Separating bounded arithmetical theories by Herbrand consistency

SAEED SALEHI, *Department of Mathematics, University of Tabriz, POBox 51666–17766, Tabriz, Iran.*
E-mail: root@SaeedSalehi.ir

Journal of Logic and Computation, Volume 22, Issue 3 (2012) pp. 545--560.

Abstract

The problem of Π_1-separating the hierarchy of bounded arithmetic has been studied in the article. It is shown that the notion of Herbrand consistency, in its full generality, cannot Π_1-separate the theory $\mathrm{I}\Delta_0+\bigwedge_j\Omega_j$ from $\mathrm{I}\Delta_0$; though it can Π_1-separate $\mathrm{I}\Delta_0+\mathrm{Exp}$ from $\mathrm{I}\Delta_0$. Namely, we show the unprovability of the Herbrand consistency of $\mathrm{I}\Delta_0$ in the theory $\mathrm{I}\Delta_0+\bigwedge_j\Omega_j$. This partially extends a result of L. A. Kołodziejczyk who showed that for a finite fragment $\mathrm{S}\subseteq\mathrm{I}\Delta_0$, the Herbrand consistency of $\mathrm{S}+\Omega_1$ is not provable in $\mathrm{I}\Delta_0+\bigwedge_j\Omega_j$.

Keywords: Bounded arithmetics, Herbrand consistency, Π_1-conservative extensions.

1 Introduction

One of the consequences of Gödel's Incompleteness Theorems is the separation of Truth and Provability, in the sense that there are true sentences which are not provable, in sufficiently strong theories. Moreover, those true and unprovable sentences could be Π_1 (see Section 3.2). Thus, Truth is not Π_1-conservative over Provable. Gödel's Second Incompleteness Theorem provides a concrete candidate for Π_1-separating a theory T over its subtheory S, and that is the consistency statement of S; when T proves the consistency of S, then T is not a Π_1-conservative extension over S, since by the second incompleteness theorem of Gödel, S cannot prove its own consistency. Indeed, there are lots of Π_1-separate examples of theories (see Section 2.2), and there are some difficult open problems relating to Π_1-separation or Π_1-conservativeness of arithmetical theories. One of the well-known ones was the Π_1-separation of $\mathrm{I}\Delta_0+\mathrm{Exp}$, elementary arithmetic, from $\mathrm{I}\Delta_0$, bounded arithmetic. Here Gödel's Second Incompleteness Theorem cannot be applied directly, since $\mathrm{I}\Delta_0+\mathrm{Exp}$ does not prove the consistency of $\mathrm{I}\Delta_0$. For this Π_1-separation, Paris and Wilkie [8] suggested the notion of cut-free consistency instead of the usual, Hilbert style, consistency predicate. Here one can show the provability of the cut-free consistency of $\mathrm{I}\Delta_0$ in the theory $\mathrm{I}\Delta_0+\mathrm{Exp}$, and it was presumed that $\mathrm{I}\Delta_0$ should not derive its own cut-free consistency (see [11, 12] for some historical accounts). But this generalization of Gödel's Second Incompleteness Theorem, that of unprovability of the weak notions of consistency of weak theories in themselves, took a long time to be established. For example, it was shown that $\mathrm{I}\Delta_0$ cannot prove the Herbrand consistency of itself augmented with the axiom of the totality of the squaring function ($\forall x\exists y[y=x\cdot x]$) – see [12]; and then, by a completely different proof, it is shown in [11] the unprovability of the Herbrand consistency of $\mathrm{I}\Delta_0$ in itself, when its standard axiomatization is taken. Thus, one line of research was opened for investigating the status of Gödel's Second Incompleteness Theorem for weak notions of consistencies in weak arithmetics. In another direction, one can ask whether weak notions of consistencies can Π_1-separate the hierarchies of weak arithmetics. One prominent result here is of L. A. Kołodziejczyk [5]; one consequence of which is that the notion of Herbrand consistency cannot Π_1-separate the theory $\mathrm{I}\Delta_0+\bigwedge\Omega_j$ (see Section 2.2) from $\mathrm{I}\Delta_0+\Omega_1$. We conjectured in [11] that by using our techniques and methods one can (partially)

Vol. 22 No. 3,
Published online March 15, 2011 doi:10.1093/logcom/exr005

extend this result by showing the unprovability of the Herbrand consistency of $\mathrm{I}\Delta_0$ in $\mathrm{I}\Delta_0+\bigwedge\Omega_{\mathrm{j}}$ (Conjecture 39). In this article, we prove the conjecture. The ideas of the arguments are heavily based on the papers [1] and [5]; one new trick is a more efficient Skolemization which allows us to extract a Skolem function symbol for squaring ($x\mapsto x^2$) from an induction axiom of $\mathrm{I}\Delta_0$. This obstacle could have been overcome by injecting a function symbol for squaring into the language of arithmetic or by accepting an axiom like $\forall x\exists y(y=x\cdot x)$, which gives out a Skolem function for squaring. Here we have avoided those ways, and used the standard language of arithmetic and standard axiomatization of $\mathrm{I}\Delta_0$. The arguments of the paper go rather quickly, nevertheless some explanations and examples are presented for clarifying them. No familiarity with the papers cited in the references is assumed for reading this paper; the classic book of Peter Hájek and Pavel Pudlák [4] is more than enough.

2 Herbrand consistency and bounded arithmetic

2.1 Herbrand consistency

For Skolemizing formulas, it is convenient to work with formulas in *negation normal form*, which are formulas built up from atomic and negated atomic formulas using $\wedge,\vee,\forall$ and $\exists$. For having more comfort we consider *rectified* formulas, which have the property that different quantifiers refer to different variables, and no variable appears both bound and free. Let us note that any formula can be negation normalized uniquely by converting implication ($A\rightarrow B$) to disjunction ($\neg A\vee B$) and using de Morgan's laws. And renaming the variables can rectify the formula. Thus, any formula can be rewritten in the rectified negation normal form (RNNF) in a somehow unique way (up to a variable renaming). For any (not necessarily RNNF) existential formula of the form $\exists xA(x)$, let $\mathfrak{f}_{\exists xA(x)}$ be a new $m-$ary function symbol where m is the number of the free variables of $\exists xA(x)$. When $m=0$ then $\mathfrak{f}_{\exists xA(x)}$ will obviously be a new constant symbol (cf. [3]). For any RNNF formula φ define φ^{S} by induction:

- $\varphi^{\mathsf{S}}=\varphi$ for atomic or negated-atomic φ;
- $(\varphi\circ\psi)^{\mathsf{S}}=\varphi^{\mathsf{S}}\circ\psi^{\mathsf{S}}$ for $\circ\in\{\wedge,\vee\}$ and RNNF formulas φ,ψ;
- $(\forall x\varphi)^{\mathsf{S}}=\forall x\varphi^{\mathsf{S}}$;
- $(\exists x\varphi)^{\mathsf{S}}=\varphi^{\mathsf{S}}[\mathfrak{f}_{\exists x\varphi(x)}(\overline{y})/x]$ where $\overline{y}$ are the free variables of $\exists x\varphi(x)$ and the formula $\varphi^{\mathsf{S}}[\mathfrak{f}_{\exists x\varphi(x)}(\overline{y})/x]$ results from the formula φ^{S} by replacing all the occurrences of the variable x with the term $\mathfrak{f}_{\exists x\varphi(x)}(\overline{y})$.

Finally, the Skolemized form of a formula ψ is obtained by

(1) negation normalizing and rectifying it to φ;
(2) getting φ^{S} by the above inductive procedure;
(3) removing all the remaining (universal) quantifiers in φ^{S}.

We denote thus resulted Skolemized form of ψ by ψ^{Sk}. Note that our way of Skolemizing did not need prenex normalizing a formula. And it results in a unique (up to a variable renaming) Skolemized formula.

EXAMPLE 2.1
Let 0 be a constant symbol, $\mathfrak{s}$ be a unary function symbol, $+$ and $\cdot$ be two binary function symbols, and $\leqslant$ be a binary predicate symbol. Let A be the sentence $\forall x\forall y(x\leqslant y\leftrightarrow\exists z[z+x=y])$ which is an axiom of Robinson's Arithmetic Q (see Example 2.3), and let $B=\theta(0)\wedge\forall x[\theta(x)\rightarrow\theta(x+1)]\Rightarrow\forall x\theta(x)$, where $\theta(x)=\exists y(y\leqslant x\cdot x\wedge y=x\cdot x)$. This is an axiom of the theory $\mathrm{I}\Delta_0$ (see Section 2.2). The rectified

negation normalized forms of these sentences can be obtained as follows:

$$C = A^{\mathrm{RNNF}} = \forall x \forall y \big([x \not\leqslant y \vee \exists u(u+x=y)] \wedge [\forall z(z+x \neq y) \vee x \leqslant y]\big), \text{ and}$$

$$D = B^{\mathrm{RNNF}} = \forall u(u \not\leqslant 0{\cdot}0 \vee u \neq 0{\cdot}0) \bigvee$$
$$\exists w\big[(\exists z[z \leqslant w{\cdot}w \wedge z = w{\cdot}w]) \wedge (\forall v[v \not\leqslant (\mathfrak{s}w){\cdot}(\mathfrak{s}w) \vee v \neq (\mathfrak{s}w){\cdot}(\mathfrak{s}w)])\big] \bigvee$$
$$\forall x \exists y[y \leqslant x{\cdot}x \wedge y = x{\cdot}x].$$

For simplifying the notation, let $\mathfrak{h}$ stand for $\mathfrak{f}_{\exists u(u+x=y)}$, $\mathfrak{c}$ abbreviate the Skolem constant symbol for the sentence $\exists w\big[(\exists z[z \leqslant w{\cdot}w \wedge z = w{\cdot}w]) \wedge (\forall v[v \not\leqslant (\mathfrak{s}w){\cdot}(\mathfrak{s}w) \wedge v \neq (\mathfrak{s}w){\cdot}(\mathfrak{s}w)])\big]$, and $\mathfrak{q}(\xi)$ be the Skolem function symbol for the formula $\exists z[z \leqslant \xi \cdot \xi \wedge z = \xi \cdot \xi]$. Then C^{S} and D^{S} are as follows:

$$C^{\mathsf{S}} = \forall x \forall y \big([x \not\leqslant y \vee (\mathfrak{h}(x,y)+x=y)] \wedge [\forall z(z+x \neq y) \vee x \leqslant y]\big), \text{ and}$$

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

Finally the Skolemized forms of A and B are obtained as:

$$A^{\mathrm{Sk}} = [x \not\leqslant y \vee (\mathfrak{h}(x,y)+x=y)] \wedge [(z+x \neq y) \vee x \leqslant y], \text{ and}$$

$$B^{\mathrm{Sk}} = (u \not\leqslant 0{\cdot}0 \vee u \neq 0{\cdot}0) \bigvee$$
$$\big[(\mathfrak{q}(\mathfrak{c}) \leqslant \mathfrak{c}{\cdot}\mathfrak{c} \wedge \mathfrak{q}(\mathfrak{c}) = \mathfrak{c}{\cdot}\mathfrak{c}) \wedge (v \not\leqslant (\mathfrak{s}\mathfrak{c}){\cdot}(\mathfrak{s}\mathfrak{c}) \vee v \neq (\mathfrak{s}\mathfrak{c}){\cdot}(\mathfrak{s}\mathfrak{c}))\big] \bigvee$$
$$(\mathfrak{q}(x) \leqslant x{\cdot}x \wedge \mathfrak{q}(x) = x{\cdot}x).$$

◇

An *Skolem instance* of a formula ψ is any formula resulted from substituting the free variables of ψ^{Sk} with some terms. So, if $x_1, \ldots, x_n$ are the free variables of ψ^{Sk} (thus written as $\psi^{\mathrm{Sk}}(x_1, \ldots, x_n)$) then an Skolem instance of ψ is $\psi^{\mathrm{Sk}}[t_1/x_1, \cdots, t_n/x_n]$ where $t_1, \ldots, t_n$ are terms (which could be constructed from the Skolem functions symbols). The Skolemized form of a theory T is by definition $T^{\mathrm{Sk}} = \{\varphi^{\mathrm{Sk}} \mid \varphi \in T\}$. A version of Herbrand's fundamental theorem reads as follows (cf. [3]).

