FUNDAMENTA
MATHEMATICAE
212 (2011)

Lower and upper bounds for the provability of Herbrand consistency in weak arithmetics

by

Zofia Adamowicz and **Konrad Zdanowski** (Warszawa)

Abstract. We prove that for $i \geq 1$, the arithmetic $\mathrm{I}\Delta_0 + \Omega_i$ does not prove a variant of its own Herbrand consistency restricted to the terms of depth in $(1+\varepsilon)\log^{i+2}$, where ε is an arbitrarily small constant greater than zero.

On the other hand, the provability holds for the set of terms of depths in $\log^{i+3}$.

1. Introduction. One of the main methods of showing that one set of axioms, say T, is strictly stronger than another one, say $S \subseteq T$, is to show that $T \vdash \mathrm{Con}_S$. However, as proved by Wilkie and Paris [WP87], this method does not work for theories of bounded arithmetic if we use the usual Hilbert style provability predicate. Indeed, they proved that even the strong arithmetic $\mathrm{I}\Delta_0 + \exp$ does not prove the Hilbert style consistency of Robinson's arithmetic Q, that is, $\mathrm{I}\Delta_0 + \exp$ does not prove that there is no Hilbert proof of $0 \neq 0$ from Q. Thus, if we hope to prove that one bounded arithmetic is stronger than another one by using consistency statements we should use some other provability notions, like tableaux or Herbrand provability. Indeed, for these notions it is usually easier to show that a given theory is consistent since, e.g., Herbrand proofs are of larger size than Hilbert ones. Thus, it may happen in a model of $\mathrm{I}\Delta_0 + \exp$ that a theory S is inconsistent in the Hilbert sense and consistent in the Herbrand sense. Only when we know that the superexponentiation function is total we can prove the equivalence of the above notions of provability. (The superexponentiation function is defined by the inductive conditions $\mathrm{supexp}(0) = 1$ and $\mathrm{supexp}(x+1) = \exp(2, \mathrm{supexp}(x))$.) For some time it has even been unknown whether the second Gödel incompleteness theorem holds for the arithmetics $\mathrm{I}\Delta_0 + \Omega_i$ and the Herbrand style provability predicate. Adamowicz and Zbierski [AZ01] proved, for $i \geq 2$, the second incompleteness theorem for $\mathrm{I}\Delta_0 + \Omega_i$ and the Herbrand notion

2010 *Mathematics Subject Classification*: Primary 03F30; Secondary 03F40.
Key words and phrases: Herbrand consistency, unprovability, weak arithmetics.

DOI: 10.4064/fm212-3-1

of consistency, and later Adamowicz [A01] proved this result for $\mathrm{I}\Delta_0 + \Omega_1$. Recently, Kołodziejczyk [K06a] showed a strengthening of these results. He proved that there is a finite fragment S of $\mathrm{I}\Delta_0 + \Omega_1$ such that no theory $\mathrm{I}\Delta_0 + \Omega_i$ proves the Herbrand consistency of S. Thus, if one wants to prove strict hierarchy of bounded arithmetics by means of provability of Herbrand consistency one should consider a thinner notion, e.g., Herbrand proofs restricted to some definable cuts of a given model of a bounded arithmetic. Such a study is the main subject of our paper.

Now, we state our main result. Let $\log^n$ be the set of elements a such the nth iteration of exp on a exists. If exp is not provably total in T then there are models of T in which not all elements are in $\log^n$. For $C > 0$, $C\log^n$ is the set of elements a such that there exists b in $\log^n$ such that a is less than or equal to Cb. Let us observe that the above notions are definable by existential formulas.

We consider Herbrand consistency statements parametrized by a standard integer N. This parameter tells us how good a Herbrand evaluation is on a given set of terms. Namely, the truth value of formulas with codes less than N has to be decided. We show that for some fixed N, for each $i \geq 1$, $\mathrm{I}\Delta_0 + \Omega_i$ does not prove its Herbrand consistency, restricted to the terms of depth in $(1+\varepsilon)\log^{i+2}$, where $\varepsilon > 0$ (see Theorem 4.12). That is, for some N, for each $i \geq 1$ and each $\varepsilon > 0$,

$$\mathrm{I}\Delta_0 + \Omega_i \nvdash \mathrm{HCons}(N, \mathrm{I}\Delta_0 + \Omega_i, (1+\varepsilon)\log^{i+2}).$$

On the other hand it may be proved by standard methods that for each i,

$$\text{for each } N, \quad \mathrm{I}\Delta_0 + \Omega_i \vdash \mathrm{HCons}(N, \mathrm{I}\Delta_0 + \Omega_i, \log^{i+3}),$$

that is, $\mathrm{I}\Delta_0 + \Omega_i$ proves its Herbrand consistency restricted to terms of depth $\log^{i+3}$ (see Theorem 3.2).

It is tempting to close the gap by proving, at least for some $i \geq 1$, either that

$$\text{for each } N, \quad \mathrm{I}\Delta_0 + \Omega_i \vdash \mathrm{HCons}(N, \mathrm{I}\Delta_0 + \Omega_i, \log^{i+2}) \tag{1.1}$$

or

$$\mathrm{I}\Delta_0 + \Omega_i \nvdash \mathrm{HCons}(N, \mathrm{I}\Delta_0 + \Omega_i, A\log^{i+3}) \quad \text{for some } N, A \in \mathbb{N}. \tag{1.2}$$

Indeed both conjectures (1.1) and (1.2) have interesting consequences for bounded arithmetics. If (1.1) holds then $\mathrm{I}\Delta_0 + \Omega_{i+1}$ would not be Π_1-conservative over $\mathrm{I}\Delta_0 + \Omega_i$. This is so because $\log^{i+2}$ is closed under addition in the presence of Ω_{i+1}. Thus, in $\mathrm{I}\Delta_0 + \Omega_{i+1}$ the cuts $\log^{i+2}$ and $(1+\varepsilon)\log^{i+2}$ are the same. It would then follow from (1.1) that $\mathrm{I}\Delta_0 + \Omega_{i+1} \vdash \mathrm{HCons}(\mathrm{I}\Delta_0 + \Omega_i, A\log^{i+2})$ for each $A \in \mathbb{N}$.

On the other hand, if (1.2) holds this would mean that we cannot mimic the proof of Theorem 3.2 for the cut $A\log^{i+3}$. But the only tool needed in

that proof which is unavailable in this situation is the existence of a suitable truth definition for Δ_0 formulas. It would follow that there is no such truth definition for Δ_0 formulas whose suitable properties are provable in $\mathrm{I}\Delta_0+\Omega_i$. This is related to a major open problem in bounded arithmetics: how much exponentiation is needed for a truth definition for bounded formulas.

The paper is structured as follows. In Section 2 we introduce all the main notions needed to define Herbrand consistency, which is the main object of our study. In that section we present and prove the main technical facts about these notions. In Section 3 we show which cases of Herbrand consistency are indeed provable. Then, in Section 4 we use the tools assembled in Section 2 to prove our main result about unprovability of Herbrand consistency.

2. Basic notions and facts. In this section we present some basic notions and we prove their properties which will be used throughout the paper. We state some properties of coding of terms and formulas. Then we define evaluations on a set of terms and describe how an evaluation can be used to define a model. We also state some combinatorial properties of evaluations like the Estimation Lemma. They will be used in proving our main result but the reader may want to skip some combinatorial arguments in this section on a first reading. We end this section with a suitable definition of Herbrand consistency.

For a detailed treatment of bounded arithmetics we refer to [HP93]. We consider bounded arithmetics theories $\mathrm{I}\Delta_0+\Omega_i$, for $i \geq 1$. $\mathrm{I}\Delta_0$ is just the first order arithmetic with the induction axioms restricted to bounded formulas, i.e. formulas with quantification of the form $Qx \leq t(\bar{z})$, where $Q \in \{\exists, \forall\}$, $x \notin \{\bar{z}\}$ and t is a term in a language of $\mathrm{I}\Delta_0 + \Omega_i$ (that is, using only $+, \times$ and ω_i function symbols). For $i \geq 1$, the axiom Ω_i states the totality of the function ω_i. The functions ω_i are defined as follows. Let $\log(x)$ be the logarithm with base 2. Let $\mathrm{lh}(x)$ be the length of the binary representation of x,

$$\mathrm{lh}(x) = \ulcorner \log(x+1) \urcorner.$$

Now,

$$\omega_1(x) = \begin{cases} 0 & \text{if } x = 0, \\ 2^{(\mathrm{lh}(x)-1)^2} & \text{if } x > 0, \end{cases}$$

$$\omega_{i+1}(x) = \begin{cases} 0 & \text{if } x = 0, \\ 2^{\omega_i(\mathrm{lh}(x)-1)} & \text{if } x > 0. \end{cases}$$

Let $\exp(x) = 2^x$. The following relation between exp and ω_i will be important for us: for all $i \geq 1$ and all k,

$$\omega_i^k(\exp^{i+2}(0)) = \exp^{i+2}(k). \tag{2.1}$$

This equality allows us to infer the existence of the $(i+2)$th iterated exp on

a number k from the existence of an interpretation for a term $\omega_i^k(\exp^{i+2}(0))$. We will also need the $\mathrm{supexp}(x)$ function defined by the conditions $\mathrm{supexp}(0) = 1$ and $\mathrm{supexp}(x+1) = \exp(\mathrm{supexp}(x))$, and the $\log^*(x)$ function, which is a kind of inverse of supexp, defined as

$$\log^*(x) = \max(\{i \le x : \mathrm{supexp}(i) \le x\} \cup \{0\}).$$

We extend the language by adding a function symbol $s^{\exists x\,\varphi}$ of arity n for each formula $\exists x\,\varphi$ with n free variables.

We identify terms and formulas with their Gödel numbers. A numeral for a number i is denoted $\underline{i}$ and its Gödel number is assumed to be $2i$. We take the *tree depth* of $\underline{i}$ to be $\log^2(i)$. The tree depth of other terms is defined by the inductive condition

$$\mathrm{tr}(f(t_1,\dots,t_k)) = 1 + \max\{\mathrm{tr}(t_i) : i \le k\}.$$

By the *depth* of a term t we define the maximum of its tree depth and the size of the greatest function symbol in t, that is,

$$\mathrm{dp}(t) = \max(\{f : f \text{ occurs in } t\} \cup \{\mathrm{tr}(t)\}).$$

For a set of terms Λ, the depth of Λ is $\mathrm{dp}(\Lambda) = \max\{\mathrm{dp}(t) : t \in \Lambda\}$.

It may seem arbitrary to define the tree depth of $\underline{i}$ to be $\mathrm{tr}(\underline{i}) = \log^2(i)$, especially when we recall that the tree depth of the usual binary representation of i is $\log(i)$. However, we can construct a canonical term $\underline{i}$ denoting a number i of tree depth $O(\log^2(i))$. Indeed, if for each number $k \le i$ there is a term $\underline{k}$ of tree depth x then each number $r < i^2$ may be decomposed as $r = ki + m$ with $k, m < i$ and we can write the term for r as $\underline{r} = \underline{k} \cdot \underline{i} + \underline{m}$ of tree depth $x+2$. This leads to a recursive dependence, that terms for numbers less than 2^{2^n} can be written with tree depth $2n = 2\log^2(2^{2^n})$. For simplicity of arguments we define the tree depth of a term $\underline{i}$ to be just $\log^2(i)$.

We do not fix one particular coding method. Indeed, any usual, efficient coding is good for our purpose. We only state some properties which we require from a coding.

