}/p, \ldots, \underline{j_m}/p), \underline{i_1}/p, \ldots, \underline{i_n}/p)$$

Thus, by the induction hypothesis we have

$$\mathscr{M} \models \mathtt{s}(j_1, \ldots, j_m) \leqslant \mathtt{t}(i_1, \ldots, i_n), \quad \text{and} \quad \mathscr{M} \models \varphi(\mathtt{s}(j_1, \ldots, j_m), i_1, \ldots, i_n).$$

So, we conclude that $\mathscr{M} \models \exists x \leqslant \mathtt{t}(i_1, \ldots, i_n)\varphi(x, i_1, \ldots, i_n)$. □

Let us repeat where we are now: in looking for a finite fragment $T \subseteq \mathrm{I}\Delta_0$ such that $\mathrm{I}\Delta_0 \nvdash \mathrm{HCon}(T)$ we found a finite fragment $T_0 \subseteq \mathrm{I}\Delta_0$ and a bounded formula $\theta(x)$ such that $T_0 \vdash \neg\exists x \in \log \mathscr{I}\theta(x)$ but the theory $\mathrm{I}\Delta_0 + \exists x \in \mathscr{I}\theta(x)$ is consistent and has a model $\mathscr{M} \models \mathrm{I}\Delta_0 + [a \in \mathscr{I} \wedge \theta(a)]$. Then we aim at showing that $\mathscr{M} \not\models \mathrm{HCon}(T)$. If $\mathscr{M} \models \mathrm{HCon}(T)$ then we form the set of terms $\Gamma = \{\underline{i} \mid i \leqslant \omega_1(a)\}$ for which $\omega_2(\ulcorner\Gamma\urcorner)$ exists (by the very definition of $\mathscr{I}$ and the assumption that $a \in \mathscr{I}$), and so we can form the model $\mathfrak{M}(\Gamma, p)$ where p is an T-evaluation on $\Gamma^{\langle j \rangle}$ (where $j \leqslant \log^4(\ulcorner\Gamma\urcorner)$ can be taken to be non-standard if a is so). The theory T_1 had the property that $\mathfrak{M}(\Gamma, p) \models \theta(\underline{a}/p)$ (by Theorem 3), and in the next subsection we introduce a finite fragment $T_2 \subseteq \mathrm{I}\Delta_0$ such that for a suitable $\Lambda \supseteq \Gamma$ (to be defined later) we will have $\mathfrak{M}(\Lambda, p) \models \underline{a}/p \in \log \mathscr{I}$. Then by taking T to be any finite fragment of $\mathrm{I}\Delta_0$ which extends $T_0 \cup T_1 \cup T_2$ we will conclude that $\mathscr{M} \models \neg\mathrm{HCon}(T)$.

3.3 The third finite fragment

The fragments T_0 and T_1 were chosen not by their axioms but by their implications; T_0 had to prove $\neg\exists x \in \log \mathscr{I}\theta(x)$ (Definition 11), and T_1 had to prove some certain arithmetical statements (Definition 12). But for T_2 we require that it contains one of the following sentences as (one of) its (explicit) axioms (not only its consequences).

Definition 13 (*Axioms for Totality of the Squaring Function*)

1. The induction principle for the bounded formula $\psi(x) =$ "$\exists y \leqslant x^2[y = x \cdot x]$" is denoted by $\mathrm{Ind}_\square : \ \psi(\mathsf{0}) \wedge \forall x\big(\psi(x) \rightarrow \psi(\mathsf{S}(x))\big) \rightarrow \forall x\psi(x)$. Or, in other words (cf. Examples 1,3) $\mathrm{Ind}_\square$, which is an axiom of the theory $\mathrm{I}\Delta_0$, is the sentence:

$$\exists y \leqslant \mathsf{0}^2[y = \mathsf{0} \cdot \mathsf{0}] \wedge \forall x\big(\exists y \leqslant x^2[y = x \cdot x] \\ \rightarrow \exists y \leqslant \mathsf{S}(x)^2[y = \mathsf{S}(x) \cdot \mathsf{S}(x)]\big) \Longrightarrow \forall x \exists y \leqslant x^2[y = x \cdot x].$$