THEOREM 2.2 (Gödel - Herbrand - Skolem)
Any theory T is equiconsistent with its Skolemized theory T^{Sk}. That is, T is consistent if and only if every finite set of Skolem instances of T is (propositionally) satisfiable.

Our means of propositional satisfiability is by *evaluations*, which are defined to be any function p whose domains are the set of all atomic formulas constructed from a given set of terms Λ and whose ranges are the set $\{0,1\}$ such that

(1) $p[t=t]=1$ for all $t \in \Lambda$; and for any terms $t, s \in \Lambda$,
(2) if $p[t=s]=1$ then $p[\varphi(t)] = p[\varphi(s)]$ for any atomic formula $\varphi(x)$.

The relation $\backsim_p$ on Λ is defined by $t \backsim_p s \iff p[t=s]=1$ for $t, s \in \Lambda$. One can see that the relation $\backsim_p$ is an equivalence relation, and moreover is a congruence relation as well. That is, for any set of terms t_i and s_i (for $i = 1, \ldots, n$) and any n-ary function symbol f, if $p[t_1 = s_1] = \cdots p[t_n = s_n] = 1$ then $p[f(t_1, \ldots, t_n) = f(s_1, \ldots, s_n)] = 1$.

The $\backsim_p$−class of a term t is denoted by t/p; and the set of all such p−classes for each $t\in\Lambda$ is denoted by Λ/p. For simplicity, we write $p\models\varphi$ instead of $p[\varphi]=1$; thus $p\not\models\varphi$ stands for $p[\varphi]=0$. This definition of *satisfying* can be generalized to other open (RNNF) formulas as usual.

If all terms appearing in an Skolem instance of φ belong to the set Λ, that formula is called an Skolem instance of φ *available* in Λ. An evaluation defined on Λ is called a φ−*evaluation* if it satisfies all the Skolem instances of φ which are available in Λ. Similarly, for a theory T, a T−*evaluation* on Λ is an evaluation on Λ which satisfies every Skolem instance of every formula of T which is available in Λ. By Herbrand's Theorem, a theory T is consistent if and only if for every set of terms Λ (constructed from the Skolem terms of axioms of T) there exists a T−evaluation on Λ. We will use this reading of Herbrand's Theorem for defining the notion of *Herbrand consistency*. Thus, *Herbrand Provability* of a formula φ in a theory T is equivalent to the existence of a set of terms on which there cannot exist any $(T\cup\{\neg\varphi\})$−evaluation.

EXAMPLE 2.3
Let Q denote Robinson's Arithmetic over the language of arithmetic $\langle 0,\mathfrak{s},+,\cdot,\leqslant\rangle$, where 0 is a constant symbol, $\mathfrak{s}$ is a unary function symbol, $+,\cdot$ are binary function symbols and $\leqslant$ is a binary predicate symbol, whose axioms are as follows:

$A_1:\ \forall x(\mathfrak{s}x\neq 0)$ $\qquad A_2:\ \forall x\forall y(\mathfrak{s}x=\mathfrak{s}y\rightarrow x=y)$

$A_3:\ \forall x(x\neq 0\rightarrow\exists y[x=\mathfrak{s}y])$ $\qquad A_4:\ \forall x\forall y(x\leqslant y\leftrightarrow\exists z[z+x=y])$

$A_5:\ \forall x(x+0=x)$ $\qquad A_6:\ \forall x\forall y(x+\mathfrak{s}y=\mathfrak{s}(x+y))$

$A_7:\ \forall x(x\cdot 0=0)$ $\qquad A_8:\ \forall x\forall y(x\cdot\mathfrak{s}y=x\cdot y+x)$

Let $\varphi=\forall x(x\leqslant 0\rightarrow x=0)$. We can show $\mathrm{Q}\vdash\varphi$; this will be proved below by Herbrand provability. Suppose Q has been Skolemized as below:

$A_1^{\mathrm{Sk}}:\ \mathfrak{s}x\neq 0$ $\qquad A_2^{\mathrm{Sk}}:\ \mathfrak{s}x\neq\mathfrak{s}y\vee x=y$

$A_3^{\mathrm{Sk}}:\ x=0\vee x=\mathfrak{s}\mathfrak{p}x$ $\qquad A_4^{\mathrm{Sk}}:\ [x\not\leqslant y\vee\mathfrak{h}(x,y)+x=y]\wedge[z+x\neq y\vee x\leqslant y]$

$A_5^{\mathrm{Sk}}:\ x+0=x$ $\qquad A_6^{\mathrm{Sk}}:\ x+\mathfrak{s}y=\mathfrak{s}(x+y)$

$A_7^{\mathrm{Sk}}:\ x\cdot 0=0$ $\qquad A_8^{\mathrm{Sk}}:\ x\cdot\mathfrak{s}y=x\cdot y+x$

Here $\mathfrak{p}$ abbreviates $\mathfrak{f}_{\exists y(x=\mathfrak{s}y)}$ and $\mathfrak{h}$ stands for $\mathfrak{f}_{\exists z(z+x=y)}$. Suppose that $\neg\varphi$ has been Skolemized as $(\mathfrak{c}\leqslant 0\wedge\mathfrak{c}\neq 0)$ where $\mathfrak{c}$ is the Skolem constant symbol for $\exists x(x\leqslant 0\wedge x\neq 0)$. Put

$$\Lambda=\{0,\mathfrak{c},\mathfrak{h}(\mathfrak{c},0),\mathfrak{h}(\mathfrak{c},0)+\mathfrak{c},\mathfrak{s}\mathfrak{p}\mathfrak{c},\mathfrak{s}(\mathfrak{h}(\mathfrak{c},0)+\mathfrak{p}\mathfrak{c}),\mathfrak{h}(\mathfrak{c},0)+\mathfrak{s}\mathfrak{p}\mathfrak{c}\}.$$

We show that there cannot exist a $(\mathrm{Q}+\neg\varphi)$−evaluation on Λ. Assume (for the sake of contradiction) that p is such an evaluation. Then by A_3 we have $p\models\mathfrak{c}=\mathfrak{s}\mathfrak{p}\mathfrak{c}$. On the other hand, by A_4 we have $p\models\mathfrak{h}(\mathfrak{c},0)+\mathfrak{c}=0$, and so $p\models\mathfrak{h}(\mathfrak{c},0)+\mathfrak{s}\mathfrak{p}\mathfrak{c}=0$. Then by A_6 we get $p\models\mathfrak{s}(\mathfrak{h}(\mathfrak{c},0)+\mathfrak{p}\mathfrak{c})=0$ which is a contradiction with A_1. ◇

Let us note that finding a suitable set of terms Λ for which there cannot exist a $(T+\neg\psi)$−evaluation on Λ is as complicated as finding a proof of $T\vdash\psi$ (even more complicated, see Section 3.1). The following is another example for illustrating the concepts of Skolem instances and evaluations, which will be used later.

EXAMPLE 2.4
Let B be as in the Example 2.1, in the language $\langle 0, \mathfrak{s}, +, \cdot, \leqslant \rangle$. Thus,

$$B = \theta(0) \wedge \forall x[\theta(x) \to \theta(\mathfrak{s}x)] \to \forall x \theta(x) \text{ where } \theta(x) = \exists y(y \leqslant x \cdot x \wedge y = x \cdot x).$$

We saw that the Skolemized form of B is

$$\begin{aligned} B^{\mathrm{Sk}} = & (u \not\leqslant 0 \cdot 0 \vee u \neq 0 \cdot 0) \bigvee \\ & \big[(\mathfrak{q}(\mathfrak{c}) \leqslant \mathfrak{c} \cdot \mathfrak{c} \wedge \mathfrak{q}(\mathfrak{c}) = \mathfrak{c} \cdot \mathfrak{c}) \wedge (v \not\leqslant (\mathfrak{s}\mathfrak{c}) \cdot (\mathfrak{s}\mathfrak{c}) \vee v \neq (\mathfrak{s}\mathfrak{c}) \cdot (\mathfrak{s}\mathfrak{c}))\big] \bigvee \\ & (\mathfrak{q}(x) \leqslant x \cdot x \wedge \mathfrak{q}(x) = x \cdot x), \end{aligned}$$

where $\mathfrak{q}(\xi)$ is the Skolem function symbol for the formula $\exists z[z \leqslant \xi \cdot \xi \wedge z = \xi \cdot \xi]$ and $\mathfrak{c}$ is the Skolem constant of $\exists w\big[(\exists z[z \leqslant w \cdot w \wedge z = w \cdot w]) \wedge (\forall v[v \not\leqslant (\mathfrak{s}w) \cdot (\mathfrak{s}w) \wedge v \neq (\mathfrak{s}w) \cdot (\mathfrak{s}w)])\big]$. Define the set of terms Υ by $\Upsilon = \{0, 0+0, 0^2, \mathfrak{c}, \mathfrak{c}^2, \mathfrak{c}^2+0, \mathfrak{s}\mathfrak{c}, \mathfrak{q}\mathfrak{c}, (\mathfrak{s}\mathfrak{c})^2, (\mathfrak{s}\mathfrak{c})^2+0\}$ and suppose p is an $(Q+B)-$evaluation on the set of terms $\Upsilon \cup \{t, t^2, \mathfrak{q}(t)\}$. The notation ϱ^2 is a shorthand for $\varrho \cdot \varrho$. Then p must satisfy the following Skolem instance of B which is available in the set $\Upsilon \cup \{t, t^2, \mathfrak{q}(t)\}$:

$$\begin{aligned} (1) \qquad & (0 \not\leqslant 0^2 \vee 0 \neq 0^2) \bigvee \\ & \Big((\mathfrak{q}\mathfrak{c} \leqslant \mathfrak{c}^2 \wedge \mathfrak{q}\mathfrak{c} = \mathfrak{c}^2) \wedge \big((\mathfrak{s}\mathfrak{c})^2 \not\leqslant (\mathfrak{s}\mathfrak{c})^2 \vee (\mathfrak{s}\mathfrak{c})^2 \neq (\mathfrak{s}\mathfrak{c})^2\big)\Big) \bigvee \\ & \big(\mathfrak{q}(t) \leqslant t^2 \wedge \mathfrak{q}(t) = t^2\big). \end{aligned}$$