We assume that if φ' is a subformula of φ then the length of the code of φ' is not greater than that of φ, $\mathrm{lh}(\varphi') \le \mathrm{lh}(\varphi)$. We assume that a code of a term $s^{\varphi}(t_1,\dots,t_k)$ is not greater than $(\varphi \prod_{i\le k} t_i)^{O(1)}$. The last expression should be read simply as a product of numbers coding the formula φ and terms t_i, for $i \le k$. Let us remark that usual efficient codings possess this property. Indeed, for the length of a term $t = s^{\varphi}(t_1,\dots,t_k)$ we have

$$\mathrm{lh}(\varphi) + \sum_{i\le k} \mathrm{lh}(t_i) \le \mathrm{lh}(t) \le A\Big(\mathrm{lh}(\varphi) + \sum_{i\le k} \mathrm{lh}(t_i)\Big) \tag{2.2}$$

for some integer $A \in \mathbb{N}$. Thus, $t \leq 2^{A(\mathrm{lh}(\varphi)+\sum_{i\leq k}\mathrm{lh}(t_i))}$ and

$$t \leq \Big(\varphi \prod_{i\leq k} t_i\Big)^A. \tag{2.3}$$

Later we will refer to the constant A from the above formulas. However, note that the precise value of A depends on the coding method one uses.

An *evaluation* p on a set of terms Λ is a boolean function from Λ^2 into $\{0,1\}$. A given evaluation p tells us which terms are equal under p. That is, t equals t' under p if $p(t,t') = 1$. For an evaluation p on Λ we define a model $M(\Lambda, p)$. The equality relation of $M(\Lambda, p)$ is given by $p(t,t') = 1$. Then, for a function symbol f and terms $t_1, \ldots, t_k$ the value of f on $t_1, \ldots, t_k$ is just $f(t_1, \ldots, t_k)$. We define the ordering in $M(\Lambda, p)$ as: $t \leq t'$ if and only if there is $s \in \Lambda$ such that $p(t+s, t') = 1$. Thus, we adopt the standard method of defining the ordering relation.

Let us observe that $M(\Lambda, p)$ is a well defined model if and only if Λ has the property that if f is a function symbol in our language and $t_1, \ldots, t_{\mathrm{ar}(f)} \in \Lambda$, then $f(t_1, \ldots, t_{\mathrm{ar}(f)}) \in \Lambda$. To have the equality relation well defined we have to require that the relation on terms given by $p(t,t') = 1$ is reflexive, symmetric and transitive. Moreover, it should be a congruence relation with respect to the operation of applying a function symbol, that is, for each $t_1, \ldots, t_n$ and $s_1, \ldots, s_n$ and for each n-ary function symbol f,

$$\text{if for each } i \leq n,\ p(t_i, s_i) = 1 \text{ then } p(f(t_1, \ldots, t_n), f(s_1, \ldots, s_n)) = 1.$$

We assume that all the evaluations to be considered satisfy the above conditions.

Let us note that, assuming some coding, all the above notions are expressible in arithmetic. Let $M \models \mathrm{I}\Delta_0$. We will say that Λ is a *set of terms in* M if Λ is an element of M which satisfies the definition of being a set of terms. Since such a definition may be written as a Δ_0 formula, say $\varphi(x)$, we have $M \models \varphi(\Lambda)$. We use the same convention to state that an element $p \in M$ is an evaluation on Λ, that is, p and Λ satisfy in M a Δ_0 arithmetical formula which states that x is an evaluation of y.

If Λ is a set of terms in M, it may contain some elements which are not terms in our sense. Nevertheless, it is possible to treat all elements of Λ as terms by using the satisfaction relation of M. E.g. if $(2a) \in \Lambda$ for some nonstandard a, then $2a$ can be thought of as a numeral $\underline{a}$ denoting a.

Now, let $\Lambda \in M$ be a set of terms in M. For $I \subseteq M$ we denote by $\Lambda{\restriction}I$ the subset of terms from Λ with depths in I, that is,

$$\Lambda{\restriction}I = \{t \in \Lambda : \mathrm{dp}(t) \in I\}.$$

If I is a cut in M (that is, I is downward closed and closed under successors) then $M(\Lambda{\restriction}I, p)$ is a well defined model (where we also restrict the evaluation

p to $\Lambda\restriction I$). Indeed, $M(\Lambda\restriction I, p)$ is closed under the operations of the arithmetical signature. If terms t_1, t_2 are in $\Lambda\restriction I$ then the result of adding t_1 and t_2 is the term $t_1 + t_2$. Its depth $\mathrm{dp}(t_1 + t_2) = \max\{\mathrm{dp}(t_1), \mathrm{dp}(t_2)\} + 1$ is in I, provided that $\mathrm{dp}(t_1), \mathrm{dp}(t_2) \in I$. Thus $M(\Lambda\restriction I, p)$ is closed under addition (and the same argument applies to multiplication, successor and ω_i functions). Let us observe that $\Lambda\restriction I$ is not an element of M even if Λ is.

In the definition below and in the rest of this article we deal with formulas in prenex normal form only. Thus, if we write $\neg\varphi$ for φ in prenex normal form we assume that negation is pushed into the quantifier free part of φ using the rules $\neg\exists x\,\gamma \equiv \forall x\,\neg\gamma$ and $\neg\forall x\,\gamma \equiv \exists x\,\neg\gamma$.

Definition 2.1. For a formula $\varphi(x_1, \ldots, x_n)$ and terms $t_1, \ldots, t_n$ we define a sequence of terms $s_1, \ldots, s_r$ by induction on the complexity of φ.

If φ is quantifier free then the sequence for φ and $t_1, \ldots, t_n$ is just $t_1, \ldots, t_n$.

If $\varphi = \exists x\,\psi(x_1, \ldots, x_n, x)$ then the sequence for φ and $t_1, \ldots, t_n$ is the sequence for $\psi(x_1, \ldots, x_n, x)$ and terms $t_1, \ldots, t_n, s^{\exists x\,\psi(x_1,\ldots,x_n,x)}(t_1, \ldots, t_n)$.

If $\varphi = \forall x\,\psi(x_1, \ldots, x_n, x)$ then the sequence for φ and $t_1, \ldots, t_n$ is the sequence for $\neg\psi$ and terms $t_1, \ldots, t_n, s^{\exists x\,\neg\psi(x_1,\ldots,x_n,x)}(t_1, \ldots, t_n)$.

We then call $s_1, \ldots, s_r$ the *terms needed to evaluate* $\varphi(x_1, \ldots, x_n)$ *on* $t_1, \ldots, t_n$ or just the *sequence of terms for* φ and $t_1, \ldots, t_n$.

As we will see, the definition of the relation $p \models \varphi[t_1, \ldots, t_n]$, in order to work properly, requires all terms $s_1, \ldots, s_r$ to be in Λ.

Definition 2.2. Let $t_1, \ldots, t_n \in \Lambda$ and $\varphi(x_1, \ldots, x_n)$ be a formula. We say that $(\varphi, t_1, \ldots, t_n)$ is *good enough* (g.e. for short) *for* Λ if all terms from the sequence for φ and $t_1, \ldots, t_n$ are in Λ.

Since the sequence of terms needed to evaluate φ on $t_1, \ldots, t_n$ is the same as the sequence needed to evaluate $\neg\varphi$ on $t_1, \ldots, t_n$ we have an obvious fact.

Fact 2.3. *For each* Λ, φ *and* $t_1, \ldots, t_n \in \Lambda$, $(\varphi, t_1, \ldots, t_n)$ *is g.e. for* Λ *if and only if* $(\neg\varphi, t_1, \ldots, t_n)$ *is g.e. for* Λ.

Now, we define the notion of satisfaction for evaluations. Later, we relate this notion to the satisfaction relation in a model $M(\Lambda, p)$.

Definition 2.4. Let p be an evaluation on Λ. By induction on φ we define $p \models \varphi[\bar{t}]$ for $\bar{t} \in \Lambda$ such that $(\varphi, \bar{t})$ is g.e. for Λ:

- $p \models t = t'$ if $p(t, t') = 1$,
- $p \models t \leq t'$ if there is $s \in \Lambda$ such that $p \models (t + s = t')$,
- for φ quantifier free, $p \models \varphi[\bar{t}]$ if p makes φ true in the sense of propositional logic,
- $p \models \exists x\,\varphi(\bar{x}, x)[\bar{t}]$ if $p \models \varphi(\bar{x}, x)[\bar{t}, s^{\exists x\,\varphi}(\bar{t})]$,

- $p \models \forall x\, \varphi(\bar{x}, x)[\bar{t}]$ if for all terms $t \in \Lambda$ such that $(\varphi, \bar{t}, t)$ is g.e. for Λ, $p \models \varphi(\bar{x}, x)[\bar{t}, t]$.

Of course, whenever we write $p \models \varphi[\bar{t}]$ we assume that $(\varphi, \bar{t})$ is g.e. for Λ.

One can easily prove the following fact by induction on the complexity of φ.

FACT 2.5. *Let p be an evaluation on Λ and let φ and $\bar{t}$ be g.e. for Λ. If if $p \models \varphi[\bar{t}]$ then $p \not\models \neg\varphi[\bar{t}]$.*

DEFINITION 2.6. Let T be a theory and let p be an evaluation on Λ. We call p a *T-evaluation* if for all $\varphi \in T$ such that φ is g.e. for Λ, $p \models \varphi$.

If T has a Δ_0 definable set of axioms then the notion of *T-evaluation* is definable by a Δ_0 formula.

We have the following relation between $p \models \varphi$ and $M(\Lambda, p) \models \varphi$.

PROPOSITION 2.7. *Let Λ be a set of terms such that for any formula $\psi(x_1, \ldots, x_n, y)$ and $t_1, \ldots, t_n \in \Lambda$, the term $s^{\exists y\, \psi}(t_1, \ldots, t_n)$ is also in Λ. Let p be an evaluation on Λ such that $M(\Lambda, p)$ is well defined. Then for a formula φ and $\bar{t} \in \Lambda$,*

$$\text{if } p \models \varphi[\bar{t}] \text{ then } M(\Lambda, p) \models \varphi[\bar{t}].$$

Proof. The proof is a straightforward induction on the complexity of φ. For quantifier free formulas the statement is obvious. If $\varphi = \exists y\, \psi(\bar{t}, y)$ then from $p \models \varphi[\bar{t}]$ we deduce that $p \models \psi[\bar{t}, s]$, where $s = s^{\exists y\, \psi(\bar{x}, y)}(\bar{t})$, and we may use our inductive assumption to conclude that $M(\Lambda, p) \models \psi[\bar{t}, s]$ and that $M \models \varphi[\bar{t}]$.

For $\varphi = \forall y\, \psi(\bar{t}, y)$ we observe that, by the condition on Λ, for all $s \in \Lambda$, $(\psi, \bar{t}, s)$ is g.e. for Λ. It follows that for all $s \in \Lambda$, $p \models \psi[\bar{t}, s]$. Since the universe of $M(\Lambda, p)$ is made from terms in Λ, the inductive assumption implies that for all $a \in M$, $M(\Lambda, p) \models \psi[\bar{t}, a]$ and $M(\Lambda, p) \models \varphi[\bar{t}]$. ∎

Let us observe that it can happen that neither $p \models \varphi[\bar{t}]$ nor $p \models \neg\varphi[\bar{t}]$. This is the case when e.g. for some $\psi(x, y)$, $p \models \neg\psi[t, s^{\exists y\, \psi}(t)]$ and $p \models \psi[t, s]$, for some term s. In this case $p \not\models \exists y\, \psi(x, y)[t]$ and $p \not\models \forall y\, \neg\psi(x, y)[t]$. This is why we need the following definition which describes the situation when p satisfies, for a given formula $\varphi(\bar{x})$, the law of excluded middle.