2. The Π_1-sentence expressing the totality of the squaring function is denoted by $\Omega_0 : \ \forall x \exists y \leqslant x^2[y = x \cdot x]$. ◇

We denote by $\mathfrak{q}(x)$ the Skolem function symbol of the formula $\exists y \leqslant x^2[y = x \cdot x]$ (cf. Examples 1,3). Then the Skolemized forms of the axioms of Definition 13 will be

1. $[u \not\leqslant \mathsf{0}^2 \vee u \neq \mathsf{0}\cdot\mathsf{0}]\ \bigvee$
   $\big[[\mathfrak{q}(\mathfrak{c}) \leqslant \mathfrak{c}^2 \wedge \mathfrak{q}(\mathfrak{c}) = \mathfrak{c}\cdot\mathfrak{c}] \wedge [v \not\leqslant \mathsf{S}(\mathfrak{c})^2 \vee v \neq \mathsf{S}(\mathfrak{c})\cdot\mathsf{S}(\mathfrak{c})]\big]\ \bigvee$
   $[\mathfrak{q}(x) \leqslant x^2 \wedge \mathfrak{q}(x) = x\cdot x]$,

   where u, v, x are free variables and $\mathfrak{c}$ is the Skolem constant as in Example 1.
2. $\mathfrak{q}(x) \leqslant x^2 \wedge \mathfrak{q}(x) = x\cdot x$.

Define the terms q_i's by induction: $\mathsf{q}_0 = \mathsf{S}(\mathsf{S}(\mathsf{0}))$ and $\mathsf{q}_{i+1} = \mathfrak{q}(\mathsf{q}_i)$. By using a Δ_0 definition for exp (see the fourth paragraph of Sect. 3), for sufficiently small i's, it can be easily seen that q_i represents the number $\exp^2(i)$, while for the code of q_i we have $\log(\ulcorner\mathsf{q}_i\urcorner) \in \mathscr{O}\big(\log(\exp(i))\big)$. That is to say that while the value of the term q_i is of double exponential, the code of it is of (single) exponential. This (one) exponential gap, will make our proof to go through.

Formulating the statement "$x \in \mathfrak{log}^2$" can be stated as "there exists a sequence p such that $(p)_0 = 2$ and $|p| = x+1$ and for any $i < x$ we have $(p)_{i+1} = (p)_i \cdot (p)_i$". And "$y \in \log\mathscr{I}$" can be stated as "$4y^4 \in \mathfrak{log}^2$". Put $\Upsilon = \{\mathsf{q}_i \mid i \leqslant 4a^4\}$. Then any $\Omega_0(\mathrm{Ind}_\Box)$-evaluation on $\Upsilon^{\langle\infty\rangle}$ must satisfy $\mathsf{q}_{i+1} = \mathsf{q}_i\cdot\mathsf{q}_i$ for any $i < 4a^4$. If p is any such evaluation, then $\mathfrak{M}(\Upsilon, p) \models \forall i < 4(\underline{a}/p)^4[\mathsf{q}_{i+1}/p = \mathsf{q}_i/p\cdot\mathsf{q}_i/p]$. We require the finite fragment $T_2 \subseteq \mathrm{I}\Delta_0$ to have the property that for any model $\mathscr{K} \models T_2$, if there are elements $q_0, q_1, \ldots, q_b \in \mathscr{K}$ such that $\mathscr{K}$ satisfies $q_0 = 2$ and $q_{i+1} = q_i^2$ for any $i < b$, then $\mathscr{K} \models b \in \mathfrak{log}^2$. Let us note that the code of the sequence $\langle\exp^2(0), \exp^2(1), \ldots, \exp^2(b)\rangle$ is roughly bounded by

$$\prod_{i\leqslant b}\exp^2(i) \approx (\exp^2(b))^2 = \exp^2(b+1).$$

So, in the presence of $q_0, q_1, \ldots, q_b \in \mathscr{K}$ with the above property, the (code of the) sequence p in $\mathscr{M}$ with the property "$(p)_0 = 2$, $|p| = b+1$ and for any $i < b$, $(p)_{i+1} = (p)_i\cdot(p)_i$" must exist.