Now since $p \models 0 \cdot 0 = 0+0 = 0$ then, by Q's axioms, $p \models 0 \leqslant 0^2 \wedge 0 = 0^2$, and so p cannot satisfy the first disjunct of (1). Similarly, since $p \models (\mathfrak{s}\mathfrak{c})^2 + 0 = (\mathfrak{s}\mathfrak{c})^2$ then $p \models (\mathfrak{s}\mathfrak{c})^2 \leqslant (\mathfrak{s}\mathfrak{c})^2$, thus p cannot satisfy the second disjunct of (1) either, because $p \models (\mathfrak{s}\mathfrak{c})^2 = (\mathfrak{s}\mathfrak{c})^2$. Whence, p must satisfy the third disjunct of (1), then necessarily $p \models \mathfrak{q}(t) = t^2$ must hold. ◇

2.2 Bounded arithmetic hierarchy

First-order Peano arithmetic (PA) is the theory in the language $\langle 0, \mathfrak{s}, +, \cdot, \leqslant \rangle$ axiomatized by Robinson's Arithmetic Q (Example 2.3) plus induction schema $\psi(0) \wedge \forall x[\psi(x) \to \psi(\mathfrak{s}(x))] \Rightarrow \forall x \psi(x)$ for any formula $\psi(x)$. This theory is believed to encompass a large body of arithmetical truth in mathematics; the most recent conjecture (due to H. Friedman) is that a proof of Fermat's Last Theorem can be carried out inside PA ([2]), and indeed Andrew Wiles's proof of the theorem has been claimed to be formalized in it ([6]). To see a simpler example, we note that primality can be expressed in the language of arithmetic by the following formula: $\mathsf{Prime}(x) \equiv \forall y, z(y \cdot z = x \to y = 1 \vee z = 1)$. Then Euclid's theorem on the infinitude of the primes can be written as $\forall x \exists y[y > x \wedge \mathsf{Prime}(y)]$. It can be shown that Euclid's proof can be formalized completely in PA. One would wish to see how much strength of PA is necessary for proving the infinitude of the primes. An important subtheory of Peano's Arithmetic is introduced by R. Parikh ([7]) as follows. A formula is called bounded if its every quantifier is bounded, i.e. is either of the form $\forall x \leqslant t(\ldots)$ or $\exists x \leqslant t(\ldots)$ where t is a term; they are read as $\forall x(x \leqslant t \to \ldots)$ and $\exists x(x \leqslant t \wedge \ldots)$, respectively. The class of bounded formulas is denoted by Δ_0. It is easy to see that bounded formulas are decidable. The theory $\mathrm{I}\Delta_0$, also called bounded arithmetic, is axiomatized by Q plus the induction schema for bounded formulas. An important property of this arithmetic is that whenever $\mathrm{I}\Delta_0 \vdash \forall \overline{x} \exists y\, \theta(\overline{x})$ for a bounded formula θ, then there exists a term (polynomial) $t(\overline{x})$ such that $\mathrm{I}\Delta_0 \vdash \forall \overline{x} \exists y \leqslant t(\overline{x})\, \theta(\overline{x})$ (see e.g. [4]). An open problem in the theory of weak arithmetics is that whether or not the infinitude of the primes can be proved inside $\mathrm{I}\Delta_0$. However, it is known that much of elementary number theory cannot be proved inside $\mathrm{I}\Delta_0$; the theory is too weak to even recognize the totality of the exponentiation function. The exponentiation function exp is defined by $\exp(x)=2^x$; the formula Exp expresses its totality: $(\forall x \exists y[y=\exp(x)])$. We note that the formula '$y=\exp(x)$' can be written by a bounded formula in two free variables x, y in the language of arithmetic (see [4]). The theory $\mathrm{I}\Delta_0$ cannot prove Exp but is able to prove some basic properties of the exp function (see again [4]). The theory $\mathrm{I}\Delta_0+\mathrm{Exp}$, sometimes called Elementary Arithmetic, is able to formalize much of number theory. It can surely prove the infinitude of the primes. Note that in Euclid's proof, for getting a prime number greater than x one can use $x!+1$, which should have a prime factor greater than x (no number non-greater than x can divide it). And it can be seen that $x! < \exp\exp(x)$. Between $\mathrm{I}\Delta_0$ and $\mathrm{I}\Delta_0+\mathrm{Exp}$, a hierarchy of theories is considered in the literature, which has close connections with computational complexity. They are sometimes called weak arithmetics, and sometimes bounded arithmetics. The hierarchy is defined below. The converse of exp is denoted by log which is formally defined as $\log x = \min\{y \mid x \leqslant \exp(y)\}$; thus $\exp(\log x - 1) < x \leqslant \exp(\log x)$. The superscripts above the function symbols indicate the iteration of the functions; e.g. $\exp^2(x) = \exp\exp(x)$ and $\log^3 x = \log\log\log x$. Define the function ω_m to be $\omega_m(x) = \exp^m\big((\log^m x)\cdot(\log^m x)\big)$. It is customary to define this function by induction: $\omega_0(x)=x^2$ and $\omega_{\mathrm{n}+1}(x)=\exp(\omega_{\mathrm{n}}(\log x))$. Let Ω_{m} express the totality of ω_{m} (i.e. $\Omega_{\mathrm{m}} \equiv \forall x \exists y[y=\omega_{\mathrm{m}}(x)]$). It can be more convenient to consider the function $\omega_{-1}(x)=2x$ as well (cf. [5]). The hierarchy between $\mathrm{I}\Delta_0$ and $\mathrm{I}\Delta_0+\mathrm{Exp}$ is $\{\mathrm{I}\Delta_0+\Omega_{\mathrm{m}}\}_{\mathrm{m}\geqslant 1}$. For example, the theory $\mathrm{I}\Delta_0+\Omega_1$ can prove the infinitude of the primes (the proof is not easy at all, see [9]). We first review some basic properties of the ω_n functions: ω_1 dominates all the polynomials and ω_{m+1} dominates all the (finite) iterations of ω_m. Let us note that $\omega_0^N(x) = x^{\exp(N)}$ and $\omega_m^N(x) = \exp^m([\log^m x]^{\exp(N)})$, also $\omega_{j+1}^N(x) = \exp(\omega_j^N(\log x))$, for $N \geqslant 1$.

LEMMA 2.5
For any natural $m \geqslant 0$ and $N > 2$, and any $x > \exp^{m+2}(N)$, we have $\omega_m^N(x) < \omega_{m+1}(x)$.

PROOF. For $m=0$ we note that $2^N \cdot \log x < (\log x)^2$ for any $x > \exp^2(N)$. Thus $\exp(2^N \log x) < \exp((\log x)^2)$, which implies that $\omega_0^N(x) < \omega_1(x)$.

For $m \geqslant 1$ we can use an inductive argument. For any $x > \exp^{m+2}(N)$ we have $\log x > \exp^{m+1}(N)$, so by the induction hypothesis $\omega_{m-1}^N(\log x) < \omega_m(\log x)$. Then $\exp[\omega_{m-1}^N(\log x)] < \exp[\omega_m(\log x)]$, and so $\omega_m^N(x) < \omega_{m+1}(x)$. ■

The following lemma will be used later in the article.

LEMMA 2.6
For any $m \geqslant -1, N \geqslant 1$ and $x > \exp^{m+2}(4N+4)$, there exists some y ($\leqslant x$) such that

$$\omega_m^N(y) < x \leqslant \omega_{m+1}(y).$$

PROOF. We first show the lemma for $m=-1$: for any $x > \exp(4N+4)$, there exists a least y such that $y^2 \geqslant x$; so $(y-1)^2 < x$. Also from $y^2 > 2^{4N+4}$ we have $y > 2^{2N+2}$. Whence we have $x \leqslant y^2 = \omega_0(y)$, and also $\omega_{-1}^N(y) = 2^N \cdot y < \sqrt{y}\cdot y \leqslant (y-1)^2 < x$. Let us note that $\sqrt{y}\cdot y \leqslant (y-1)^2$ holds for any $y \geqslant 4$ and we have $y > 2^{2N+2} > 4$.

For $m=0$, we use the above argument for $\log x$, noting that $\log x > \exp(4N+4)$ holds by the assumption $x > \exp^2(4N+4)$. There must exist some z such that $2^N \cdot z < \log x \leqslant z^2$. Let $y = \exp(z)$, so $z = \log y$. Thus, from $2^N \log y < \log x \leqslant (\log y)^2$ it follows that $\omega_0^N(y) = y^{\exp(N)} \leqslant \exp[\exp(N)\cdot(\log y)] \leqslant \exp(\log x - 1) < x \leqslant \exp(\log x) \leqslant \exp([\log y]^2) = \omega_1(y)$.

For $m \geqslant 1$, we use induction on m with a straightforward argument. For $x > \exp^{m+3}(4N+4)$, we have $\log x > \exp^{m+2}(4N+4)$, and so by the induction hypothesis there exists some z such that $\omega_m^N(z) < \log x \leqslant \omega_{m+1}(z)$. Put $y = \exp(z)$, so we have $\omega_m^N(\log y) < \log x \leqslant \omega_{m+1}(\log y)$. Thus, we finally get $\omega_{m+1}^N(y) = \exp(\omega_m^N(\log y)) \leqslant \exp(\log x - 1) < x \leqslant \exp(\log x) \leqslant \exp(\omega_{m+1}(\log y)) = \omega_{m+2}(y)$. ■

Whence the hierarchy $\{\mathrm{I\Delta_0}\} \cup \{\mathrm{I\Delta_0} + \Omega_\mathrm{m}\}_{\mathrm{m} \geqslant 1} \cup \{\mathrm{I\Delta_0} + \bigwedge \Omega_\mathrm{j}, \mathrm{I\Delta_0} + \mathrm{Exp}\}$ is proper:

$$(2) \qquad \mathrm{I\Delta_0} \subsetneqq \cdots \mathrm{I\Delta_0} + \Omega_\mathrm{n} \subsetneqq \mathrm{I\Delta_0} + \Omega_\mathrm{n+1} \subsetneqq \cdots \subsetneqq \mathrm{I\Delta_0} + \bigwedge \Omega_\mathrm{j} \subsetneqq \mathrm{I\Delta_0} + \mathrm{Exp}.$$

The notation $\mathrm{I\Delta_0} + \bigwedge \Omega_\mathrm{j}$ abbreviates $\bigcup_{\mathrm{n} \geqslant 1}(\mathrm{I\Delta_0} + \Omega_\mathrm{n})$. The class of Σ_n–formulas and Π_n–formulas are defined as follows: Σ_1-formulas are equivalently (in first-order logic) in the form $\exists \overline{x} \theta(\overline{x})$, where $\theta \in \Delta_0$, and Π_1-formulas are equivalently in the form $\forall \overline{x} \theta(\overline{x})$, for some $\theta \in \Delta_0$. Then $\Sigma_{n+1}-$formulas are (logically) equivalent to $\exists \overline{x} \varphi(\overline{x})$ for some $\varphi \in \Pi_n$, and $\Pi_{n+1}-$formulas are equivalent to $\forall \overline{x} \varphi(\overline{x})$ for some $\varphi \in \Sigma_n$. The above hierarchy is not Π_2-conservative, i.e. there exists a Π_2-formula (namely Ω_{m+1}) which is provable in $\mathrm{I\Delta_0} + \Omega_\mathrm{m+1}$ but not in $\mathrm{I\Delta_0} + \Omega_\mathrm{m}$. Though, the (difficult) open problem here is the Π_1-conservativity of the hierarchy:

PROBLEM 2.7
Is there a Π_1-sentence ψ such that $\mathrm{I\Delta_0} + \Omega_\mathrm{m+1} \vdash \psi$ and $\mathrm{I\Delta_0} + \Omega_\mathrm{m} \nvdash \psi$? ◇

As for the above hierarchy (2) it is (only) known that $\mathrm{I\Delta_0} + \mathrm{Exp}$ is not Π_1-conservative over $\mathrm{I\Delta_0} + \bigwedge \Omega_\mathrm{j}$ (see [4], Corollary 5.34 and the afterward explanation).