DEFINITION 2.8. Let $(\varphi(x_1, \ldots, x_k), t_1, \ldots, t_k)$ be g.e. for Λ.
An evaluation p on Λ *decides* $(\varphi, t_1, \ldots, t_k)$ if

$$p \models \varphi[t_1, \ldots, t_k] \quad \text{or} \quad p \models \neg\varphi[t_1, \ldots, t_k].$$

An evaluation p *decides a formula* $\varphi(\bar{x})$ if for each sequence of terms $\bar{t} \in \Lambda$ such that $(\varphi, \bar{t})$ is g.e. for Λ, p decides $(\varphi, \bar{t})$.

Let N be an integer. An evaluation p on Λ is *N-deciding* if p decides all formulas φ with codes less than N.

For formulas which are decided by an evaluation p the satisfaction relation behaves in a way which is easy to handle.

LEMMA 2.9. *Let $(\forall x\,\varphi, \bar{t})$ be g.e. for Λ and let p decide $\forall x\,\varphi$. Then*

$$p \models \forall x\,\varphi[\bar{t}] \;\Leftrightarrow\; p \models \varphi[\bar{t}, s^{\exists x\,\neg\varphi}(\bar{t})].$$

Proof. The direction from left to right is obvious. So let us assume $p \models \varphi[\bar{t}, s^{\exists x\,\neg\varphi}(\bar{t})]$. Since p decides $\forall x\,\varphi$ we have either

$$p \models \forall x\,\varphi[\bar{t}] \quad \text{or} \quad p \models \exists x\,\neg\varphi[\bar{t}].$$

But if the latter is true then $p \models \neg\varphi[\bar{t}, s^{\exists x\,\neg\varphi}(\bar{t})]$, which is impossible by our assumption and Fact 2.5. ∎

We have the following proposition:

PROPOSITION 2.10. *Let $\varphi = Q_1 x_1 \dots Q_n x_n\,\psi(\bar{z}, x_1, \dots, x_n)$, where ψ is quantifier free, and let $(\varphi, \bar{t})$ be g.e. for Λ. Let p, an evaluation on Λ, decide $\varphi(\bar{t})$. Then*

$$p \models \varphi[\bar{t}] \;\Leftrightarrow\; p \models \psi[\bar{t}, s_1/x_1, \dots, s_n/x_n],$$

where $\bar{t}, s_1, \dots, s_n$ is the sequence for $(\varphi, \bar{t})$.

Proof. The proof is an easy induction on the complexity of φ. For the only nontrivial step for the universal quantifier, one uses Lemma 2.9. ∎

The relation $p \models \varphi[\bar{t}]$ is preserved when going to some subsets of the original set of terms Λ. As a consequence, if $\Lambda' \subseteq \Lambda$ and $M(\Lambda', p)$ is a well defined model, then its properties may be deduced from the properties of p considered as an evaluation on Λ.

Let us recall that for a cut $I \subseteq M$, $\Lambda{\restriction}I = \{t \in \Lambda : \mathrm{dp}(t) \in I\}$.

PROPOSITION 2.11. *Let $M \models \mathrm{I}\Delta_0$ and let $p \in M$ be an evaluation on a set of terms $\Lambda \in M$. Let $I \subseteq M$ be a cut in M and let $p{\restriction}I$ be an evaluation p restricted to $\Lambda{\restriction}I$. If $\bar{t} \in \Lambda{\restriction}I$ and $p \models \varphi[\bar{t}]$ then $p{\restriction}I \models \varphi[\bar{t}]$. Consequently, if $p \models \varphi[\bar{t}]$, then $M(\Lambda{\restriction}I, p{\restriction}I) \models \varphi[\bar{t}]$.*

Proof. We need to show that for all $\bar{t} \in \Lambda{\restriction}I$, if $p \models \varphi[\bar{t}]$ then $p{\restriction}I \models \varphi[\bar{t}]$.

For φ quantifier free the conclusion is obvious. For $\varphi = \exists y\,\psi(\bar{t}, y)$ one uses the fact that the term for Skolem witness, $s^{\exists y\,\psi(\bar{x}, y)}(\bar{t})$, is a member of $\Lambda{\restriction}I$, together with the inductive assumption. For the universal quantifier step one uses the fact that $\Lambda{\restriction}I$ is a subset of Λ. ∎

We will write p for an evaluation on a set of terms Λ as well as for its restriction to any subset of Λ. The last proposition shows that in order to establish that $M(\Lambda{\restriction}I, p) \models \varphi$, it suffices to show that $p \models \varphi$ when we treat p as an evaluation on the whole Λ.

The next lemma shows that if an evaluation p decides a formula $\exists y\,\varphi(y, \bar{t})$ then to check whether $p \models \exists y\,\varphi(y, \bar{t})$, it suffices to check whether $p \models \varphi(s, \bar{t})$

for some term $s \in \Lambda$. Indeed, any s is as good as the canonical witness for $\exists y\, \varphi(\bar{t})$ which is $s^{\exists y\, \varphi}(\bar{t})$.

LEMMA 2.12. *Let p be an evaluation on Λ and let p decide $\exists y\, \varphi(y, \bar{t})$. Then*

$$p \models \exists y\, \varphi(y, \bar{t}) \quad \textit{if and only if}$$
$$\textit{there is } s \in \Lambda \textit{ such that } (\varphi, s, \bar{t}) \textit{ is g.e. for } \Lambda \textit{ and } p \models \varphi(s, \bar{t}).$$

Proof. To prove the direction from left to right it suffices to take $s = s^{\exists y\, \varphi}(\bar{t})$. For the direction from right to left let us assume that there is $s_0 \in \Lambda$ such that $(\varphi, s_0, \bar{t})$ is g.e. for Λ and $p \models \varphi(s_0, \bar{t})$. By definition we have

$$p \models \exists y\, \varphi(y, \bar{t}) \quad \text{if and only if} \quad p \models \varphi(s^{\exists y\, \varphi}(\bar{t}), \bar{t}).$$

Thus let us assume, for the sake of contradiction, that

$$p \not\models \varphi(s^{\exists y\, \varphi}(\bar{t}), \bar{t}).$$

Since p decides $\exists y\, \varphi(y, \bar{t})$, it follows that

$$p \models \neg \exists y\, \varphi(y, \bar{t}).$$

This is equivalent to saying that for all $s' \in \Lambda$ such that $(\varphi, s', \bar{t})$ is g.e. for Λ,

$$p \models \neg \varphi(s', \bar{t}).$$

But this contradicts our assumption that $p \models \varphi(s_0, \bar{t})$. ∎

In the next lemma we show a kind of closedness of the relation $p \models \varphi$ under the Hilbert notion of provability. This lemma will be useful in establishing that a given T-evaluation p will satisfy some consequences of T.

LEMMA 2.13. *Let $T \vdash \varphi$, let $M \models \mathrm{I}\Delta_0$ and let $p \in M$ be a T-evaluation on Λ, where Λ contains all terms of standard depth. If p decides φ, then $p \models \varphi$.*

Proof. In the proof we use the fact that if $M \models \mathrm{I}\Delta_0$ then p and Λ have all the properties proven above for evaluations.

Let $\Lambda' \subseteq \Lambda$ be the set of all terms in Λ of standard depth. Then $M(\Lambda', p)$ is a well defined model. Moreover, since Λ' contains all standard terms, each axiom of T is g.e. for Λ'. Then, as p is a T-evaluation, $M(\Lambda', p) \models T$. Now, if $p \models \neg\varphi$, then $M(\Lambda', p) \models \neg\varphi$, which is impossible. ∎

Let T_0 be a finite set of axioms of $\mathrm{I}\Delta_0$ which characterize the recursive properties of successor, addition and multiplication and basic properties of ordering. We have to put in T_0 all axioms which are used in the proof of Lemma 2.14 below, e.g., such as $\forall x\, \forall y\, (x \leq y + 1 \Rightarrow (x \leq y \vee x = y + 1))$.

LEMMA 2.14. *Let $M \models \mathrm{I}\Delta_0$, let Λ be a set of terms from M such that $\{\underline{0}, \ldots, \underline{k}\} \subseteq \Lambda$ and let $p \in M$ be a T_0-evaluation on Λ. Then:*

(i) *for each* $t \in \Lambda$, $i \leq k$, *if* $p \models t \leq \underline{i}$ *then there exists* $j \leq i$, $p \models t = \underline{j}$;
(ii) *for each* $i, j \leq k$, $i \leq j$ *if and only if* $p \models \underline{i} \leq \underline{j}$;
(iii) *for each* $i, l, m \leq k$,

- $i + j = m \Leftrightarrow p \models \underline{i} + \underline{j} = \underline{m}$,
- $ij = m \Leftrightarrow p \models \underline{i}\,\underline{j} = \underline{m}$.

Proof. The proof of (i) is an easy induction on $i \leq k$. For $i = 0$ one uses the fact that p makes true the following axioms of T: $\forall x\,(0 \leq x)$ and $\forall x \forall y\,((x \leq y \wedge y \leq x) \Rightarrow x = y)$. Thus, if $p \models \underline{i} \leq \underline{0}$ then $p \models \underline{i} = 0$. The induction step follows easily from the fact that p makes true the following axiom: $\forall x \forall y\,(x \leq y + 1 \Rightarrow (x \leq y \vee x = y + 1))$.

For (ii) and (iii) one uses the inductive definitions of addition and multiplication and the properties of the ordering. ∎

The next lemma is a strengthening of Lemma 2.14. It shows that if $\{\underline{0}, \ldots, \underline{k}\} \subseteq \Lambda$ then any T_0-evaluation on Λ has to reflect the truth for Δ_0 formulas on $\{\underline{0}, \ldots, \underline{k}\}$, not only equalities between terms. We will use this for $a \in M$ and $\{\underline{0}, \ldots, \underline{a}\} \subseteq \Lambda$. Then the Δ_0 theory of M about $\{\underline{0}, \ldots, \underline{a}\}$ has to be reflected in $M(\Lambda, p)$.

Lemma 2.15 (Absoluteness Lemma). *Let* $M \models \mathrm{I}\Delta_0$, *let* $\Lambda \in M$ *be a set of terms such that* $\{\underline{0}, \ldots, \underline{k}\} \subseteq \Lambda$ *and let* $p \in M$ *be a* T_0*-evaluation on* Λ. *Let* φ *be a* Δ_0 *formula with only variables as bounds of quantifiers, such that values of terms in* $\varphi(\bar{x})$ *are not greater than* $\max\{\bar{x}\}$. *For all* $i_1, \ldots, i_m \leq k$ *such that* $(\varphi, \underline{i_m}, \ldots, \underline{i_m})$ *is g.e. for* Λ, *the following holds in M: if p decides* $(\varphi, \underline{i_m}, \ldots, \underline{i_m})$ *then*

$$\varphi(i_1, \ldots, i_m) \;\Leftrightarrow\; p \models \varphi[\underline{i_1}, \ldots, \underline{i_m}].$$

Proof. The proof is by induction on the complexity of φ. The case of atomic formulas holds by Lemma 2.14(ii)&(iii). The bounded quantifier step can be carried out by using Lemma 2.14(i). ∎

Now, we will estimate the size of terms which occur in the sequence for a given formula φ and $\bar{t}$. This lemma will be useful for ensuring that terms needed to evaluate $\varphi(t_1, \ldots, t_k)$ are elements of a given Λ.