Note also that

$$\mathrm{I}\Delta_0 \vdash \forall i[i \in \mathfrak{log}^2 \rightarrow i+1 \in \mathfrak{log}^2]. \qquad (\circledast)$$

Definition 14 (*The Third Fragment T_2*)

1. If the usual axiomatization of $\mathrm{I}\Delta_0$ is taken into account, then let T_2 be a finite fragment of it which contains the axiom $\mathrm{Ind}_\Box$ and has the property $(\circledast)$ above. That is to say, for any model $\mathscr{K} \models T_2$, if there are elements $q_0, q_1, \ldots, q_b \in \mathscr{K}$ such that $\mathscr{K}$ satisfies $q_0 = 2$ and $q_{i+1} = q_i^2$ for any $i < b$, then $\mathscr{K} \models b \in \mathfrak{log}^2$.
2. If $\mathrm{I}\Delta_0$ has been axiomatized all by Π_1-formulas, where the induction axioms are in the form $\forall y\big(\varphi(\mathsf{0}) \wedge \forall x < y[\varphi(x) \rightarrow \varphi(\mathsf{S}(x))] \rightarrow \forall x \leqslant y\varphi(x)\big)$ for bounded φ, then we take the theory T_2 to be a finite fragment of $\mathrm{I}\dot{\Delta}_0 + \Omega_0$, where $\mathrm{I}\dot{\Delta}_0$ is the above Π_1-axiomatization of $\mathrm{I}\Delta_0$, together with the axiom Ω_0, such that it has the property $(\circledast)$ above (i.e., for any model $\mathscr{K} \models T_2$, if there are elements $q_0, q_1, \ldots, q_b \in \mathscr{K}$ such that $\mathscr{K}$ satisfies $q_0 = 2$ and $q_{i+1} = q_i^2$ for any $i < b$, then $\mathscr{K} \models b \in \mathfrak{log}^2$). So, in this case T_2 is a Π_1-theory. $\diamond$

Let us reiterate the main property of T_2 again.

The Main Property of T_2 For a model $\mathscr{K} \models T_2$, if there are $q_0, q_1, \ldots, q_b \in \mathscr{K}$ such that for any $j < b$ we have $\mathscr{K} \models q_{j+1} = q_j^2$, then $\mathscr{K} \models$"$b \in \mathfrak{log}^2$".

We note that whenever we have elements $q_0, \ldots, q_b$ (for $b \in \mathfrak{log}^2$) in a model of T_2, we also have a code for the whole sequence $\langle q_0, \ldots, q_b \rangle$ (in that model).

3.4 The proof of the main result

Let T be any finite fragment of $\mathrm{I}\Delta_0$ or $\mathrm{I}\dot{\Delta}_0 + \Omega_0$ such that $T \supseteq T_0 \cup T_1 \cup T_2$. If T_2 is taken as in the clause (1) of Definition 14 then T is truly a finite fragment of $\mathrm{I}\Delta_0$, and if T_2 is taken as in the clause (2) of Definition 14 then T is a finite $\mathrm{I}\Delta_0$-derivable Π_1-theory, whose conjunction (denoted by U) is a $\mathrm{I}\Delta_0$-derivable Π_1-sentence.

Theorem 4 (The Main Theorem)

(1) There exists a finite fragment T of $\mathrm{I}\Delta_0$ such that $\mathrm{I}\Delta_0 \nvdash \mathrm{HCon}(T)$.
(2) There exists an $\mathrm{I}\Delta_0$-derivable Π_1-sentence U such that $\mathrm{I}\Delta_0 \nvdash \mathrm{HCon}(U)$.