Examples of Π_1-separation abound in mathematics and logic: Zermelo-Frankel Set Theory ZFC is not Π_1-conservative over Peano's Arithmetic PA, because we have $\mathrm{ZFC} \vdash \mathrm{Con(PA)}$ but, by Gödel's Second Incompleteness Theorem, $\mathrm{PA} \nvdash \mathrm{Con(PA)}$; where $\mathrm{Con}(-)$ is the consistency predicate. Inside PA the Σ_n-hierarchy is not a Π_1-conservative hierarchy, since $\mathrm{I\Sigma_{n+1}} \vdash \mathrm{Con(I\Sigma_n)}$ though $\mathrm{I\Sigma_n} \nvdash \mathrm{Con(I\Sigma_n)}$ (see e.g. [4]). Then below the theory $\mathrm{I\Sigma_1}$, things get more complicated: for Π_1-separating $\mathrm{I\Delta_0} + \mathrm{Exp}$ over $\mathrm{I\Delta_0}$ the candidate $\mathrm{Con(I\Delta_0)}$ does not work as expected, because $\mathrm{I\Delta_0} + \mathrm{Exp} \nvdash \mathrm{Con(I\Delta_0)}$ (see [4] Corollary 5.29). For this Π_1-separation, Paris and Wilkie [8] suggested the notion of cut-free consistency instead of the usual - Hilbert style - consistency predicate. Here one can show that $\mathrm{I\Delta_0} + \mathrm{Exp} \vdash \mathrm{CFCon(I\Delta_0)}$, and then it was presumed that $\mathrm{I\Delta_0} \nvdash \mathrm{CFCon(I\Delta_0)}$, where CFCon stands for cut-free consistency. In 2006, L. A. Kołodziejczyk [5] showed that the notion of Herbrand consistency (and thus, more probably, other cut-free consistencies, like Tableaux, etc.) will not work for Π_1-separating the hierarchy above $\mathrm{I\Delta_0} + \Omega_1$ either. Namely, $\mathrm{I\Delta_0} + \bigwedge \Omega_\mathrm{j} \nvdash \mathrm{HCon}(\mathrm{I\Delta_0} + \Omega_1)$, where $\mathrm{HCon}(-)$ is the predicate of Herbrand consistency (see Section 2.3). Actually the main result of [5] is stronger, in the sense that it proves the existence of a finite fragment $\mathrm{S} \subseteq \mathrm{I\Delta_0}$ such that $\mathrm{I\Delta_0} + \bigwedge \Omega_\mathrm{j}$ cannot prove $\mathrm{HCon}(\mathrm{S} + \Omega_1)$ (cf. Conjecture 4.1 in Section 4). In this article, we partially extend this rather negative result one step further, by proving $\mathrm{I\Delta_0} + \bigwedge \Omega_\mathrm{j} \nvdash \mathrm{HCon(I\Delta_0)}$.

2.3 *Herbrand consistency in bounded arithmetics*

For a theory T, when Λ is the set of all terms (constructed from the function symbols of the language of T and also the Skolem function symbols of the formulas of T) any $T-$evaluation on Λ induces a model of T, which is called a *Herbrand Model*. Let $\mathcal{L}$ be a language and Λ be a set of (ground) terms (constructed by the Skolem constant symbols and the Skolem function symbols of $\mathcal{L}$).

Put $\Lambda^{\langle 0\rangle}=\Lambda$, and define inductively

$$\Lambda^{\langle k+1\rangle}=\Lambda^{\langle k\rangle}\cup\{f(t_1,\ldots,t_m)\,|\,f\in\mathcal{L}\,\&\,t_1,\ldots,t_m\in\Lambda^{\langle k\rangle}\}$$
$$\cup\{\mathfrak{f}_{\exists x\psi(x)}(t_1,\ldots,t_m)\,|\,\ulcorner\psi\urcorner\leqslant k\,\&\,t_1,\ldots,t_m\in\Lambda^{\langle k\rangle}\}.$$

Let $\Lambda^{\langle\infty\rangle}$ denote the union $\bigcup_{k\in\mathbb{N}}\Lambda^{\langle k\rangle}$.
Suppose p is an evaluation on $\Lambda^{\langle\infty\rangle}$. Define $\mathfrak{M}(\Lambda,p)=\{t/p\,|\,t\in\Lambda^{\langle\infty\rangle}\}$ and put the $\mathcal{L}-$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) $R^{\mathfrak{M}(\Lambda,p)}=\{(t_1/p,\ldots,t_m/p)\,|\,p\models R(t_1,\ldots,t_m)\}$,

for function symbol f, relation symbols R, and terms $t_1,\ldots,t_m\in\Lambda^{\langle\infty\rangle}$.

LEMMA 2.8
The definition of $\mathcal{L}-$structure on $\mathfrak{M}(\Lambda,p)$ is well-defined, and when p is an $T-$evaluation on $\Lambda^{\langle\infty\rangle}$, for an $\mathcal{L}-$theory T, then $\mathfrak{M}(\Lambda,p)\models T$.

PROOF. That the definitions of $f^{\mathfrak{M}(\Lambda,p)}$ and $R^{\mathfrak{M}(\Lambda,p)}$ are well-defined follows directly from the definition of an evaluation. By the definition of $\Lambda^{\langle\infty\rangle}$, the structure $\mathfrak{M}(\Lambda,p)$ is closed under all the Skolem functions of $\mathcal{L}$, and moreover it satisfies an atomic (or negated atomic) formula $A(t_1/p,\ldots,t_m/p)$ if and only if $p\models A(t_1,\ldots,t_m)$. Then it can be shown, by induction on the complexity of formulas, that for every RNNF formula ψ, we have $\mathfrak{M}(\Lambda,p)\models\psi$ whenever p satisfies all the available Skolem instances of ψ in $\Lambda^{\langle\infty\rangle}$. Whence, if p is a $T-$evaluation, then $\mathfrak{M}(\Lambda,p)\models T$. ■

For arithmetizing the notion of Herbrand consistency, we adopt an efficient Gödel coding, introduced e.g. in Chapter V of [4]. For convenience, and shortening the computations, we introduce the $\mathcal{P}$ notation: we say x is of $\mathcal{P}(y)$, when x is bounded above by a polynomial of y; and we write this as $x\leqslant\mathcal{P}(y)$, meaning that for some natural n the inequality $x\leqslant y^n+n$ holds. Let us note that $x\leqslant\mathcal{P}(y)$ is equivalent to the old (more familiar) $O-$notation '$\log x\in\mathcal{O}(\log y)$'. Here, we collect some very basic facts about this fixed efficient coding that will be needed later.

REMARK 2.9
Let A,B be sets of terms and let $|A|,|B|$ denote their cardinality. Then

- $\ulcorner A\cup B\urcorner\leqslant 64\cdot(\ulcorner A\urcorner\cdot\ulcorner B\urcorner)$ (Proposition 3.29 page 311 of [4]); and
- $|A|\leqslant\log(\ulcorner A\urcorner)$ (Section (e), pp. 304–310 of [4]);

where $\ulcorner A\urcorner$ denotes the Gödel code of the set A. ◇

Let $\mathcal{L}_A=\langle 0,\mathfrak{s},+,\cdot,\leqslant\rangle$ be the language of arithmetics (see Example 2.3). If we let $\mathcal{L}_A^{\rm Sk}$ be the closure of $\mathcal{L}_A$ under Skolem function and constant symbols, i.e. let $\mathcal{L}_A^{\rm Sk}$ be the smallest set that contains $\mathcal{L}_A$ and for any $\mathcal{L}_A^{\rm Sk}-$formula $\exists x\phi(x)$ we have $\mathfrak{f}_{\exists x\phi(x)}\in\mathcal{L}_A^{\rm Sk}$, then this new countable language can also be re-coded, and this recoding can be generalized to $\mathcal{L}_A^{\rm Sk}-$terms and $\mathcal{L}_A^{\rm Sk}-$formulas. We wish to compute an upper bound for the codes of evaluations on a set of terms Λ. For a given Λ, all the atomic formulas, in the language $\mathcal{L}_A$, constructed from terms of Λ are either of the form $t=s$ or of the form $t\leqslant s$ for some $t,s\in\Lambda$. And every member of an evaluation p on Λ is an ordered pair like $\langle t=s,i\rangle$ or $\langle t\leqslant s,i\rangle$ for some $t,s\in\Lambda$ and $i\in\{0,1\}$. Thus, the code of any member of p is a constant multiple of $(\ulcorner t\urcorner\cdot\ulcorner s\urcorner)^2$, and so the code of p is bounded above by $\mathcal{P}(\prod_{t,s\in\Lambda}\ulcorner t\urcorner\cdot\ulcorner s\urcorner)$. Let us also note that $\prod_{t,s\in\Lambda}\ulcorner t\urcorner\cdot\ulcorner s\urcorner=\prod_{t\in\Lambda}(\ulcorner t\urcorner)^{2|\Lambda|}=(\prod_{t\in\Lambda}\ulcorner t\urcorner)^{2|\Lambda|}\leqslant\mathcal{P}(\ulcorner\Lambda\urcorner)^{2\log\ulcorner\Lambda\urcorner}\leqslant\mathcal{P}(\ulcorner\Lambda\urcorner^{\log\ulcorner\Lambda\urcorner})$ and that $\ulcorner\Lambda\urcorner^{\log\ulcorner\Lambda\urcorner}\leqslant\omega_1(\ulcorner\Lambda\urcorner)$. Thus, we have $\ulcorner p\urcorner\leqslant\mathcal{P}\big(\omega_1(\ulcorner\Lambda\urcorner)\big)$ for any evaluation p on any set of terms Λ. As noted in [11] there are $\exp(2|\Lambda|^2)$ different evaluations on the set Λ, and by $|\Lambda|\leqslant\log\ulcorner\Lambda\urcorner$ we get $\exp(2|\Lambda|^2)\leqslant\mathcal{P}\big(\exp((\log\ulcorner\Lambda\urcorner)^2)\big)\leqslant\mathcal{P}\big(\omega_1(\ulcorner\Lambda\urcorner)\big)$. So, only when $\omega_1(\ulcorner\Lambda\urcorner)$ exists, can we have all the evaluations on Λ in our disposal. We need an upper bound on the size (cardinal) and the code of $\Lambda^{\langle j\rangle}$ defined above.