Lemma 2.16 (Estimation Lemma). *Let* $\varphi(x_1, \ldots, x_k)$ *be a formula, let* $t_1, \ldots, t_k$ *be arbitrary terms and let* $t_1, \ldots, t_k, w_1, \ldots, w_r$ *be the sequence of terms needed to evaluate* φ *on* $t_1, \ldots, t_k$. *Then, for all* $i \leq r$,

$$w_i \leq \max\{t_j : j \leq k\}^{(\varphi^E)} \varphi^{(\varphi^E)},$$

where E *is a standard constant.*

Proof. First, we prove by induction on $i \le r$ that

$$w_i \le \Big(\varphi \prod_{j\le k} t_j\Big)^{(2A)^i}.$$

By (2.2), let A be such that

$$\mathrm{lh}(s^\varphi(t_1,\dots,t_k)) \le A\Big(\mathrm{lh}(\varphi) + \sum_{j\le k} \mathrm{lh}(t_j)\Big).$$

Then

$$w_1 \le 2^{A(\mathrm{lh}(\varphi)+\sum_{j\le k}\mathrm{lh}(t_j))} \le \varphi^A\Big(\prod_{j\le k} t_j\Big)^A \le \varphi^{2A}\Big(\prod_{j\le k} t_j\Big)^{2A}.$$

Now, let

$$w_i = s^\psi(t_1,\dots,t_k,w_1,\dots,w_{i-1}),$$
$$w_{i+1} = s^{\psi'}(t_1,\dots,t_k,w_1,\dots,w_i),$$

for ψ and ψ' being subformulas of φ. Then we have the following inequalities (the second follows from the fact that ψ' is a subformula of ψ; the third uses the left inequality of (2.2)):

$$\begin{aligned}\mathrm{lh}(w_{i+1}) &\le A\Big(\mathrm{lh}(\psi') + \sum_{j\le k}\mathrm{lh}(t_j) + \sum_{j\le i}\mathrm{lh}(w_j)\Big)\\ &\le A\Big(\mathrm{lh}(\psi) + \sum_{j\le k}\mathrm{lh}(t_j) + \sum_{j\le i-1}\mathrm{lh}(w_j)\Big) + A\,\mathrm{lh}(w_i)\\ &\le A(\mathrm{lh}(w_i)+\mathrm{lh}(w_i)) = 2A\,\mathrm{lh}(w_i).\end{aligned}$$

Thus, by the inductive assumption,

$$w_{i+1} \le 2^{2A\,\mathrm{lh}(w_i)} \le (w_i)^{2A} \le \Big(\Big(\varphi\prod_{j\le k} t_j\Big)^{(2A)^i}\Big)^{2A} \le \Big(\varphi\prod_{j\le k} t_j\Big)^{(2A)^{i+1}}.$$

But $r, k \le \log(\varphi)$, so

$$\begin{aligned}w_i &\le \Big(\prod_{j\le k} t_j\Big)^{O(1)^{\log(\varphi)}} \varphi^{O(1)^{\log(\varphi)}}\\ &\le (\max\{t_i : i \le r\})^{\log(\varphi)O(1)^{\log(\varphi)}}\varphi^{O(1)^{\log(\varphi)}}\\ &\le (\max\{t_i : i \le r\})^{\log(\varphi)(\varphi^{O(1)})}\varphi^{(\varphi^{O(1)})}\\ &\le (\max\{t_i : i \le r\})^{(\varphi^{O(1)})}\varphi^{(\varphi^{O(1)})}. \ \blacksquare\end{aligned}$$

The theorem below is Theorem 1.1 from [A02]. Below, we write $x \in \log^n$ to indicate that x is within the nth logarithm of a universe. This can be expressed as $\exists x_1 \dots \exists x_n\,[\mathrm{Exp}(x,x_i)\wedge\bigwedge_{1\le i<n}\mathrm{Exp}(x_i,x_{i+1})]$, where $\mathrm{Exp}(x,y,z)$ is a Δ_0 formula defining (provably in $\mathrm{I}\Delta_0$) the graph of the exponentiation

function (see [HP93]). For a sequence $\bar{x} = x_1, \ldots, x_k$, $\bar{x} \in \log^n$ should be read as $\bigwedge_{1 \le i \le k} x_i \in \log^n$. Consequently, $\exists \bar{x} \in \log^n \varphi(\bar{x})$ is shorthand for $\exists \bar{x} \, (\bar{x} \in \log^n \wedge \varphi(\bar{x}))$.

THEOREM 2.17 (Adamowicz, [A02]). *For each $m, n \in \mathbb{N}$ there is a bounded formula $\theta(\bar{x})$ such that*

$$\mathrm{I}\Delta_0 + \Omega_n + \exists \bar{x} \in \log^m \theta(\bar{x}) \text{ is consistent}$$

and

$$\mathrm{I}\Delta_0 + \Omega_n + \exists \bar{x} \in \log^{m+1} \theta(\bar{x}) \text{ is inconsistent.}$$

Let us recall that an evaluation p on Λ is a T-evaluation if for each $\varphi \in T$ such that φ is g.e. for Λ, $p \models \varphi$. Then, for an integer N, p is N-deciding if p decides all formulas φ with codes less than N, that is, for each $\bar{t}$ such that $(\varphi, \bar{t})$ is g.e. for Λ, either $p \models \varphi[\bar{t}]$ or $p \models \neg\varphi[\bar{t}]$.

We define the following version of Herbrand consistency.

DEFINITION 2.18. Let N be a standard constant. $\mathrm{HCons}(N, T, i)$ is a Π_1 arithmetical formula which states that for each set of terms Λ of depth not greater than i, there exists an N-deciding T-evaluation on Λ.

Let us comment on the above definition. Usually, Herbrand consistency is formulated as follows. Let $\mathrm{sk}(\varphi)$ be the quantifier free formula obtained after skolemization of φ and removing its universal quantifiers. E.g.

$$\mathrm{sk}(\forall x \, \exists y \, \forall z \, \exists w \, P(x, y, z, w)) = P(x, s_1(x), z, s_2(x, z)),$$

where s_1 and s_2 are the Skolem functions for $\exists y \, \forall z \, \exists w \, P(x, y, z, w)$ and $\exists w \, P(x, s_1(x), z, w)$, respectively. Then, for a sequence of terms $\bar{t}$ and a formula $\varphi(\bar{x})$ with free variables $\bar{x}$ let $\varphi[\bar{t}]$ be the formula obtained by substitution of terms $\bar{t}$ for the variables $\bar{x}$. Then the Herbrand theorem can be stated as follows: a formula φ is provable in first order logic if and only if there exists a finite set of terms Λ such that $\bigvee_{\bar{t} \in \Lambda} \neg \mathrm{sk}(\neg\varphi)[\bar{t}]$ is provable in propositional logic.

Since we are concerned with consistency, we use an equivalent form: a formula φ is consistent if and only if for each finite set of terms Λ, the formula $\bigwedge_{\bar{t} \in \Lambda} \mathrm{sk}(\varphi)[\bar{t}]$ is consistent in propositional logic.

If a given theory T has equality as the only predicate, a boolean evaluation only needs to state which pairs of terms for a skolemized language are equal. Hence a theory T is consistent if and only if for each finite set of terms Λ in the skolemized language there is an evaluation p on pairs of terms from Λ such that p makes true all axioms of T which are g.e. for Λ.

In proving the above equivalence one needs to note that T is consistent if and only if its skolemization $\mathrm{sk}(T)$ is consistent. The left-to-right direction of the above equivalence may be easily proved by taking any model M for

$\mathrm{sk}(T)$. Indeed, we can interpret in M any term from $\mathrm{sk}(T)$ and define an evaluation from the satisfaction relation of M.

To prove the other direction one needs to observe that $\mathrm{sk}(T)$ is a purely universal theory. Now, let p be a boolean evaluation on the set of atomic formulas of $\mathrm{sk}(T)$ with all possible substitutions of terms of $\mathrm{sk}(T)$. If p makes all axioms from $\mathrm{sk}(T)$ true then, by universality of $\mathrm{sk}(T)$, p defines a model of $\mathrm{sk}(T)$ on a set of terms. In order to obtain such a p, it is enough, by compactness, to find good evaluations for each finite set of terms Λ.

Our definition of Herbrand consistency deviates from the above in two ways. Firstly, we add Skolem functions for all formulas of the form $\exists y\, \psi$, not only for subformulas of axioms of T. This alone does not change the difficulty of proving Herbrand consistency. Indeed, if a Skolem term does not occur in the axioms of $\mathrm{sk}(T)$ then we can interpret it freely e.g. as denoting zero. Thus, more terms do not introduce any difficulty in proving Herbrand consistency. However, this changes with the requirement that an evaluation should decide some finite set of formulas less than N.

If $\mathrm{sk}(T)$ has a model, we can interpret any term in this model and construct an evaluation from the satisfaction relation of this model. It is easy to see that such an evaluation will decide any formula. Thus, the consistency of $\mathrm{sk}(T)$ gives the existence of an N-deciding evaluation, for any N. However, in the bounded arithmetic world the requirement of being N-deciding may increase the difficulty of proving Herbrand consistency. It will follow that some formulas provable in T will have to be true also in the sense of the constructed evaluation. Therefore our evaluations will be somewhat better behaved. However, as we will see in Theorem 3.2 this additional condition does not restrict the provability of some cases of Herbrand consistency while it allows us to have an interesting and still natural unprovability result.

We do not specify what is the size of the constant N. We do not need to fix it because for each i, $\mathrm{I}\Delta_0 + \Omega_i \vdash \mathrm{HCons}(N, \mathrm{I}\Delta_0 + \Omega_i, \log^{i+3})$ for an arbitrary constant N (Theorem 3.2). On the other hand for our unprovability result one should take N so large that evaluations decide all relevant formulas which occur in the course of the proof of the unprovability of $\mathrm{HCons}(N, \mathrm{I}\Delta_0 + \Omega_i, (1+\varepsilon)\log^{i+2})$. It will be a large constant but its precise value is irrelevant. It may seem that N should have some self-referential properties. However, as we will comment during the proof, it is enough that N is a large constant definable by a short formula. E.g. it may be of the form $\exp^n(\underline{2})$, which is definable by a formula $\exists x_0 \ldots \exists x_n \,(x_0 = \underline{2} \wedge \bigwedge_{1 \leq i \leq n} x_i = \exp(x_{i-1}))$ which has length linear in n.

We believe that our two theorems on provability and unprovability of our notion of Herbrand consistency (see Theorems 3.2 and 4.12) justify our choice. They show that we can come close to some border cases of provability of Herbrand consistency. Moreover, our discussion at the end of the

Introduction shows that either way of closing the gap would be interesting. Nevertheless, one should be aware that in the bounded arithmetic context not only cut-free proof methods are not equivalent to proofs with cuts but even various cut-free methods may not be equivalent when we do not have the totality of exp. This was proven e.g. for tableaux and Herbrand proofs by Kołodziejczyk ([K06b]).

Finally, let us note that a similar notion of evaluation was used in [AK04]. Evaluations defined there had to decide all Σ^b_n formulas not greater than a fixed parameter N and had to reflect the truth of a model for Σ^b_n formulas with numerals. Such a notion was related in [AK04] to instances of Σ^b_{n+1} induction on some $\log^k$-part of a model.