Proof By Theorem 2 there exists a (fixed) bounded formula $\theta(x)$, for the cut $\mathscr{I}$ defined in Definition 10, such that $\mathrm{I}\Delta_0 \nvdash \neg\exists x \in \mathscr{I}\theta(x)$ and $T_0 \vdash \neg\exists x \in \log \mathscr{I}\theta(x)$ (see Definition 11 of the theory T_0). Fix $\mathscr{M} \models \mathrm{I}\Delta_0 + [a \in \mathscr{I} \wedge \theta(a)]$. For the part (1) take T_2 as in clause (1) of Definition 14, and for part (2) take T_2 as in clause (2) of Definition 14, and let U be the conjunction of the axioms of T. In each case we will have the Skolem function symbol $\mathfrak{q}(x)$ for the squaring function $x \mapsto x^2$.

We show that $\mathscr{M} \not\models \mathrm{HCon}(T)$: Assume, for the sake of contradiction, that we have $\mathscr{M} \models \mathrm{HCon}(T)$. Define the terms $\underline{i}$'s and q_i's by induction:

$$\underline{0} = \mathsf{0}, \underline{i+1} = \mathsf{S}(\underline{i}), \mathsf{q}_0 = \underline{2}, \mathsf{q}_{i+1} = \mathfrak{q}(\mathsf{q}_i).$$

Let Λ be the set of terms $\{\underline{i} \mid i \leqslant \omega_1(a)\} \cup \{\mathsf{q}_i \mid i \leqslant \omega_1(a)\}$ in $\mathscr{M}$. As we saw earlier, the code of $\underline{i}$ (and q_i) are bounded by some polynomial of $\exp(i)$ and the code of Λ is polynomially bounded by $\exp\big((\omega_1(a)^2)\big)$ or $\exp^2\big(2(\log a)^2\big)$, and finally $\omega_2(\ulcorner\Lambda\urcorner)$ is polynomially bounded by $\exp^2\big(4(\log a)^4\big)$; which exists by the assumption $a \in \mathscr{I}$. We note that a is non-standard, because otherwise we would have $a \in \log \mathscr{I}$ and whence $\mathscr{M}$ would be a model of $\mathrm{I}\Delta_0 + \exists x \in \log \mathscr{I}\theta(x)$, and this theory is inconsistent; a contradiction. The existence of $\omega_2(\ulcorner\Lambda\urcorner)$ assures the existence of a non-standard element $j (\leqslant \log^4(\ulcorner\Lambda\urcorner))$ for which $\Lambda^{\langle j\rangle}$ exists, and so by the assumption $\mathscr{M} \models \mathrm{HCon}(T)$ there must exist some T-evaluation p on $\Lambda^{\langle j\rangle}$ (hence, on $\Lambda^{\langle\infty\rangle}$) in $\mathscr{M}$. So, we can form the model $\mathfrak{M}(\Lambda, p)$. For this model we have $\mathfrak{M}(\Lambda, p) \models T$ by Lemma 4. Since $\mathscr{M} \models \theta(a)$ (and $\mathfrak{M}(\Lambda, p) \models T_1$) then $\mathfrak{M}(\Lambda, p) \models \theta(\underline{a}/p)$ by Theorem 3. Also, since $\mathfrak{M}(\Lambda, p) \models T_2$ and $\mathsf{q}_0, \mathsf{q}_1, \ldots, \mathsf{q}_b$ (for $b = 4a^4$) are elements of $\mathfrak{M}(\Lambda, p)$ such that $\mathfrak{M}(\Lambda, p) \models \mathsf{q}_0 = 2$ and $\mathfrak{M}(\Lambda, p) \models \mathsf{q}_{i+1} = \mathsf{q}_i^2$ for any $i < b$, then (by the main property of T_2) $\mathfrak{M}(\Lambda, p) \models$ "$b \in \mathfrak{log}^2$". Or in other words we have $\mathfrak{M}(\Lambda, p) \models$ "$\underline{a}/p \in \log \mathscr{I}$". Whence, $\mathfrak{M}(\Lambda, p) \models [\underline{a}/p \in \log \mathscr{I} \wedge \theta(\underline{a}/p)]$. So, $\mathfrak{M}(\Lambda, p)$ is a model of $T + \exists x \in \log \mathscr{I}\theta(x)$, and this is in contradiction with the assumption of $T \supseteq T_0$ and the inconsistency of the theory $T_0 + \exists x \in \log \mathscr{I}\theta(x)$. Thus, $\mathscr{M} \not\models \mathrm{HCon}(T)$; and so $\mathrm{I}\Delta_0 \nvdash \mathrm{HCon}(T)$. □