THEOREM 2.10
If for a set of terms Λ with non-standard $\ulcorner\Lambda\urcorner$ the value $\omega_2(\ulcorner\Lambda\urcorner)$ exists, then for some non-standard j the value $\ulcorner\Lambda^{\langle j\rangle}\urcorner$ will exist.

PROOF. We first show that the following inequalities hold when $\ulcorner\Lambda\urcorner$ and $|\Lambda|$ are sufficiently larger than n: (1) $|\Lambda^{\langle n\rangle}|\leqslant\mathcal{P}\Big(|\Lambda|^{n!}\Big)$ and (2) $\ulcorner\Lambda^{\langle n\rangle}\urcorner\leqslant\mathcal{P}\Big(\big(\ulcorner\Lambda\urcorner\big)^{|\Lambda|^{(n+1)!}}\Big)$.
Denote $\ulcorner\Lambda^{\langle k\rangle}\urcorner$ by λ_k (thus $\ulcorner\Lambda\urcorner=\lambda_0=\lambda$) and $|\Lambda^{\langle k\rangle}|$ by σ_k (and thus $|\Lambda|=\sigma_0=\sigma$). We first note that $\sigma_{k+1}\leqslant\sigma_k+M\sigma_k^M+k\sigma_k^k$ for a fixed M. Thus $\sigma_{k+1}\leqslant\mathcal{P}(\sigma_k^{k+1})$, and then, by an inductive argument, we have $\sigma_n\leqslant\mathcal{P}(\sigma^{n!})$. For the second statement, we first compute an upper bound for the code of the Cartesian power A^m for a set A. Now we have $\ulcorner A^{k+1}\urcorner\leqslant\mathcal{P}\big(\prod_{t\in A^k\&s\in A}\ulcorner t\urcorner\cdot\ulcorner s\urcorner\big)\leqslant\mathcal{P}\big(\ulcorner A^k\urcorner^{|A|}\cdot\ulcorner A\urcorner^{|A|^k}\big)$, and thus $\ulcorner A^m\urcorner\leqslant\mathcal{P}\big(\ulcorner A\urcorner^{|A|^m}\big)$ can be shown by induction on m. We also have $\lambda_{k+1}\leqslant\mathcal{P}\big(\ulcorner\Lambda^{\langle k\rangle}\urcorner\cdot\ulcorner(\Lambda^{\langle k\rangle})^M\urcorner\cdot\ulcorner(\Lambda^{\langle k\rangle})^k\urcorner\big)$ for a fixed M. So, $\lambda_{k+1}\leqslant\mathcal{P}\big(\lambda_k^{\sigma_k{}^k}\big)$ and finally our desired conclusion $\lambda_m\leqslant\mathcal{P}\big(\lambda^{\sigma^{(m+1)!}}\big)$ follows by induction.

Now since $\ulcorner\Lambda\urcorner$ is a non-standard number, there must exist a non-standard j such that $j\leqslant\log^4(\ulcorner\Lambda\urcorner)$. Thus $2(j+1)!\leqslant\exp^2(j)\leqslant\log^2(\ulcorner\Lambda\urcorner)$. Now, by the inequality (2) above we can write $\ulcorner\Lambda^{\langle j\rangle}\urcorner\leqslant\mathcal{P}\Big((\ulcorner\Lambda\urcorner)^{|\Lambda|^{(j+1)!}}\Big)\leqslant\mathcal{P}\Big((2^{2\log\ulcorner\Lambda\urcorner})^{(\log\ulcorner\Lambda\urcorner)^{(j+1)!}}\Big)\leqslant\mathcal{P}\Big(\exp((\log\ulcorner\Lambda\urcorner)^{2(j+1)!})\Big)$, and so $\ulcorner\Lambda^{\langle j\rangle}\urcorner\leqslant\mathcal{P}\big(\exp(\omega_1(\log\ulcorner\Lambda\urcorner))\big)\leqslant\mathcal{P}\big(\omega_2(\ulcorner\Lambda\urcorner)\big)$. ■

The reason that Theorem 2.10 is stated for non-standard $\ulcorner\Lambda\urcorner$ is that the set $\Lambda^{\langle\infty\rangle}$, needed for constructing the model $\mathfrak{M}(\Lambda,p)$, is not definable in $\mathcal{L}_A$. But the existence of the definable $\Lambda^{\langle j\rangle}$ for a non-standard j can guarantee the existence of $\Lambda^{\langle\infty\rangle}$ and thus of $\mathfrak{M}(\Lambda,p)$. This non-standard j exists for non-standard $\ulcorner\Lambda\urcorner$. Finally, we formalize the notion of Herbrand consistency as follows.

DEFINITION 2.11
A theory T is called *Herbrand consistent* if for any set of terms Λ (constructed from the Skolem terms of T) for which $\omega_1(\ulcorner\Lambda\urcorner)$ exists, there is a $T-$evaluation on Λ.
This notion can be formalized in the language of arithmetic, denoted by HCon(T). ◇

REMARK 2.12
The above formalization of the notion of Herbrand consistency may seem unnatural, as one would like to have a $T-$evaluation on *any* set of terms Λ. The requirement for the existence of $\omega_1(\ulcorner\Lambda\urcorner)$ is only to assure the existence (availability) of all the evaluations on Λ. As it was noted before Theorem 2.10, the size of an evaluation on a given set of terms Λ may be roughly bounded by $\omega_1(\ulcorner\Lambda\urcorner)$. ◇

3 Separating bounded arithmetical hierarchy

3.1 Separating by Herbrand consistency

Let us recall that the (usual) Hilbert Provability $T\vdash\varphi$ is, by definition, the existence of a sequence of formulas whose last element is (the Gödel code of) φ and every other element is either a logical axiom or an axiom of T, or has been resulted from two previous elements by means of *modus ponens*. Thus, Hilbert consistency means the non-existence of such a sequence whose last element is a contradiction. Let us note that Herbrand consistency is, in a sense, a weaker notion of consistency; some more explanation is in order. The super-exponentiation function is defined by the equation $\sup\_\exp(x)=\exp^x(x)$; let Sup_Exp be the sentence which expresses the totality of this function ($\mathrm{Sup\_Exp}=\forall x\exists y[y=\sup\_\exp(x)]$). By the techniques of cut elimination (see e.g. [4]), the equivalence $\mathrm{I}\Delta_0+\mathrm{Sup\_Exp}\vdash\mathrm{Con}(T)\leftrightarrow\mathrm{HCon}(T)$ can be shown for any theory T. Though the theory $\mathrm{I}\Delta_0+\mathrm{Exp}$ it too weak to recognize this equivalence, since $\mathrm{I}\Delta_0+\mathrm{Exp}\vdash\mathrm{HCon}(\mathrm{Q})$ but $\mathrm{I}\Delta_0+\mathrm{Exp}\nvdash\mathrm{Con}(\mathrm{Q})$ ([4], Theorem 5.20 and Corollary 5.29). So, $\mathrm{I}\Delta_0+\mathrm{Exp}\nvdash\mathrm{HCon}(T)\to\mathrm{Con}(T)$ in general, though it can be shown that $\mathrm{I}\Delta_0+\mathrm{Exp}\vdash\mathrm{Con}(T)\to\mathrm{HCon}(T)$ (see [4]). Thus, showing the unprovability of Herbrand consistency of weak theories in themselves is an interesting generalization of Gödel's Second Incompleteness Theorem. We are interested here in the question as to whether the notion of Herbrand consistency can Π_1-separate the hierarchy (2) above. We already know (only) that $\mathrm{I}\Delta_0+\mathrm{Exp}$ is not Π_1-conservative over $\mathrm{I}\Delta_0+\bigwedge\Omega_\mathrm{j}$, but we do not yet know if these two theories can be Π_1-separated by the Herbrand consistency of $\mathrm{I}\Delta_0+\mathrm{Exp}$. In other words, we do not know whether $\mathrm{I}\Delta_0+\mathrm{Exp}$ is able to derive the Herbrand consistency of the theory $\mathrm{I}\Delta_0+\bigwedge\Omega_\mathrm{j}$.

CONJECTURE 3.1
The notion of Herbrand consistency cannot Π_1-separate the (already Π_1-distinct) theories $\mathrm{I}\Delta_0+\mathrm{Exp}$ and $\mathrm{I}\Delta_0+\bigwedge\Omega_\mathrm{j}$; that is $\mathrm{I}\Delta_0+\mathrm{Exp}\nvdash\mathrm{HCon}(\mathrm{I}\Delta_0+\bigwedge\Omega_\mathrm{j})$. ◇

Though, for any $m\geqslant 1$, Herbrand consistency can Π_1-separate $\mathrm{I}\Delta_0+\mathrm{Exp}$ from the theory $\mathrm{I}\Delta_0+\Omega_\mathrm{m}$, and also from $\mathrm{I}\Delta_0$. Since already $\mathrm{I}\Delta_0+\Omega_\mathrm{m}\nvdash\mathrm{HCon}(\mathrm{I}\Delta_0+\Omega_\mathrm{m})$ for any $m\geqslant 1$ (see [1, 10]) and also the following theorem hold.

THEOREM 3.2
For any $m\geqslant 1$ we have $\mathrm{I}\Delta_0+\mathrm{Exp}\vdash\mathrm{HCon}(\mathrm{I}\Delta_0+\Omega_\mathrm{m})$.