3. Provability of Herbrand consistency. In this section we exhibit a case for which Herbrand consistency is provable in bounded arithmetic. To show that for each $M \models \mathrm{I}\Delta_0 + \Omega_i$, for a given set terms $\Lambda \in M$, there exists in M an N-deciding $\mathrm{I}\Delta_0 + \Omega_i$-evaluation p, we will construct in M a set of interpretations H for terms in Λ. Then an evaluation p will be defined according to H. Intuitively, elements of H are just the real values for terms in Λ. However, to construct H we should be able to compute values for all Skolem functions from terms in Λ. This means that we should have a suitable truth definition. Below, we state the existence of such a truth definition (see Theorem 5.4 in [HP93]).

THEOREM 3.1. *There exists a Δ_0 formula $\mathrm{Tr}_{\Delta_0}(\varphi, y, p)$ with a parameter p which is, provably in $\mathrm{I}\Delta_0$, a truth definition for Δ_0 formulas, whenever a sufficiently large parameter is substituted for p. Namely, for $\varphi(x_1, \ldots, x_n)$ and $y = \langle b_1, \ldots, b_n\rangle$ one should take $p \geq \exp(y)^{\varphi^c}$ for some standard constant c.*

We will use the above truth definition in the proof of the next theorem. All Δ_0 formulas φ to which we will apply Tr_{Δ_0} will be smaller than $\log^2(y)$. Thus, it will be enough for us to take $\omega_1(\exp(y))$ as parameter.

THEOREM 3.2. *For each $N \in \omega$, $\mathrm{I}\Delta_0 + \Omega_i$ proves its Herbrand consistency restricted to the terms of depth not greater than $\log^{i+3}$ for N-deciding evaluations, that is,*

$$\mathrm{I}\Delta_0 + \Omega_i \vdash \mathrm{HCons}(N, \mathrm{I}\Delta_0 + \Omega_i, \log^{i+3}).$$

Proof. We prove the theorem for the case of $i = 1$. The proof for $i > 1$ is essentially the same.

Let $T = \mathrm{I}\Delta_0 + \Omega_1$. Let $M \models T$ and let $\Lambda = \{t_1, \ldots, t_k\}$ be a set of terms of depth not greater than some $d \in \log^4(M)$. For simplicity we assume that if t_m is a subterm of t_j then $m \leq j$. We prove by induction on $m \leq k$ that

$$\exists H_m = \{h_{t_1}, \dots, h_{t_m}\}\, \forall j \le m\, [\forall a \le \Lambda\, (t_j = \underline{a} \Rightarrow h_{t_j} = a) \wedge$$
$$\forall r\, \forall \varphi \le \Lambda\, \forall (n_1, \dots, n_r) \le \Lambda\, (t_j = s^{\exists y\, \varphi}(t_{n_1}, \dots, t_{n_r}) \Rightarrow$$
$$h_{t_j} = \text{the least witness of } \exists y\, \varphi(h_{t_{n_1}}, \dots, h_{t_{n_r}}) \text{ or 0 otherwise})].$$

Since the theory T is Π_1, it is easy to see that the greatest element of H_i may be only for a term $\omega_1^d(\underline{0})$ and is less than $\exp^3(d) \in \log(M)$ because $d \in \log^4(M)$. (We assume here that symbols for $\omega_1(x)$ as well as for multiplication and addition are of the form $s^{\exists y\, \varphi}(\bar{x})$.) Thus, to compute the witness for $\exists y \ \le\ t\, \varphi(h_{t_{n_1}}, \dots, h_{t_{n_r}})$ we may use a universal formula $\mathrm{Tr}_{\Delta_0}(x, y, a)$ for $a = \omega_1(\exp^4(d)) \in M$.

It is worth mentioning that this is the only place where we use the relation between the rate of growth of the ω_1 function and the 4th logarithm (or, more generally, between the rate of growth of ω_i and the $(i+3)$th logarithm).

It is also easy to see that H_m is small enough to be in M. The number of terms of depth below d is not greater than $d^{\log(d)^d}$. Indeed, the number of nodes in the tree for a term of depth not greater than d is at most $\log(d)^d$ ($\log(d)$ is the branching of the tree and d is the depth of the tree). Since we have only d labels for these nodes, the number of terms is at most $d^{\log(d)^d}$. Thus,

$$\mathrm{card}(H_i) \le \log^4(M)^{(\log^5(M)^{\log^4(M)})} \le 2^{2^{\log^6(M)(\log^4(M)+1)}} \le 2^{2^{\log^3(M)}} \le \log(M).$$

It follows that the size of H_m, the set of $\log(M)$ elements of sizes in $\log(M)$, is not greater than

$$\log(M)^{\log(M)} \le 2^{(\log(M))^2},$$

which is an element of M. Thus we can take an element of M to bound the quantifier $\exists H_m$ in the induction formula.

Now, we define an evaluation p on $\Lambda = \{t_1, \dots, t_k\}$ according to $H_k = \{h_{t_1}, \dots, h_{t_k}\}$:

$$p(t, t') = 1 \ \Leftrightarrow\ h_t = h_{t'}.$$

It suffices to show that p is an N-deciding T-evaluation. By induction on the complexity of formulas we show that p decides all standard formulas. Indeed, we show something stronger: for each formula φ and for all terms $s_1, \dots, s_r \in \Lambda$,

$$M \models \varphi[h_{s_1}, \dots, h_{s_r}] \ \Leftrightarrow\ p \models \varphi[s_1, \dots, s_r].$$

For atomic formulas the statement is obvious, as it also is for all quantifier free formulas. Now, let us take a formula $\varphi = \exists y\, \psi(y, \bar{x})$ and $\bar{s} \in \Lambda$, $\bar{s} = s_1, \dots, s_r$, such that $(\varphi, \bar{s})$ is g.e. for Λ. If $M \models \exists y\, \psi[h_{s_1}, \dots, h_{s_r}]$, then for $s = s^{\exists y\, \psi}(s_1, \dots, s_r)$, $M \models \psi[h_s, h_{s_1}, \dots, h_{s_r}]$ and, by the inductive assumption, $p \models \psi[s, s_1, \dots, s_r]$. So, $p \models \exists y\, \psi[s_1, \dots, s_r]$.

On the other hand, if $M \models \neg\exists y\, \psi[h_{s_1}, \ldots, h_{s_r}]$, then for all $h \in H$, $M \models \neg\psi[h, h_{s_1}, \ldots, h_{s_r}]$. It easily follows by the inductive assumption that $p \models \neg\exists y\, \psi[s_1, \ldots, s_r]$.

Let us observe that the above argument also works for all nonstandard Δ_0 formulas if we change our statement to: for all terms $s_1, \ldots, s_r \in \Lambda$,

$$M \models \mathrm{Tr}_{\Delta_0}(\varphi, \langle h_{s_1}, \ldots, h_{s_r}\rangle, \omega_1(\exp^4(d))) \;\Leftrightarrow\; p \models \varphi[s_1, \ldots, s_r].$$

As we stated above, $\mathrm{Tr}_{\Delta_0}(\varphi, \langle h_{s_1}, \ldots, h_{s_r}\rangle, \omega_1(\exp^4(d)))$ has in M all the properties of a Δ_0 truth definition.

Now we show that p satisfies bounded induction axioms. Let $\varphi(x)$ be a Δ_0 formula. We want to show that

$$p \models \forall z\, (\neg\varphi(\underline{0}) \vee \exists x \leq z\, (\varphi(x) \wedge \neg\varphi(x+1)) \vee \varphi(z)).$$

Let us assume that

$$M \models \mathrm{Tr}_{\Delta_0}(\varphi, 0, \exp^4(d))$$

and

$$M \models \forall x \leq h_{s_i}\, (\mathrm{Tr}_{\Delta_0}(\varphi, x, \exp^4(d)) \Rightarrow \mathrm{Tr}_{\Delta_0}(\varphi, x+1, \exp^4(d))),$$

where h_{s_i} is an arbitrary, fixed element of H. If not, then by the remark above, we could easily show that either $p \models \neg\varphi(\underline{0})$ or $p \models \exists x \leq s_i\, (\varphi(x) \wedge \neg\varphi(x+1))$. Now, by Δ_0 induction in M for $\mathrm{Tr}_{\Delta_0}(\varphi, x, \exp^4(d))$ we infer that $M \models \mathrm{Tr}_{\Delta_0}(\varphi, h_{s_i}, \exp^4(d))$. Thus, $p \models \varphi[s_i]$. Since s_i is arbitrary we have shown that the induction axiom holds under p. ∎

4. Unprovability of Herbrand consistency. In this section we prove that for $T_i = \mathrm{I}\Delta_0 + \Omega_i$, there exists an integer N such that T_i does not prove its Herbrand consistency restricted to terms of depth in $(1+\varepsilon)\log^{i+2}$, for any $\varepsilon > 0$. However, for simplicity, we present the proof only for the subtlest case of $\mathrm{I}\Delta_0 + \Omega_1$. Indeed, only in this case should we take care that all the objects we construct are inside the model and that the main inductive argument can be carried out in a bounded induction arithmetic. We encourage the reader to review the proof after reading it to check how it behaves for $i > 1$. Indeed, all estimations then become easier. One should only replace $\log^3$ with $\log^{i+2}$, $\log^4$ with $\log^{i+3}$ and ensure that all elements needed in the proof are in the model.

Therefore, from now on $T = \mathrm{I}\Delta_0 + \Omega_1$. For the sake of contradiction, till the end of this section we assume that for all integers N there exists $\varepsilon > 0$ such that

$$T \vdash \mathrm{HCons}(T, (1+\varepsilon)\log^3, N).$$

In fact we will use this assumption for a fixed large value of N and one fixed $\varepsilon > 0$ chosen for this N.

Let us also fix a model $M \models T$ and an element $a \in \log^3(M)$. We will consider only evaluations for the set of Skolem terms for formulas $\varphi \leq \log^*(a)$. Since our result is about unprovability of HCons, such a restriction only makes our result stronger.

The main idea of the proof is the following. Under the assumption that $T \vdash \mathrm{HCons}(N, T, (1+\varepsilon)\log^3)$, we show that for any model $M \models T$ and any $a \in \log^3(M)$ we can construct a model M' such that

$$M{\restriction}\{0,\dots,a\} \cong M'{\restriction}\{0,\dots,a\} \quad \text{and} \quad M' \models a \in \log^4 .$$

Together with Theorem 2.17 this will allow us to obtain a contradiction when we suitably choose an element $a \in M$ which cannot be (provably in T) in $\log^4$ due to its Δ_0 properties.

In order to construct M' we work in M. We construct a sequence of sets of terms and evaluations on them, $\{(\Lambda_i, p_i)\}_{i \leq a/C^2}$, for some standard constant C. The key property of the sequence will be that under p_i the element $\exp^4((1+\varepsilon/2)^i a)$ exists. Then, the desired model M' will be, roughly speaking, the model defined by Λ_{a/C^2} and p_{a/C^2}.

We should say a word on how we choose the constant C. Again the reader should think about C as a fixed large integer. Our construction is uniform in C so that a particular choice of C is not important. We only require that C is so large that

- $C > 4E\log(C)$,
- $C \geq (\models)\log(2A) + \log(A) + 1$,

where E is the constant from the Estimation Lemma, A is the constant from (2.2), and $\models$ is understood as a Gödel number for the formula $x \models y[z]$.

We start with a definition which will be used in the main inductive argument. Since the definition is quite involved we comment on it below.

Definition 4.1. Let $M \models \mathrm{I}\Delta_0$, let $a \in \log^3(M)$, and let $\varepsilon > 0$ be a small standard constant. The elements a and ε are parameters of the definition which are fixed during the whole proof.