Acknowledgments This research is partially supported by grant no. 89030062 of the Institute for Research in Fundamental Sciences (IPM), Niavaran, Tehran, Iran.

References

1. Adamowicz, Z.: On Tableaux consistency in weak theories. Preprint # 618, Institute of Mathematics, Polish Academy of Sciences (2001). http://www.impan.pl/Preprints/p618.ps
2. Adamowicz, Z., Zbierski, P.: On Herbrand consistency in weak arithmetic. Arch. Math. Logic **40**, 399–413 (2001). doi:10.1007/s001530000072
3. Adamowicz, Z.: Herbrand consistency and bounded arithmetic. Fund. Math. **171**, 279–292 (2002). http://journals.impan.gov.pl/fm/Inf/171-3-7.html
4. Adamowicz, Z., Zdanowski, K.: Lower bounds for the unprovability of Herbrand consistency in weak arithmetics. Fund. Math. **212**, 191–216 (2011). doi:10.4064/fm212-3-1
5. Buss, S.R.: On Herbrand's theorem. In: Maurice, D., Leivant, R. (eds.) Selected Papers from the International Workshop on Logic and Computational Complexity. Indianapolis, IN, USA, October 13–16, 1994, Lecture Notes in Computer Science, vol. 960, pp. 195–209. Springer, Berlin (1995). http://math.ucsd.edu/~sbuss/ResearchWeb/herbrandtheorem/
6. Hájek, P., Pudlák, P.: Metamathematics of First-Order Arithmetic. Springer, Berlin (1998) (2nd printing). http://projecteuclid.org/handle/euclid.pl/1235421926
7. Kołodziejczyk, L.A.: On the Herbrand notion of consistency for finitely axiomatizable fragments of bounded arithmetic theories. J. Symb. Log. **71**, 624–638 (2006). doi:10.2178/jsl/1146620163
8. Paris, J.B., Wilkie, A.J.: Δ_0 sets and induction. In: Guzicki, W., Marek, W., Plec, A., Rauszer, C. (eds.) Proceedings of Open Days in Model Theory and Set Theory, Jadwisin, Poland 1981, pp. 237–248. Leeds University Press, Leeds (1981).
9. Salehi, S.: Unprovability of Herbrand consistency in weak arithmetics. In: Striegnitz, K. (ed.) Proceedings of the Sixth ESSLLI Student Session, European Summer School for Logic, Language, and Information, pp. 265–274 (2001). http://saeedsalehi.ir/pdf/esslli.pdf
10. Salehi, S.: Herbrand consistency in arithmetics with bounded induction. Ph.D. Dissertation, Institute of Mathematics, Polish Academy of Sciences (2002). http://saeedsalehi.ir/pphd.html
11. Salehi, S.: Separating bounded arithmetical theories by Herbrand consistency. J. Log. Comput. **22**, 545–560 (2012). doi:10.1093/logcom/exr005
12. Salehi, S.: Herbrand consistency of some arithmetical theories. J. Symb. Log. **77**, 807–827 (2012). doi:10.2178/jsl/1344862163
13. Willard, D.E.: How to extend the semantic Tableaux and cut-free versions of the second incompleteness theorem almost to Robinson's arithmetic Q. J. Symb. Log. **67**, 465–496 (2002). doi:10.2178/jsl/1190150055
14. Willard, D.E.: Passive induction and a solution to a Paris–Wilkie open question. Ann. Pure Appl. Log. **146**, 124–149 (2007). doi:10.1016/j.apal.2007.01.003