PROOF. Reason inside a model $\mathcal{M}\models\mathrm{I}\Delta_0+\mathrm{Exp}$. For any set of terms $\Lambda\in\mathcal{M}$, assume it has been rearranged in a non-decreasing order $\Lambda=\{t_0,t_1,t_2,\cdots,t_j\}$. Then for some terms $u_1,u_2,\cdots,u_j$ we have $t_1\leqslant\omega_m^{u_1}(t_0), t_2\leqslant\omega_m^{u_2}(t_1),\cdots,t_j\leqslant\omega_m^{u_j}(t_{j-1})$. Let $u=\sum_i u_i$; then $t_i\leqslant\omega_m^u(t_0)$ for each $i\leqslant j$. On the other hand, $\omega_m^u(t_0)=\exp^m([\log^m(t_0)]^{\exp(u)})\leqslant\exp^{m+1}(u\cdot t_0)$; and since $u\leqslant(\ulcorner\Lambda\urcorner)^2$ and exp is available for all elements, then every term in Λ has a realization inside $\mathcal{M}$. Denote the realization of t_i by $t_i^{\mathcal{M}}$. Then the evaluation p defined on Λ by

(1) $p\models t_k=t_l$ if and only if $t_k^{\mathcal{M}}=t_l^{\mathcal{M}}$, and (2) $p\models t_k\leqslant t_l$ if and only if $t_k^{\mathcal{M}}\leqslant t_l^{\mathcal{M}}$,

is an $(\mathrm{I}\Delta_0+\Omega_\mathrm{m})-$evaluation on Λ (note also that $\mathcal{M}\models\mathrm{I}\Delta_0+\Omega_\mathrm{m}$). Now the desired conclusion $\mathcal{M}\models\mathrm{HCon}(\mathrm{I}\Delta_0+\Omega_\mathrm{m})$ follows from the classical fact that there exists a well-behaved truth definition for bounded formulas which can be used in induction inside $\mathcal{M}$. ■

REMARK 3.3
The above proof also shows that $\mathrm{I}\Delta_0+\mathrm{Exp}\vdash\mathrm{HCon}(\mathrm{I}\Delta_0)$ and indeed it is shown in [11] that $\mathrm{I}\Delta_0$ does not prove $\mathrm{HCon}(\mathrm{I}\Delta_0)$. Thus, HCon(−) can Π_1-separate $\mathrm{I}\Delta_0+\mathrm{Exp}$ and $\mathrm{I}\Delta_0$ as well. ◇

REMARK 3.4
A reason that the proof of the above theorem cannot be applied for showing the presumably false deduction $\mathrm{I}\Delta_0+\mathrm{Exp}\vdash\mathrm{HCon}(\mathrm{I}\Delta_0+\bigwedge\Omega_\mathrm{j})$ in the conjecture, is that for the set of terms $\Xi=\{v_0,v_1,\cdots,v_j\}$ defined by $v_0=4$ and $v_{i+1}=\omega_{i+1}(v_i)$ for each $i<j$, we have $v_j=\exp^j(4)$ (the equality $v_i=\exp^i(4)$ follows by induction on i). Thus a model of $\mathrm{I}\Delta_0+\mathrm{Exp}$ can contain a big j, and the set Ξ above, for which $\exp^j(4)$ does not exist. So, some terms of Ξ may not have a realization in

the model; and a suitable evaluation could not be defined in it. Note that $\exp^j(4)$ is a super-exponential term and cannot be obtained by applying a finite number of the exponential function. ◇

3.2 *Unprovability of Herbrand consistency of* $\mathrm{I}\Delta_0$ *in* $\mathrm{I}\Delta_0+\bigwedge\Omega_{\mathrm{j}}$

Here, we show the unprovability of the Herbrand consistency of $\mathrm{I}\Delta_0$ in $\mathrm{I}\Delta_0+\bigwedge\Omega_{\mathrm{j}}$. The proof is by a technique of logarithmic shortening of bounded witnesses, introduced by Z. Adamowicz in [1], and also employed in [5, 11]. The following is an outline of the proof. If $\mathrm{I}\Delta_0+\bigwedge\Omega_{\mathrm{j}}\vdash\mathrm{HCon}(\mathrm{I}\Delta_0)$, then there is an $\boldsymbol{m}\geqslant 2$ such that

$$\text{(3)}\qquad \mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}}\vdash\mathrm{HCon}(\mathrm{I}\Delta_0).$$

From now on fix this $\boldsymbol{m}$.We first show that one cannot always logarithmically shorten the witness of a bounded formula inside $\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}}$. Or in other words, for any cut (i.e. a definable initial segment) like I and its logarithm $J=\{\log x|x\in I\}$, there exists a bounded formula $\eta(x)$ such that the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x\in I\eta(x)$ is consistent, but the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x\in J\eta(x)$ is not consistent; or in other words we have $\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}}\vdash\forall x\in J\neg\eta(x)$ and $\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}}\nvdash\forall x\in I\neg\eta(x)$. For a similar statement on $\mathrm{I}\Delta_0+\Omega_1$ see Theorem 5.36 of [4]. Second we show that, under the assumption (3) above, for any bounded $\theta(x)$, if the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x\in I\theta(x)$ is consistent, then so is $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x\in J\theta(x)$. This immediately contradicts (3). The first theorem is a classical result in the theory of bounded arithmetic, which can be proved without using the assumption (3). The second theorem uses the assumption (3) to be able to logarithmically shorten a witness $\alpha\in I\wedge\theta(\alpha)$ for the formula $x\in I\wedge\theta(x)$ in a model $\mathcal{M}\models(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x\in I\theta(x)$ by constructing a model $\mathcal{N}\models(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x\in J\theta(x)$. And for that we will use the assumption (3) to infer $\mathcal{M}\models\mathrm{HCon}(\mathrm{I}\Delta_0)$, which implies the existence of an $\mathrm{I}\Delta_0-$evaluation on any set of terms Λ for which $\omega_1(\ulcorner\Lambda\urcorner)$ exists. That evaluation on a suitable Λ will give us a model of $\mathrm{I}\Delta_0+\exists x\in J\theta(x)$ (see Lemma 2.8). Then by a trick of [5] we will construct a model for $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x\in J\theta(x)$. The suitable set of terms Λ should contain a term for representing α and all the polynomials (i.e. arithmetical terms) of α. Define the terms $\underline{j}$'s by induction: $\underline{0}=0$, and $\underline{j+1}=\mathfrak{s}(\underline{j})$. The term $\underline{j}$ represents the (standard or non-standard) number j. We require the inclusion $\Lambda\supseteq\{\underline{j}|j\leqslant\omega_1(\alpha)\}=\digamma$. The code of $\digamma$ is bounded above by $\ulcorner\digamma\urcorner\leqslant\mathcal{P}\left(\prod_{j=0}^{j=\omega_1(\alpha)}2^j\right)\leqslant\mathcal{P}\left(\exp(\omega_1(\alpha)^2)\right)$. And the value $\omega_2(\ulcorner\digamma\urcorner)$ is bounded above by

$$\omega_2(\ulcorner\digamma\urcorner)\leqslant\mathcal{P}\left(\omega_2(\exp(\omega_1(\alpha)^2))\right)\leqslant\mathcal{P}\left(\exp(\omega_1(\omega_1(\alpha)^2))\right)\leqslant\mathcal{P}\left(\exp^2\left(4(\log\alpha)^4\right)\right).$$

Thus, by Theorem 2.10, for some non-standard j the (code of the) set $\Lambda=\digamma^{\langle j\rangle}$ exists, and the value $\omega_1(\ulcorner\Lambda\urcorner)=\omega_1(\ulcorner\digamma^{\langle j\rangle}\urcorner)$ is bounded above by

$$\omega_1\left(\omega_2(\ulcorner\digamma\urcorner)\right)\leqslant\mathcal{P}\left(\omega_1(\exp^2\left(4(\log\alpha)^4\right))\right)\leqslant\mathcal{P}\left(\exp^2\left(8(\log\alpha)^4\right)\right).$$

DEFINITION 3.5
Let the cut $\mathcal{I}$ be defined by $\mathcal{I}=\{x\mid\exists y[y=\exp^2\left(8(\log\alpha)^4\right)]\}$ and let $\mathcal{J}$ be the logarithm of its elements: $\mathcal{J}=\{x\mid\exists y[y=\exp^2\left(8\alpha^4\right)]\}$. ◇

Note that $\forall x[\exp(x)\in\mathcal{I}\iff x\in\mathcal{J}]$. The two mentioned theorems are the following.

THEOREM 3.6
There exists a bounded formula $\eta(x)$ such that the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{I}\eta(x)$ is consistent, but the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{J}\eta(x)$ is not consistent.

THEOREM 3.7
If $\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}}\vdash\mathrm{HCon}(\mathrm{I}\Delta_0)$, then for any bounded formula $\theta(x)$, the consistency of the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{I}\theta(x)$ implies the consistency of $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{J}\theta(x)$.

Having proved the theorems below, we conclude our main result.

COROLLARY 3.8
For any $m{\in}\mathbb{N}$, $\mathrm{I}\Delta_0+\Omega_m\nvdash\mathrm{HCon}(\mathrm{I}\Delta_0)$; thus $\mathrm{I}\Delta_0+\bigwedge\Omega_{\mathrm{j}}\nvdash\mathrm{HCon}(\mathrm{I}\Delta_0)$.

We have already proved Theorem 3.6, which is an interesting theorem in its own right.

PROOF (OF THEOREM 3.6.). The proof is rather long and we will sketch the main ideas, cf. the proof of Theorem 5.36 in [4]. We will follow [1] here. If the theorem does not hold, then for *any* bounded formula $\theta(x)$, the consistency of the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{I}\theta(x)$ will imply the consistency of $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{J}\theta(x)$. Now let $\psi(x)$ be a bounded formula such that $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{I}\psi(x)$ is consistent. Then $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{J}\psi(x)$ is consistent also. The formula $\exists x{\in}\mathcal{J}\psi(x)$ is equivalent to $\exists y{\in}\mathcal{I}\psi'(y)$ where $\psi'(y)=\exists x{\leqslant}y(y{=}\exp(x)\wedge\psi(x))$ is clearly a bounded formula. So, the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists y{\in}\mathcal{I}\psi'(y)$ is consistent, and by the assumption, the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists y{\in}\mathcal{J}\psi'(y)$ must be consistent too. Again $\exists y{\in}\mathcal{J}\psi'(y)$ is equivalent to $\exists z{\in}\mathcal{I}\exists x{\leqslant}z(z{=}\exp^2(x)\wedge\psi(x))$. Continuing this way, we infer that the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists u{\in}\mathcal{I}\exists x{\leqslant}u(u{=}\exp^k(x)\wedge\psi(x))$ is consistent for any natural $k{\in}\mathbb{N}$. Let $\mathfrak{b}$ be a constant symbol. By the above argument, the theory

$$(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\{\exists z[z=\exp^k(\mathfrak{b})\wedge\psi(\mathfrak{b})]\mid k\in\mathbb{N}\}$$

is finitely consistent, and whence it is consistent. Thus there exists a model $\mathcal{K}\models\mathrm{I}\Delta_0$ such that for some element $b{\in}\mathcal{K}$, $\mathcal{K}\models\exists z[z=\exp^k(b)\wedge\psi(b)]$ for any $k{\in}\mathbb{N}$. The initial segment $\mathcal{M}$ of $\mathcal{K}$ determined by $\{a{\in}\mathcal{K}\mid\exists k{\in}\mathbb{N}{:}a{\leqslant}\exp^k(b)\}=\exp^{\mathbb{N}}(b)$ is a model of $\mathrm{I}\Delta_0+\mathrm{Exp}$ for which $\mathcal{M}\models\psi(b)$. Thus the theory $(\mathrm{I}\Delta_0+\mathrm{Exp})+\exists x\psi(x)$ is consistent. Hence, if the theorem is not true, then for *any* bounded formula $\psi(x)$, if the theory $(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\exists x{\in}\mathcal{I}\theta(x)$ is consistent, then $(\mathrm{I}\Delta_0+\mathrm{Exp})+\exists x\psi(x)$ is also consistent. Contrapositing this statement, we get: if for a Π_1-formula $\forall x\theta(x)$ (with bounded θ) we have $\mathrm{I}\Delta_0+\mathrm{Exp}\vdash\forall x\theta(x)$, then we must also have $\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}}\vdash\forall x{\in}\mathcal{I}\theta(x)$. Since for any $x{\in}\mathcal{I}$ the value $\exp^3(x)$ exists, and all finite applications of $\omega_{\boldsymbol{m}}$ are dominated by one use of exp, then

$$\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}}\vdash\forall x{\in}\mathcal{I}\theta(x)$$

implies that

$$\mathrm{I}\Delta_0\vdash\forall x[\exists y(y=\exp^4(x))\rightarrow\theta(x)].$$

All in all, from the falsity of the theorem we inferred that whenever

$$\mathrm{I}\Delta_0+\mathrm{Exp}\vdash\forall x\theta(x)$$

for a bonded $\theta(x)$, then

$$\mathrm{I}\Delta_0\vdash\forall x[\exists y(y=\exp^4(x))\rightarrow\theta(x)].$$

Or in other words, four times application of Exp is enough to deduce all the Π_1-theorems of the theory $\mathrm{I}\Delta_0+\mathrm{Exp}$! And this contradicts Theorem 5.36 of [4]. ■

The rest of the article will be dedicated to proving Theorem 3.7.