Let Λ be a set of terms, let p be an evaluation on Λ and let $k, b \in M$. The sequence (Λ, p, k, b) is *suitable* when

(i) $k, b > \mathbb{N}$,

(ii) Λ is the set of terms of the form

$$\Lambda = \{t : \mathrm{Term}(t) \wedge \mathrm{dp}(t) \leq b \wedge t \leq 2^{2^k}\},$$

(iii) p is a T-evaluation on Λ and p is N-deciding,

(iv) $k + (\varepsilon/4)a < b$ and $b(\varepsilon/4)a < 2^k$.

During the induction we will consider only suitable sequences $(\Lambda_i, p_i, k_i, b_i)$ where $k_i = a - iC$ and $b_i = (1+\varepsilon/2)^{i+1}a$, for $i \leq a/C^2$. In item (iv) of the definition it would suffice that $k_i + \mathbb{N} < b_i$ and $b_i\mathbb{N} < 2^{k_i}$. However this

condition is not expressible by an arithmetical formula unless we can define the standard part of the model.

For a suitable sequence $(\Lambda_i, p_i, k_i, b_i)$ we will have $\{\underline{0}, \ldots, \underline{2^{2^{(k_i)}-1}}\} \subseteq \Lambda_i$. Since k_i's are decreasing, p_i's will reflect Δ_0 truth on smaller parts of a model. However, the element a will always be in this part. On the other hand the depths of terms in Λ_i's will grow with b_i's. It follows that we will have bigger terms of the form $\omega^{b_i}(\underline{8})$ in Λ_i. This will allow us to show that we have more exponentiation under p_i as i grows. The above will determine the needed properties of a model defined from $(\Lambda_n, p_n, k_n, b_n)$, for $n = i/C^2$.

Now, we will establish some properties of a suitable (Λ, p, k, b).

FACT 4.2. *Let (Λ, p, k, b) be suitable. Then $\{\underline{0}, \ldots, \underline{2^{2^k-1}}\} \subseteq \Lambda$.*

Proof. For $i \leq 2^{2^k-1}$, we have $\underline{i} \leq 2i \leq 2^{2^k}$ and $\mathrm{dp}(\underline{i}) \leq k < b$. ∎

In the next lemma we show which formulas with numerals as parameters are g.e. for a suitable sequence (Λ, p, k, b).

LEMMA 4.3. *Let (Λ, p, k, b) be suitable and let $\varphi(x_1, \ldots, x_r)$ be a formula less than C. Then, for $m_1, \ldots, m_k \leq 2^{2^{k-C}}$, $(\varphi, \underline{m_1}, \ldots, \underline{m_r})$ is g.e. for Λ.*

Proof. The lemma follows from the Estimation Lemma and the fact that, by our choice of C, $C > 4E\log(C)$, where E is the constant from the Estimation Lemma. Indeed, by the Estimation Lemma, the size of the greatest term needed to evaluate $\varphi(\underline{m_1}, \ldots, \underline{m_r})$ is not greater than $(\max\{\underline{m_i} : i \leq r\})^{(C^E)} C^{(C^E)}$. It follows that these terms are not greater than

$$\begin{aligned}(22^{2^{k-C}})^{(C^E)} C^{(C^E)} &\leq 2^{(2^{k-C}+1)(C^E)+\log(C)C^E} \\ &\leq 2^{(2^{k-C+1+E\log(C)+E\log(C)+\log\log(C)})} \\ &\leq 2^{2^{k-C+4E\log(C)}} \leq 2^{2^k}.\end{aligned}$$

Moreover, the depths of terms are not greater than $k + \mathbb{N}$, which is less than b. ∎

In the most important case, the φ from Lemma 4.3 will be just $x \models y[z]$.

In the lemma below we show how much exponentiation is available under an evaluation p which occurs in a suitable sequence (Λ, p, k, b). We show that $p \models$ "$\exp^3(\underline{b - C})$ exists". Ideally, we would like to have $\exp^3(\underline{b})$. Unfortunately, to show that $\exp^3(\underline{i})$ exists we need a term $\omega_1^i(\underline{8})$ and we need some formulas with this term (used in the course of proof of Lemma 4.4) to be g.e. for Λ. This is why we restrict Lemma 4.4 to $\exp^3(\underline{b - C})$.

LEMMA 4.4. *Let (Λ, p, k, b) be suitable. Then*

$$p \models \exists x\, (x = \exp^3(\underline{b - C})).$$

Proof. Let (Λ, p, k, b) be a suitable sequence. In order to show that $p \models \exists x\,(x = \exp^3(\underline{b - C}))$ we prove that for each $i \leq b - C$,

$$p \models x = \exp^3(y)[\omega_1^i(\underline{8})/x, \underline{i}/y],$$

which clearly suffices by Lemma 2.12.

By Lemma 2.13 and since N is chosen large enough,

$$p \models \forall y\,\forall x\,(x = \omega_1^y(\underline{8}) \Rightarrow x = \exp^3(y)). \tag{4.1}$$

Of course, in the formula $x = \omega_1^y(\underline{8})$, y is a free variable so "$x = \omega_1^y(\underline{8})$" should not be read as an equality between two terms but as a formula with free variables x and y.

By (4.1), to show that for all $i \leq b - C$,

$$p \models x = \exp^3(y)[\omega_1^i(\underline{8})/x, \underline{i}/y],$$

it suffices to show that for each $i \leq b - C$,

$$p \models x = \omega_1^y(\underline{8})[\omega_1^i(\underline{8})/x, \underline{i}/y].$$

For $i = 0$ there is nothing to prove. Indeed, $\underline{8} = \omega_1^0(\underline{8})$ is a true Δ_0 formula thus it has to be decided positively by p.

Now, let us assume that for some $i < b - C$, $p \models x = \omega_1^y(\underline{8})[\omega_1^i(\underline{8}), \underline{i}]$. Then, since

$$T \vdash \forall x\,\forall y\,[x = \omega_1^y(\underline{8}) \Rightarrow \omega_1(x) = \omega_1^{y+1}(\underline{8})],$$

we have, again by Lemma 2.13,

$$p \models \omega_1(x) = \omega_1^{y+1}(\underline{8})[\omega_1^i(\underline{8})/x, \underline{i}/y].$$

Since $p \models z = y + 1[\underline{i+1}/z, \underline{i}/y]$, the last display is nothing other than

$$p \models x = \omega_1^y(\underline{8})[\omega_1^{i+1}(\underline{8})/x, \underline{i+1}/y].$$

Thus, we have proved the induction step and this ends the proof of the lemma. ∎

In the next lemma we show how to construct from a suitable sequence (Λ, p, k, b) a new suitable sequence. The new sequence is of the form $(\tilde{\Lambda}, \tilde{p}, k - C, (1 + \varepsilon/2)b)$. The lemma below is, essentially, the inductive step in our construction. The key property of the new sequence is that we will have more exp available under $\tilde{p}$.

LEMMA 4.5. *Let $\varepsilon' = \varepsilon/2$, $M \models \mathrm{I}\Delta_0$, $a \in \log^3(M)$ and let $i < a/C^2$. Let (Λ, p, k, b) be suitable, where $k = a - iC$ and $b = (1 + \varepsilon')^{i+1}a$. Then there are $\tilde{\Lambda}, \tilde{p}$ such that $(\tilde{\Lambda}, \tilde{p}, \tilde{k}, \tilde{b})$ is suitable, where $\tilde{k} = k - C$ and $\tilde{b} = (1 + \varepsilon')b$.*

Proof. It is straightforward that $\tilde{k}$ and $\tilde{b}$ satisfy item (iv) of Definition 4.1. Indeed, it is enough to verify that

$$a + (\varepsilon/4)a < (1 + \varepsilon/2)a \quad \text{and} \quad (1 + \varepsilon/2)^{a/C^2}a < 2^{\frac{C-1}{C}a}.$$

We define $\tilde{\Lambda}$ in the only possible way as

$$\tilde{\Lambda} = \{t : \mathrm{Term}(t) \wedge \mathrm{dp}(t) \le \tilde{b} \wedge t \le 2^{2^{\tilde{k}}}\}.$$

CLAIM 4.6. *For each $t \le 2^{2^{k-C}}$, $t \in \tilde{\Lambda}$ if and only if*

$$p \models \mathrm{Term}(\underline{t}), \quad p \models \mathrm{dp}(\underline{t}) \le \underline{(1+\varepsilon')b} \quad \textit{and} \quad p \models \underline{t} \le \underline{2^{2^{k-C}}}.$$

Proof. For $t \le 2^{2^{k-C}}$ all the formulas: $\mathrm{Term}(x)$, $\mathrm{dp}(x) \le y$ and $x \le 2^{2^y}$ with terms $\underline{t}$ and $\underline{(1+\varepsilon')b}$ are g.e. for Λ (see Lemma 4.3). Since p decides these formulas, the claim follows from the Absoluteness Lemma. ∎

Let $\Lambda(t, x, y)$ be a formula expressing that the term t is such that $t \le 2^{2^x}$ and $\mathrm{dp}(t) \le y$. Claim 4.6 established that

$$\forall t\,(t \in \tilde{\Lambda} \;\Leftrightarrow\; p \models \Lambda[\underline{t}, \underline{k-C}, \underline{(1+\varepsilon')b}]). \tag{4.2}$$

Let $\Lambda(x, y)$ be the set of terms defined by $\Lambda(t, x, y)$. We can refer to this set by a term $s^{\exists z\,\forall t \le x\,(t \in z \equiv \Lambda(t,x,y))}$. Let

$$\gamma(x, y) := \exists z\,(z \text{ is an } N\text{-deciding } T\text{-evaluation on } \Lambda(x, y)).$$

Next, let

$$\hat{p} = s^{\exists z\,\gamma(x,y)}(\underline{k-C}, \underline{(1+\varepsilon')b}).$$

We have assumed that $T \vdash \mathrm{HCons}(N, T, (1+\varepsilon)\log^3)$. Thus, since

$$(1+\varepsilon')b \le (1+\varepsilon)(b-C),$$

and by Lemmas 4.4 and 4.3, we have

$$p \models \exists z\,(z \text{ is an } N\text{-deciding } T\text{-evaluation on } \Lambda(x, y))[\underline{k-C}/x, (1+\varepsilon')b/y],$$

and consequently

$$p \models (z \text{ is an } N\text{-deciding } T\text{-evaluation on } \Lambda(x, y))[\hat{p}/z, \underline{k-C}/x, (1+\varepsilon')b/y].$$

Of course, by our choice of N, p decides this formula ([1]).

Since p decides the formula $x_1 \models x_2[x_3]$, for each $t_1, t_2 \in \tilde{\Lambda}$ we have

$$p \models \text{“}\hat{p} \models \underline{t_1} = \underline{t_2}\text{”} \quad \text{or} \quad p \models \text{“}\neg\hat{p} \models (\underline{t_1} = \underline{t_2})\text{”},$$

and not both. Thus, we define $\tilde{p}$ on $\tilde{\Lambda}$ as follows: for each $t_1, t_2 \in \tilde{\Lambda}$,

$$\tilde{p} \models t_1 = t_2 \;\Leftrightarrow\; p \models \text{“}\hat{p} \models \underline{t_1} = \underline{t_2}\text{”}.$$

Such a $\tilde{p}$ exists in M by Δ_0 induction.

We claim that $(\tilde{\Lambda}, \tilde{p}, \tilde{k}, \tilde{b})$ is suitable. It suffices to show that $\tilde{p}$ is an N-deciding T-evaluation on $\tilde{\Lambda}$. The other conditions from Definition 4.1 are easily seen to be satisfied.

([1]) One may object that we want N to be greater than a formula which uses N as a fixed parameter. However, this is easily possible if N has a short encoding. If N is of the form $\exp^n(\underline{2})$ for some n, then a formula $\varphi(N)$ can be written in a short but equivalent form $\exists z\,(z = \exp^n(\underline{2}) \wedge \varphi(z))$.

Now, we establish the relationship between $\tilde{p} \models \varphi$ and $p \models$ "$\hat{p} \models \underline{\varphi}$". We need to show that it makes sense to ask whether $p \models$ "$\hat{p} \models \underline{\varphi}$" when we ask whether $\tilde{p} \models \varphi$.

CLAIM 4.7. *Let* $\varphi \leq \log^*(a)$ *and* $t_1, \dots, t_m \in \tilde{\Lambda}$. *If* $(\varphi, t_1, \dots, t_m)$ *is g.e. for* $\tilde{\Lambda}$ *then* $(\models, \hat{p}, \underline{\varphi}, \langle \underline{t_1}, \dots, \underline{t_m} \rangle)$ *is g.e. for* Λ.

Proof. Let us assume that $(\varphi, t_1, \dots, t_m)$ is g.e. for $\tilde{\Lambda}$. Under the usual coding the term $s^{\varphi}(t_1, \dots, t_m)$ is greater than $\varphi \prod_{i \leq m} t_i$. Thus, by the construction of $\tilde{\Lambda}$,

$$\varphi \prod_{i \leq m} t_i \leq 2^{2^{k-C}}.$$

Now, let

$$\begin{aligned} s_0 &= f_1(\hat{p}, \underline{\varphi}, \langle \underline{t_1}, \dots, \underline{t_m} \rangle), \\ s_1 &= f_2(\hat{p}, \underline{\varphi}, \langle \underline{t_1}, \dots, \underline{t_m} \rangle, s_0), \\ &\vdots \\ s_r &= f_r(\hat{p}, \underline{\varphi}, \langle \underline{t_1}, \dots, \underline{t_m} \rangle, s_0, \dots, s_{r-1}) \end{aligned}$$

be all terms in the sequence for a formula $\hat{p} \models \underline{\varphi}[\underline{t_1}, \dots, \underline{t_m}]$ besides the parameters $\hat{p}$, $\underline{\varphi}$ and $\langle \underline{t_1}, \dots, \underline{t_m} \rangle$. (Here $\langle x_1, \dots, x_m \rangle$ is the mth iteration of the pairing function.) We can estimate the depth of s_i as

$$\mathrm{dp}(s_i) \leq \max\{\mathrm{dp}(\underline{t_i}) : i \leq m\} \cup \{\mathrm{dp}(\underline{\varphi})\} \cup \{\mathrm{dp}(\hat{p})\} + \mathbb{N} \leq k + \mathbb{N} < b.$$

Thus, it suffices to show that $s_1, \dots, s_r \leq 2^{2^k}$. We estimate the size of s_r.

Since $s_1, \dots, s_r$ are terms witnessing quantifiers in $\models$, we have $r \leq \mathrm{lh}(\models)$. Firstly, we estimate the lengths of the terms s_i. We show that for $i \leq r$,

$$\mathrm{lh}(s_i) \leq 2^i A^i \,\mathrm{lh}(s_0)$$

where A is the constant from (2.2). Of course, there is nothing to prove for s_0 but we also write down the formula for $\mathrm{lh}(s_0)$ since it will be useful later:

$$\mathrm{lh}(s_0) \leq A\Big(\mathrm{lh}(\models) + \mathrm{lh}(\hat{p}) + \mathrm{lh}(\underline{\varphi}) + m\,\mathrm{lh}(\langle\rangle) + \sum_{i \leq m} \mathrm{lh}(\underline{t_i}) \Big), \tag{4.3}$$

$$\begin{aligned} \mathrm{lh}(s_1) &\leq A\Big(\mathrm{lh}(\models) + \mathrm{lh}(\hat{p}) + \mathrm{lh}(\underline{\varphi}) + m\,\mathrm{lh}(\langle\rangle) + \sum_{i \leq m} \mathrm{lh}(\underline{t_i}) + \mathrm{lh}(s_0) \Big) \\ &\leq 2A\,\mathrm{lh}(s_0), \\ \mathrm{lh}(s_2) &\leq A\Big\{ \mathrm{lh}(\models) + \mathrm{lh}(\hat{p}) + \mathrm{lh}(\underline{\varphi}) + m\,\mathrm{lh}(\langle\rangle) \\ &\qquad + \sum_{1 \leq i \leq m} \mathrm{lh}(\underline{t_i}) + \mathrm{lh}(s_0) + \mathrm{lh}(s_1) \Big\} \end{aligned}$$

$$
\begin{aligned}
&\le A\Big\{ \operatorname{lh}(\models) + \operatorname{lh}(\hat{p}) + \operatorname{lh}(\underline{\varphi}) + m\operatorname{lh}(\langle\rangle) \\
&\quad + \sum_{1\le i\le m} \operatorname{lh}(\underline{t_i})\} + A\{\operatorname{lh}(s_0) + \operatorname{lh}(s_1))\Big\} \\
&\le A\operatorname{lh}(s_0) + A\{2^0A^0\operatorname{lh}(s_0) + 2^1A^1\operatorname{lh}(s_0)\} \\
&\le \Big(A + \sum_{0\le j<2} 2^jA^j\Big)\operatorname{lh}(s_0) \le \Big(1 + \sum_{0\le j<2} 2^j\Big)A^2\operatorname{lh}(s_0) \le 2^2A^2\operatorname{lh}(s_0), \\
&\ \ \vdots \\
\operatorname{lh}(s_r) &\le A\Big\{ \operatorname{lh}(\models) + \operatorname{lh}(\hat{p}) + \operatorname{lh}(\underline{\varphi}) + m\operatorname{lh}(\langle\rangle) + \sum_{1\le i\le m} \operatorname{lh}(\underline{t_i}) + \sum_{0\le i<r} \operatorname{lh}(s_i)\Big\} \\
&\le A\Big\{ \operatorname{lh}(\models) + \operatorname{lh}(\hat{p}) + \operatorname{lh}(\underline{\varphi}) + m\operatorname{lh}(\langle\rangle) + \sum_{1\le i\le m} \operatorname{lh}(\underline{t_i})\Big\} + A\sum_{0\le i<r} \operatorname{lh}(s_i) \\
&\le A\operatorname{lh}(s_0) + A\sum_{0\le j<r} 2^jA^j\operatorname{lh}(s_0) \\
&\le \Big(1 + \sum_{0\le j<r} 2^j\Big)A^r\operatorname{lh}(s_0) \le 2^rA^r\operatorname{lh}(s_0).
\end{aligned}
$$

The length of the sequence $t_1, \dots, t_m$ is less than $\operatorname{lh}(\varphi) \le \log^*(a)$. We can estimate the elements in the sum for $\operatorname{lh}(s_0)$ from (4.3) as follows:

$$
\begin{aligned}
\operatorname{lh}(\hat{p}) &\le A(2\operatorname{lh}(\underline{k - C}) + \operatorname{lh}(s^\gamma)), \\
\operatorname{lh}(\varphi) &\le \log^*(a), \\
m\operatorname{lh}(\langle\rangle) &\le \log^*(a)\operatorname{lh}(\langle\rangle),
\end{aligned}
$$

Thus, the length of s_0 can be estimated by

$$
\begin{aligned}
\operatorname{lh}(s_0) &\le A\Big\{ \operatorname{lh}(\models) + \operatorname{lh}(\hat{p}) + \operatorname{lh}(\underline{\varphi}) + m\operatorname{lh}(\langle\rangle) + \sum_{1\le i\le m} \operatorname{lh}(t_i)\Big\} \\
&\le A\Big\{2A\log(k) + \log^*(a) + \log^*(a)\log^*(a) + \sum_{1\le i\le m} \operatorname{lh}(t_i)\Big\} + \mathbb{N}.
\end{aligned}
$$

Since $a \le 2^{2^k}$, it follows that

$$
\operatorname{lh}(s_0) \le A\Big\{3A\log(k) + \sum_{1\le i\le m} \operatorname{lh}(t_i)\Big\}.
$$

Since the length of the greatest term s_r is not greater than $2^rA^r\operatorname{lh}(s_0)$, we can bound the size of s_r by

$$
\begin{aligned}
s_r &\le 2^{(2A)^r\operatorname{lh}(s_0)} \le 2^{(2A)^rA\{3A\log(k) + \sum_{1\le i\le m}\operatorname{lh}(t_i)\}} \\
&\le 2^{(2A)^r3A^2\log(k)}\Big(\prod_{i\le m} t_i\Big)^{(2A)^rA}.
\end{aligned}
$$

Hence, because $\varphi \prod_{i\leq m} t_i \leq 2^{2^{k-C}}$,

$$s_r \leq 2^{(2A)^r 3A^2 \log(k)} (2^{2^{k-C}})^{(2A)^r A} \leq 2^{(2A)^r 3A^2 \log(k)} 2^{2^{k-C}(2A)^r A}$$
$$\leq 2^{3\log(k)(2A)^r A^2} 2^{2^{k-C+r\log(2A)+\log(A)}} \leq 2^{2^{k-C+r\log(2A)+\log(A)}+3\log(k)(2A)^r A^2}$$

and, because $k > \mathbb{N}$ and $2^{k-C+r\log(2A)+\log(A)} > 3\log(k)(2A)^r A^2$,

$$s_r \leq 2^{2^{k-C+r\log(2A)\log(A)+1}} \leq 2^{2^k}.$$

The last inequality is true since $r \leq \log(\models)$ and we have chosen C so that $C \geq (\models)\log(2A) + \log(A) + 1$.

It is also easy to see that $\mathrm{dp}(s_r) \leq k + \mathbb{N} < b$. This completes the proof of Claim 4.7. ∎

We need to show that $\tilde{p}$ reflects $\hat{p}$ not only for equality but for all formulas of size $\log^*(a)$.

Claim 4.8. *For each $\varphi \leq \log^*(a)$, for each $t_1, \ldots, t_m \in \tilde{\Lambda}$ such that $(\varphi, t_1, \ldots, t_m)$ are g.e. for $(\tilde{\Lambda}, \tilde{p})$,*

$$\tilde{p} \models \varphi[t_1, \ldots, t_m] \;\Leftrightarrow\; p \models \text{“}\hat{p} \models \underline{\varphi}[\underline{t_1}, \ldots, \underline{t_m}]\text{”}.$$

Proof. The proof is by induction on φ. We use the fact that, by Claim 4.7, if $(\varphi, t_1, \ldots, t_m)$ is g.e. for $\tilde{\Lambda}$ then $(\models, \hat{p}, \underline{\varphi}, \langle \underline{t_1}, \ldots, \underline{t_m}\rangle)$ is g.e. for Λ. So, it makes sense to write $p \models \text{“}\hat{p} \models \underline{\varphi}[\underline{t_1}, \ldots, \underline{t_m}]\text{”}$ whenever we may write $\tilde{p} \models \varphi[t_1, \ldots, t_m]$.

For φ atomic the statement follows from the definition of $\tilde{p}$.