DEFINITION 3.9
The inverse of ω_n, denoted by $\varpi_n(x)$, is defined to be the smallest y such that the inequality $\omega_n(y) \geqslant x$ holds. The cut $\mathfrak{I}_n$ is the set $\{x \mid \exists y[y = \exp^2(\varpi_{n-1}(8x^4))]\}$. ◇

Let us note that $\mathcal{J} \subset \mathfrak{I}_n \subset \mathcal{I}$ holds for any $n > 1$. Theorem 3.7 will be proved by the help of an intermediate theorem.

THEOREM 3.10
If $\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}} \vdash \mathrm{HCon}(\mathrm{I}\Delta_0)$, then for any bounded formula $\theta(x)$, the consistency of the theory $(\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists x \in \mathcal{I}\theta(x)$ implies the consistency of $(\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists x \in \mathfrak{I}_{\boldsymbol{m}}\theta(x)$.

Having proved this, Theorem 3.7 can be proved easily:

PROOF (OF THEOREM 3.7 FROM THEOREM 3.10.). Assume $\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}} \vdash \mathrm{HCon}(\mathrm{I}\Delta_0)$. Let $\theta(x)$ be a bounded formula such that $(\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists x \in \mathcal{I}\theta(x)$ is consistent. Then by Theorem 3.10, the theory $(\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists x \in \mathfrak{I}_{\boldsymbol{m}}\theta(x)$ is consistent too. Let $\theta'(y)$ be the bounded formula

$$\theta'(y) = \exists x \leqslant y[8x^4 \leqslant \omega_{\boldsymbol{m}-1}(8(\log y)^4) \wedge \theta(x)];$$

then $\exists x \in \mathfrak{I}_{\boldsymbol{m}}\theta(x)$ is equivalent to $\exists y \in \mathcal{I}\theta'(y)$. Now, since $(\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists y \in \mathcal{I}\theta'(y)$ is consistent, again by Theorem 3.10, the theory $(\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists y \in \mathfrak{I}_{\boldsymbol{m}}\theta'(y)$ must be consistent. Now we note that

$$(y \in \mathfrak{I}_{\boldsymbol{m}}) \wedge [8x^4 \leqslant \omega_{\boldsymbol{m}-1}(8(\log y)^4)] \Rightarrow (x \in \mathcal{J})$$

holds for non-standard x and y, because $\omega^2_{\boldsymbol{m}-1}(8[\log y]^4) < 8y^4$. So, $(\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists x \in \mathcal{J}\theta(x)$ must be consistent too. ■

For proving Theorem 3.10, we assume that for the bounded formula $\theta(x)$ there exists a model $\mathcal{M}$ such that

$$\mathcal{M} \models (\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + (\alpha \in \mathcal{I} \wedge \theta(\alpha)) \tag{4}$$

holds for some non-standard $\alpha \in \mathcal{M}$. We will construct another model

$$\mathcal{N} \models (\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}) + \exists x \in \mathfrak{I}_{\boldsymbol{m}}\theta(x).$$

Define the terms $\underline{j}$'s by induction: $\underline{0} = 0$, and $\underline{j+1} = \mathfrak{s}(\underline{j})$. The term $\underline{j}$ represents the (standard or non-standard) number j. Let $\mathfrak{q}$ be the Skolem function symbol for the formula $\exists y(y \leqslant x^2 \wedge y = x^2)$ and $\mathfrak{c}$ be the Skolem constant symbol for the sentence $\exists x\big(\exists w(w \leqslant x^2 \wedge w = x^2) \wedge \forall v(v \not\leqslant (\mathfrak{s}x)^2 \wedge v \neq (\mathfrak{s}x)^2)\big)$, and let $\Upsilon = \{0, 0+0, 0^2, \mathfrak{c}, \mathfrak{c}^2, \mathfrak{c}^2+0, \mathfrak{s}\mathfrak{c}, \mathfrak{q}\mathfrak{c}, (\mathfrak{s}\mathfrak{c})^2, (\mathfrak{s}\mathfrak{c})^2+0\}$ (see Example 2.4). Define the terms z_i's inductively: $\mathsf{z}_0 = \underline{2}$, and $\mathsf{z}_{j+1} = \mathfrak{q}(\mathsf{z}_j)$. Since we have $\mathfrak{q}(x) = x^2$ in $\mathrm{I}\Delta_0$−evaluations (by Example 2.4), then z_i will represent $\exp^2(i)$ (by induction on i). Take $\Lambda = \Upsilon \cup \{\underline{j} \mid j \leqslant \omega_1(\alpha)\} \cup \{\mathsf{z}_j \mid j \leqslant 8\alpha^4\}$; then $\omega_2(\ulcorner\Lambda\urcorner)$ is of order $\exp^2\big(4(\log\alpha)^4\big)$ which exists by the assumption $\mathcal{M} \models \alpha \in \mathcal{I}$ (see (4) above). Whence, for some non-standard j the set $\Lambda^{\langle j \rangle}$ exists (in $\mathcal{M}$), and moreover the value $\omega_1(\ulcorner\Lambda^{\langle j \rangle}\urcorner)$ exists because by (the proof of) Theorem 2.10, $\omega_1(\ulcorner\Lambda^{\langle j \rangle}\urcorner) \leqslant \omega_1(\exp^2\big(4(\log\alpha)^4\big)) \leqslant \exp^2\big(8(\log\alpha)^4\big)$, and $\alpha \in \mathcal{I}$. Since by the assumptions (3) and (4) we have $\mathcal{M} \models \mathrm{HCon}(\mathrm{I}\Delta_0)$, then there must exist an $\mathrm{I}\Delta_0$−evaluation $p \in \mathcal{M}$ on this $\Lambda^{\langle j \rangle}$. Now, we can build the model $\mathcal{K} := \mathfrak{M}(\Lambda, p)$.

LEMMA 3.11
With the above assumptions, $\mathcal{K} \models \theta(\underline{\alpha}/p)$.

After proving this lemma, we can finish the proof of Theorem 3.10.

PROOF (OF THEOREM 3.10 FROM LEMMA 3.11.). By Lemma 2.8 we already have $\mathcal{K}\models \mathrm{I}\Delta_0$, and by Lemma 3.11, $\mathcal{K}\models\theta(\underline{\alpha}/p)$. Also $\underline{\alpha}/p\in\mathcal{J}^{\mathcal{K}}$ by the existence of z_i/p's ($\mathcal{K}\models\mathsf{z}_{8\alpha^4}/p=\exp^2(8[\underline{\alpha}/p]^4)$). Whence $\mathcal{K}\models\alpha/p\in\mathcal{J}\wedge\theta(\underline{\alpha}/p)$. By Lemma 2.6 there exists some (non-standard) element $\beta\in\mathcal{K}$ such that the inequalities $\omega_{\boldsymbol{m}}^{\mathbb{N}}(\beta)<\mathsf{z}_{8\alpha^4}/p\leqslant\omega_{\boldsymbol{m}+1}(\beta)$ hold. Now, let $\mathcal{N}$ be the initial segment of $\mathcal{K}$ determined by $\omega_{\boldsymbol{m}}^{\mathbb{N}}(\beta)$, i.e. $\mathcal{N}=\{x\in\mathcal{K}\mid\exists k\in\mathbb{N}:x<\omega_{\boldsymbol{m}}^{k}(\beta)\}$. Then, for this model $\mathcal{N}$ we have that $\mathcal{N}\models(\mathrm{I}\Delta_0+\Omega_{\boldsymbol{m}})+\theta(\underline{\alpha}/p)$, and all we have to show is that $\mathcal{N}\models\underline{\alpha}/p\in\mathfrak{I}_{\boldsymbol{m}}$. First note that $\beta\in\mathcal{N}$, and second that $\exp^2(8[\underline{\alpha}/p]^4)\leqslant\omega_{\boldsymbol{m}+1}(\beta)$ implies $8[\underline{\alpha}/p]^4\leqslant\omega_{\boldsymbol{m}-1}(2\log^2\beta)$, and so we have $\varpi_{\boldsymbol{m}-1}(8[\underline{\alpha}/p]^4)\leqslant 2\log^2\beta$. Thus $\exp^2(\varpi_{\boldsymbol{m}-1}(8[\underline{\alpha}/p]^4))$ exists ($\leqslant\omega_1(\beta)$), and so $[\underline{\alpha}/p]\in\mathfrak{I}_{\boldsymbol{m}}$ holds. ∎

Finally, it remains (only) to prove Lemma 3.11. This is exactly Corollary 35 of [11]; and the reader is invited to consult it for more details. Here a sketch of the proof, for the sake of self-containedness, is presented.

PROOF (OF LEMMA 3.11 – A SKETCH.). Since $\theta(x)\in\Delta_0$ and $\mathcal{M}\models\theta(\alpha)$, we note that the range of the quantifiers of $\theta(\alpha)$ is the set $\{x\in\mathcal{M}\mid x\leqslant t(\alpha)\ \text{for some}\ \mathcal{L}_A\text{–term}\ t\}$. This set is the initial segment of $\mathcal{M}$ determined by $\alpha^{\mathbb{N}}$; denote it by $\mathcal{M}'$. We have $\mathcal{M}'\models\theta(\alpha)$. For any $j\in\alpha^{\mathbb{N}}$ we have the corresponding $\underline{j}\in\Lambda$, and thus $\underline{j}/p\in\mathcal{K}$. So, this suggests a correspondence between $\alpha^{\mathbb{N}}$ and the initial segment of $\mathcal{K}$ determined by $(\underline{\alpha}/p)^{\mathbb{N}}$ which we denote it by $\mathcal{K}'$. It suffices to show that this correspondence exists and is an isomorphism between $\mathcal{M}'$ and $\mathcal{K}'$. Because, then we will have $\mathcal{K}'\models\theta(\underline{\alpha}/p)$ which will immediately imply $\mathcal{K}\models\theta(\underline{\alpha}/p)$; our desired conclusion.