Since, by the choice of N,

$$p \models \text{“}\hat{p} \text{ satisfies Tarski conditions for propositional connectives”},$$

it follows that the equivalence holds for all quantifier free formulas. Now, let us consider the case where φ is of the form $\exists y\, \psi(y, \bar{z})$. Let us assume that $\tilde{p} \models \exists y\, \psi(y, t_1, \ldots, t_m)$. Then

$$\tilde{p} \models \psi[s^{\exists y\, \psi}(t_1, \ldots, t_m), t_1, \ldots, t_m].$$

By the inductive assumption,

$$p \models \text{“}\hat{p} \models \underline{\psi}[\underline{s^{\exists y\, \psi}(t_1, \ldots, t_m)}, \underline{t_1}, \ldots, \underline{t_m}]\text{”}.$$

Since all terms $\underline{s^{\exists y\, \psi}(t_1, \ldots, t_m)}, \underline{t_1}, \ldots \underline{t_m}$ are in $\{\underline{0}, \ldots, \underline{2^{2^k}}\}$ it follows that p decides that $\underline{s^{\exists y\, \psi}(t_1, \ldots, t_m)}$ is a witnessing term for $\underline{\exists y\, \psi}$ and $\underline{t_1}, \ldots, \underline{t_m}$. Thus,

$$p \models (\text{“}\hat{p} \models \underline{\psi}[\underline{s^{\exists y\, \psi}(t_1, \ldots, t_m)}, \underline{t_1}, \ldots, \underline{t_m}]\text{”} \;\Leftrightarrow\; \text{“}\hat{p} \models \underline{\exists y\, \psi}[\underline{t_1}, \ldots, \underline{t_m}]\text{”}).$$

Let us observe that in the above formula ψ and $\exists y\, \psi$ are given as terms. Thus, p does not need to decide these formulas to decide positively the above equivalence. The proof of the implication from $p \models \text{“}\hat{p} \models \underline{\exists y\, \psi}[\underline{t_1}, \ldots, \underline{t_m}]\text{”}$ to $p \models \exists y\, \psi[t_1, \ldots, t_m]$ goes exactly in the same way.

Finally, let $\tilde{p} \models \forall x\, \psi[t_1, \dots, t_m]$. This is equivalent to

$$\forall t \in \tilde{\Lambda}\, \tilde{p} \models \psi[t, t_1, \dots, t_m] \quad \text{and} \quad \forall t \in \tilde{\Lambda}\, p \models \text{“}\hat{p} \models \underline{\psi}[\underline{t}, \underline{t_1}, \dots, \underline{t_m}]\text{”}.$$

But, by (4.2),

$$\forall t\, (t \in \tilde{\Lambda} \Leftrightarrow p \models \Lambda[\underline{t}, \underline{k - C}, \underline{(1+\varepsilon')b}]).$$

Thus

$$p \models \forall z\, (z \in \Lambda(\underline{k - C}, \underline{(1+\varepsilon')b}) \Rightarrow \hat{p} \models \underline{\psi}[\underline{t_1}, \dots, \underline{t_m}, z]),$$

and this is just the definition of

$$p \models \text{“}\hat{p} \models \underline{\forall x\, \psi}[\underline{t_1}, \dots, \underline{t_m}]\text{”}.$$

Again, we skip the proof of the other implication. ∎

Since $\hat{p}$ is (under p) a T-evaluation, the following is true under p:

$$\forall x \leq \log^*(a)\, ((x \in T \wedge \text{“}x \text{ is g.e. for } \Lambda(\underline{k - C}, \underline{(1+\varepsilon')b})\text{”} \Rightarrow \hat{p} \models x).$$

Then, by Claims 4.7 and 4.8, we have in M,

$$\forall \varphi \leq \log^*(a)\, (\varphi \in T \wedge \text{“}\varphi \text{ is g.e. for } \tilde{\Lambda}\text{”} \Rightarrow \tilde{p} \models \varphi).$$

Thus $\tilde{p}$ is a T-evaluation. Moreover, $\tilde{p}$ decides formulas less than or equal to N because

$$p \models \text{“}\hat{p} \text{ decides formulas less than or equal to } \underline{N}\text{,”}$$

and this property is also transferred to $\tilde{p}$ by Claim 4.8. So, $(\tilde{\Lambda}, \tilde{p}, \tilde{k}, \tilde{b})$ satisfies the third condition for being suitable. This completes the proof of Lemma 4.5. ∎

Now, we are ready to formulate the induction.

PROPOSITION 4.9. *Let N be chosen as above, and $T = \mathrm{I}\Delta_0 + \Omega_1$. Assume that for some $\varepsilon > 0$, $T \vdash \mathrm{HCons}((1+\varepsilon)\log^3, N)$, $M \models T$ and $a \in \log^3(M)$. Then, for all $i \leq a/C^2$ there exist Λ_i, p_i such that*

$$(\Lambda_i, p_i, a - iC, (1+\varepsilon/2)^{i+1} a)$$

is suitable.

Proof. We prove the conclusion by induction on $i \leq a/C^2$. To carry out the induction we should take a sufficiently large parameter to express the induction formula as a bounded formula. To see that such a parameter exists we should estimate the size of Λ_0 and p_0 which are the greatest among Λ_i, p_i for $i \leq a/C^2$. Λ_0 is a set of terms 2^{2^a}. Thus,

$$\Lambda_0 \leq 2^{2^{2^a}}.$$

Since $a \in \log^3$, $\Lambda_0 \in M$. An evaluation p_0 is just a 0-1 function from the set of pairs of terms from Λ and we can bound p_0 as

$$p_0 \leq 2^{\mathrm{card}(\Lambda)^2} = 2^{(2^{2^a})^2} = 2^{2^{2 \cdot 2^a}} = 2^{2^{2^{a+1}}}.$$

In T, the third logarithm of a model is closed under successors. Thus, $p_0 \in M$ and we can bound the induction formula by an element from M.

To start the induction, the existence of a suitable $(\Lambda_0, p_0, a, (1+\varepsilon/2)a)$ is guaranteed by our assumption that $T \vdash \mathrm{HCons}(N, T, (1+\varepsilon)\log^3)$.

To prove the induction step let us assume that for some $i \leq a/C^2$ there are Λ_i, p_i such that the sequence $(\Lambda_i, p_i, a - iC, (1+\varepsilon/2)^{i+1}a)$ is suitable. Then $(\Lambda_i, p_i, a - iC, (1+\varepsilon/2)^{i+1}a)$ satisfies the assumptions of Lemma 4.5. It follows that there exists $(\tilde{\Lambda}, \tilde{p}, a - (i+1)C, (1+(\varepsilon/2))^{i+2}a)$ which is suitable. And this is just the $(i+1)$th sequence. ∎

LEMMA 4.10. *Let N, C be large constants chosen as above. Let $T = \mathrm{I}\Delta_0 + \Omega_1$. Assume that for some $\varepsilon > 0$, $T \vdash \mathrm{HCons}(N, T, (1+\varepsilon)\log^3)$. Then for each model $M \models T$ and for each $a \in \log^3(M)$ there exists a model $M' \models T$ such that $M' \restriction a = M \restriction a$ and $a \in \frac{C^2}{\log(1+\varepsilon/2)} \log^4(M')$.*

Proof. By Proposition 4.9, for $i = a/C^2$, there exists a suitable sequence

$$(\Lambda, p, a(1-1/C), (1+\varepsilon/2)^{a/C^2}a).$$

Now, for any cut $2a < I < (1+\varepsilon/2)^{a/C^2}a$ if we take terms from Λ of depths in I we can define from p a model M' for T. By Lemma 4.4 and Proposition 2.11,

$$M' \models \exists z\, \{z = \exp^3((1+\varepsilon/2)^{a/C^2}a - C)\}.$$

But

$$(1+\varepsilon/2)^{a/C^2}a - C = 2^{\log(1+\varepsilon/2)C^{-2}a + \log(a)} - C \geq 2^{\log(1+\varepsilon/2)C^{-2}a}.$$

It follows that

$$M' \models 2^{\log(1+\varepsilon/2)C^{-2}a} \in \log^3$$

which means that $a \in \frac{C^2}{\log(1+\varepsilon/2)} \log^4(M')$. Moreover, since $\{\underline{0}, \ldots, \underline{a}\} \subseteq \Lambda$, we conclude, by Lemma 2.14, that $M' \restriction a = M \restriction a$. ∎

THEOREM 4.11. *Let N be chosen as above, and let $T = \mathrm{I}\Delta_0 + \Omega_1$. Assume that for some $\varepsilon > 0$, $T \vdash \mathrm{HCons}(N, T, (1+\varepsilon)\log^3)$. Then for each model $M \models T$ and each $a \in \log^3(M)$ there exists a model $M' \models T$ such that $M' \restriction a = M \restriction a$ and $a \in \log^4(M')$.*

Proof. It suffices to use Lemma 4.10 twice. Indeed, let $M \models T$, $a \in \log^3(M)$ and let $D = C^2/\log(1+\varepsilon/2) \in \mathbb{N}$. Then, by Lemma 4.10, there exists $M'' \models T$ such that $M'' \restriction a = M \restriction a$ and $a \in D\log^4(M'')$. Thus, $2^{a/D} \in \log^3(M'')$ and, again by Lemma 4.10, there exists $M' \models T$ such that $M' \restriction 2^{a/D} = M \restriction 2^{a/D}$ and $2^{a/D} \in D\log^4(M')$. Thus,

$$2^{a/D}/D \in \log^4(M').$$

Since $a > \mathbb{N}$ and $D \in \mathbb{N}$, $a < 2^{a/D}/D$. So, we have $a \in \log^4(M')$. This ends the proof of the theorem. ∎

Now, we may formulate and prove the main result of the paper.

THEOREM 4.12. *Let $T = \mathrm{I}\Delta_0 + \Omega_i$. Let N be a sufficiently large integer. Then, for any $\varepsilon > 0$, T does not prove its Herbrand consistency restricted to terms of depth not greater than $(1+\varepsilon)\log^{i+2}$, that is, $T \nvdash \mathrm{HCons}(N, T, (1+\varepsilon)\log^{i+2})$.*

Proof. This is an easy consequence of Theorems 4.11 and 2.17. ∎

Acknowledgements. The authors would like to thank the anonymous referee for his/her careful reading of the paper and providing comments which significantly improved the quality of the exposition.

This research was partially supported by grant N N201 382234 of the Polish Ministry of Science and Higher Education.

References

[A01] Z. Adamowicz, *On tableaux consistency in weak theories*, preprint 618, Inst. Math., Polish Acad. Sci., 2001.

[A02] —, *Herbrand consistency and bounded arithmetic*, Fund. Math. 171 (2002), 279–292.

[AK04] Z. Adamowicz and L. A. Kołodziejczyk, *Well-behaved principles alternative to bounded induction*, Theoret. Comput. Sci. 322 (2004), 5–16.

[AZ01] Z. Adamowicz and P. Zbierski, *On Herbrand consistency in weak arithmetics*, Arch. Math. Logic 40 (2001), 399–413.

[HP93] P. Hájek and P. Pudlák, *Metamathematics of First-Order Arithmetic*, Springer, 1993.

[K06a] L. A. Kołodziejczyk, *On the Herbrand notion of consistency for finitely axiomatizable fragments of bounded arithmetic theories*, J. Symbolic Logic 71 (2006), 624–638.

[K06b] —, a personal communication.

[WP87] A. J. Wilkie and J. B. Paris, *On the scheme of induction for bounded arithmetical formulas*, Ann. Pure Appl. Logic 35 (1987), 261–302.

Zofia Adamowicz, Konrad Zdanowski
Institute of Mathematics
Polish Academy of Sciences
00-956 Warszawa, Poland
E-mail: Z.Adamowicz@impan.pl
K.Zdanowski@impan.pl

Received 11 January 2008;
in revised form 10 December 2009 and 15 November 2010