We first note that $\mathcal{M}'=\{t(i_1,\ldots,i_n)\mid i_1,\ldots,i_n\leqslant\alpha\ \&\ t\text{ is an }\mathcal{L}_A\text{–term}\}$. This follows from a more general fact:

(5) If for some model $\mathfrak{A}\models\mathrm{I}\Delta_0$ and $x,a_1,\ldots,a_n\in\mathfrak{A}$ we have $\mathfrak{A}\models x\leqslant t(a_1,\ldots,a_n)$ for an $\mathcal{L}_A$–term t, then there are some $b_1,\ldots,b_m\in\mathfrak{A}$ and some $\mathcal{L}_A$–term s such that $\mathfrak{A}\models x=s(b_1,\ldots,b_m)$; moreover $\max b_j\leqslant\max a_i$.

This can be proved by induction on the complexity of t. For $t=t_1+t_2$, distinguish two cases: (i) if $\mathfrak{A}\models x\leqslant t_1(\overline{a})$, where $\overline{a}$ is a shorthand for $(a_1,\ldots,a_n)$, then we are done by the induction hypothesis; (ii) if $\mathfrak{A}\models t_1(\overline{a})\leqslant x$ then there exists some $y\in\mathfrak{A}$ such that $\mathfrak{A}\models[x=t_1(\overline{a})+y]\wedge[y\leqslant t_2(\overline{a})]$, and the result follows from the induction hypothesis. For $t=t_1\cdot t_2$, there are some $q,r\in\mathfrak{A}$ for which we have $\mathfrak{A}\models[x=t_1(\overline{a})\cdot q+r]\wedge[r<t_1(\overline{a})]\wedge[q\leqslant t_2(\overline{a})]$. Two uses of induction hypothesis (for the terms t_1 and t_2) will finish the proof.

Second, we note that $\mathcal{K}\models\mathrm{I}\Delta_0$ by Lemma 2.8, and so $\mathcal{K}'\models\mathrm{I}\Delta_0$, whence by (5) above we can write $\mathcal{K}'=\{t(u_1,\ldots,u_n)\mid u_1,\ldots,u_n\leqslant\underline{\alpha}/p\ \&\ t\text{ is an }\mathcal{L}_A\text{–term}\}$. And it can be proved by induction on j that if $\mathcal{K}'\models u\leqslant\underline{j}/p$ (or equivalently $\mathcal{M}\models$ '$p\models u\leqslant j$') there there exists an $l\leqslant j$ (in $\mathcal{M}$) such that $\mathcal{K}'\models u=\underline{l}/p$ (or equivalently $\mathcal{M}\models$ '$p\models u=l$'). Whence, we can present $\mathcal{K}'$ as

$$\mathcal{K}'=\{t(\underline{i_1}/p,\ldots,\underline{i_n}/p)\mid i_1,\ldots,i_n\leqslant\alpha\ \&\ t\text{ is an }\mathcal{L}_A\text{–term}\}.$$

Thus a correspondence by $t(i_1,\ldots,i_n)\mapsto t(\underline{i_1}/p,\ldots,\underline{i_n}/p)$ exists between the two $\mathrm{I}\Delta_0$– models $\mathcal{M}'$ and $\mathcal{K}'$. That this mapping preserves atomic formulas of the form $u=v$ for terms u,v follows from the axioms of Q (the inductive definitions of addition and multiplication). It also preserves atomic formulas of the form $u\leqslant v$ because we have in Q that $u\leqslant v\leftrightarrow w+u=v$ for some $w\leqslant v$. The preservation of negated atomic formulas follows from the $\mathrm{I}\Delta_0$–derivable equivalences $x\neq y\leftrightarrow\mathfrak{s}y\leqslant x\vee\mathfrak{s}x\leqslant y$, and $x\not\leqslant y\leftrightarrow\mathfrak{s}y\leqslant x$. Thus the above mapping is an isomorphism. ∎

4 Conclusions

We saw one example of the provability of Herbrand consistency of a theory S in a (super-)theory (of it) T (Theorem 3.2 for $S = \mathrm{I}\Delta_0 + \Omega_{\mathrm{m}}, T = \mathrm{I}\Delta_0 + \mathrm{Exp}$) and one example of the unprovability of Herbrand consistency of S in T (Corollary 3.8 for $S = \mathrm{I}\Delta_0, T = \mathrm{I}\Delta_0 + \bigwedge \Omega_{\mathrm{j}}$). The main point common in both of the results was that, if every Skolem term of S has an evaluation in T, then T may prove the Herbrand consistency of S; but if there are some Skolem terms of S which grow too fast for T to catch them, then T could not be able to derive the Herbrand consistency of S. This is not a general law, but a rule of thumb. Note that in our proof of Corollary 3.8, the terms z_i had the code of order $\exp(i)$ but the value of $\exp^2(i)$. And the theory $\mathrm{I}\Delta_0 + \bigwedge \Omega_{\mathrm{j}}$ cannot catch the value of $\exp^2(i)$ by having the code $\exp(i)$; the gap is of exponential order. And in our proof of Theorem 3.2, the theory $\mathrm{I}\Delta_0 + \mathrm{Exp}$ could evaluate all the Skolem terms of $\mathrm{I}\Delta_0 + \Omega_{\mathrm{m}}$. A very similar argument can show that $\mathrm{I}\Delta_0 + \mathrm{Sup\_Exp} \vdash \mathrm{HCon}(\mathrm{I}\Delta_0 + \mathrm{Exp})$. An open question, asked by L. A. Kołodziejczyk, is whether showing the unprovability of Herbrand consistency is possible without making use of fast-growing terms. More explicitly, if bounded formulas are required to have only variables in their bounds, and the re-axiomatization of $\mathrm{I}\Delta_0$ by the induction scheme $\forall y\big(\theta(0) \wedge \forall x < y[\theta(x) \rightarrow \theta(\mathfrak{s}x)] \rightarrow \forall x \leqslant y \theta(x)\big)$ is taken into account, then is it possible to show the unprovability of the Herbrand consistency of (this axiomatization of) $\mathrm{I}\Delta_0$ in itself? Note that here having terms like z_i's with double exponential values could not be possible.

The proof of our main result (Corollary 3.8) is very similar to the proof of the main result of [11]—the unprovability $\mathrm{I}\Delta_0 \nvdash \mathrm{HCon}(\mathrm{I}\Delta_0)$. A major difference was the technique of Theorem 3.10 for constructing a model of $\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}$ from a model of $\mathrm{I}\Delta_0$, for which Lemma 2.6 was used. The idea of this technique is taken from [5]; note that the proof of our Theorem 3.6 is different from the proof of the corresponding theorem in [5], in that we had fixed one $\boldsymbol{m}$ and (instead) followed the lines of the corresponding proof in [1]. That way we did not need to show the theorem for the theory $\mathrm{I}\Delta_0 + \bigwedge \Omega_{\mathrm{j}}$, and instead a simplified proof of the theorem for $\mathrm{I}\Delta_0 + \Omega_{\boldsymbol{m}}$ in [1] would suffice for us. Let us note that the corresponding theorem in [5] is somehow stronger: $\mathrm{I}\Delta_0 + \bigwedge \Omega_{\mathrm{j}} \nvdash \mathrm{HCon}(\mathrm{S} + \Omega_1)$ has been shown for a finite fragment $\mathrm{S} \subseteq \mathrm{I}\Delta_0$. The question of generalizing this result, as follows, was asked by the referee.

CONJECTURE 4.1
There exists a finite fragment $\mathrm{S} \subseteq \mathrm{I}\Delta_0$ such that $\mathrm{I}\Delta_0 + \bigwedge \Omega_{\mathrm{j}} \nvdash \mathrm{HCon}(\mathrm{S})$. ◇

Let us finish the article by repeating the open question asked also in [11], which is whether Gödel's Second Incompleteness Theorem for the Herbrand consistency predicate has a uniform proof for theories containing Robinson's Arithmetic Q.

QUESTION 4.2
Can a BOOK proof (in the words of Paul Erdös) of $T \nvdash \mathcal{H}\mathfrak{Con}(T)$ be given uniformly for any theory $T \supseteq \mathrm{Q}$ and a canonical definition of Herbrand consistency $\mathcal{H}\mathfrak{Con}$? ◇

Acknowledgements

This work is a final result of a long project, pursued in different places and different times. I would like to thank Prof. Albert Visser for having me in Universiteit Utrecht as a visitor in April 2006, and Prof. Gerhard Jäger for inviting me to visit Universität Bern in September 2006. I am grateful to Dr Rashid Zaare-Nahandi in IASBS for his support during my stay there in the academic year 2006–2007. I appreciate the stimulating discussions I had with Leszek Aleksander

Kołodziejczyk during the LABCC workshop in Greifswald. This research was partially supported by the National Elite Foundation of Iran (`Bonyad Melli Nokhbegan`, www.bmn.ir). The comments and remarks of the anonymous referee have greatly improved the presentation of the article; this is highly appreciated.

References

[1] Z. Adamowicz. Herbrand consistency and bounded arithmetic. *Fundamenta Mathematicae*, **171**, 279–292, 2002.

[2] J. Avigad. Number theory and elementary arithmetic. *Philosophia Mathematica*, **3**, 257–284, 2003.

[3] S. R. Buss. On Herbrand's Theorem. In *(Selected Papers from the International Workshop on) Logic and Computational Complexity*, D. Maurice and R. Leivant, eds., 1994, Vol. 960 of *Lecture Notes in Computer Science*, pp. 195–209. Springer, 1995.

[4] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic*. 2nd printing, Springer, 1998.

[5] L. A. Kołodziejczyk. On the Herbrand notion of consistency for finitely axiomatizable fragments of bounded arithmetic theories. *Journal of Symbolic Logic*, **71**, 624–638, 2006.

[6] A. MacIntyre. The issues around proving Fermat's last theorem in Peano Arithmetic. In *Abstracts of Seminar Talks Presented at the Computing Laboratory of the University of Oxford*. Available at: http://www.comlab.ox.ac.uk/seminars/128.html. (5 December 2008); http://www.comlab.ox.ac.uk/seminars/355.html. (23 October 2009).

[7] R. Parikh. Existence and feasibility in arithmetic. *Journal of Symbolic Logic*, **36**, 494–508, 1971.

[8] J. B. Paris and A. J. Wilkie. Δ_0 sets and induction. In *Proceedings of Open Days in Model Theory and Set Theory*. W. Guzicki and W. Marek and A. Plec and C. Rauszer, eds., pp. 237–248. Leeds University Press, 1981.

[9] J. B. Paris and A. J. Wilkie and A. R. Woods. Provability of the Pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic*, **53**, 1235–1244, 1988.

[10] S. Salehi. *Herbrand Consistency in Arithmetics with Bounded Induction*. PhD Dissertation, Institute of Mathematics, Polish Academy of Sciences, 2002.

[11] S. Salehi. *Herbrand Consistency of Some Arithmetical Theories*. arXiv:1005.2654v2[math.LO], 2010. Available at: http://arxiv.org/pdf/1005.2654.

[12] D. E. Willard. Passive induction and a solution to a Paris−Wilkie open question. *Annals of Pure and Applied Logic*, **146**, 124–149, 2007.

Received 1 August 2010