

HERBRAND CONSISTENCY OF SOME ARITHMETICAL THEORIES

SAEED SALEHI

Abstract. Gödel's second incompleteness theorem is proved for Herbrand consistency of some arithmetical theories with bounded induction, by using a technique of logarithmic shrinking the witnesses of bounded formulas, due to Z. Adamowicz [*Herbrand consistency and bounded arithmetic*, *Fundamenta Mathematicae*, vol. 171 (2002), pp. 279–292]. In that paper, it was shown that one cannot always shrink the witness of a bounded formula logarithmically, but in the presence of Herbrand consistency, for theories $\text{I}\Delta_0 + \Omega_m$ with $m \geq 2$, any witness for any bounded formula can be shortened logarithmically. This immediately implies the unprovability of Herbrand consistency of a theory $T \supseteq \text{I}\Delta_0 + \Omega_2$ in T itself.

In this paper, the above results are generalized for $\text{I}\Delta_0 + \Omega_1$. Also after tailoring the definition of Herbrand consistency for $\text{I}\Delta_0$ we prove the corresponding theorems for $\text{I}\Delta_0$. Thus the Herbrand version of Gödel's second incompleteness theorem follows for the theories $\text{I}\Delta_0 + \Omega_1$ and $\text{I}\Delta_0$.

§1. Introduction. By Gödel's first incompleteness theorem Truth is not the same as Provability in sufficiently strong theories. In other words, Provable is a proper subset of True, and thus True is not conservative over Provable. It is not even Π_1 -conservative; i.e., there exists a Π_1 -formula, in theories which can interpret enough arithmetic, which is true but unprovable in those theories. Thus one way of comparing the strength of a theory T over one of its sub-theories S is considering the Π_1 -conservativeness of T over S . Indeed, Gödel's second incompleteness theorem provides a Π_1 -candidate for this purpose: $\text{Con}(S)$, the statement of the consistency of S . By that theorem we have $S \not\vdash \text{Con}(S)$, but if $T \vdash \text{Con}(S)$ then T is not Π_1 -conservative over S .

Examples abound in mathematics and logic: Zermelo–Frankel Set Theory ZFC is not Π_1 -conservative over Peano's Arithmetic PA, because $\text{ZFC} \vdash \text{Con}(\text{PA})$ but $\text{PA} \not\vdash \text{Con}(\text{PA})$. Inside PA the Σ_n -hierarchy is not a Π_1 -conservative hierarchy, since $\text{I}\Sigma_{n+1} \vdash \text{Con}(\text{I}\Sigma_n)$ though $\text{I}\Sigma_n \not\vdash \text{Con}(\text{I}\Sigma_n)$; see e.g., [7]. Then below the theory $\text{I}\Sigma_1$ things get more complicated: for Π_1 -separating $\text{I}\Delta_0 + \text{Exp}$ over $\text{I}\Delta_0$ the candidate $\text{Con}(\text{I}\Delta_0)$ does not work, because $\text{I}\Delta_0 + \text{Exp} \not\vdash \text{Con}(\text{I}\Delta_0)$. For this Π_1 -separation, Paris and Wilkie [10] suggested the notion of cut-free consistency instead of the usual—Hilbert style—consistency predicate. Here one can show that $\text{I}\Delta_0 + \text{Exp} \vdash \text{CFCon}(\text{I}\Delta_0)$, and then it was presumed that $\text{I}\Delta_0 \not\vdash \text{CFCon}(\text{I}\Delta_0)$, where

Received June 1, 2010.

2010 *Mathematics Subject Classification*. Primary 03F40, 03F30; Secondary 03F05, 03H15.

Key words and phrases. Cut-free provability, Herbrand provability, bounded arithmetics, weak arithmetics, Gödel's Second Incompleteness Theorem.

This research was partially supported by the grant N^o86030011 of the Institute for Research in Fundamental Sciences (IPM), Tehran, Iran.

CFCCon stands for cut-free consistency. But this presumption took a rather long time to be established. Meanwhile, Pudlák in [11] established the Π_1 -separation of $\text{I}\Delta_0 + \text{Exp}$ over $\text{I}\Delta_0$ by other methods, and mentioned the unprovability of $\text{CFCCon}(\text{I}\Delta_0)$ in $\text{I}\Delta_0$ as an open problem. This problem is interesting in its own right. Indeed Gödel's second incompleteness theorem has been generalized to all consistent theories containing Robinson's Arithmetic Q , in the case of Hilbert consistency; see [7]. But for cut-free consistency it is still an open problem whether the theorem holds for Q , and its not too strong extensions. This is a double strengthening of Gödel's second incompleteness theorem: weakening the theory and weakening the consistency predicate. Let us note that since cut-free provability is stronger than the usual Hilbert provability (with a super-exponential cost), then cut-free consistency is a weaker notion of consistency. Indeed, proving Gödel's second incompleteness theorem for weak notions of consistencies in weak arithmetics turns out to be a difficult problem. We do not intend here to give a thorough history of this ongoing research area, let us just mention a few results:

- Z. Adamowicz was the first one to demonstrate the unprovability of cut-free consistency in bounded arithmetics, by proving in an unpublished manuscript in 1999 (later appeared as a technical report [1]) that the tableau consistency of $\text{I}\Delta_0 + \Omega_1$ is not provable in itself. Later with P. Zbierski (2001) she proved Gödel's second incompleteness theorem for the Herbrand consistency of $\text{I}\Delta_0 + \Omega_2$ (see [3]), and a bit later she gave a model theoretic proof of it in 2002; see [2].
- D. E. Willard introduced an $\text{I}\Delta_0$ -provable Π_1 -formula V and showed that any theory whose axioms contains $\text{Q} + V$ cannot prove its own tableaux consistency. He also showed that tableaux consistency of $\text{I}\Delta_0$ is not provable in itself, see [16, 17]; this proved the conjecture of Paris and Wilkie mentioned above.
- S. Salehi (see [13] Chapter 3 and also [12]) showed the unprovability of Herbrand consistency of a re-axiomatization of $\text{I}\Delta_0$ in itself, the proof of which was heavily based on [3]. The re-axiomatization used PA^- , the theory of the positive fragment of a discretely ordered ring, as the base theory, instead of Q , and assumed two $\text{I}\Delta_0$ -derivable sentences as axioms. Also the model-theoretic proof of Z. Adamowicz in [2] was generalized to the $\text{I}\Delta_0 + \Omega_1$ case in Chapter 5 of [13]. A polished and updated proof of it appears in the present paper.
- L. A. Kołodziejczyk showed in [8] that the notion of Herbrand consistency cannot Π_1 -separate the hierarchy of bounded arithmetics (this Π_1 -separation is still an open problem). Main results are the existence of an n for any given $m \geq 3$ such that $S_m \not\vdash \text{HCon}(S_m^n)$, and the existence of a natural $n \in \mathbb{N}$ such that $\bigcup_m S_m \not\vdash \text{HCon}(S_3^n)$, where HCon stands for Herbrand consistency.
- Z. Adamowicz and K. Zdanowski have obtained some results on the unprovability of the relativized notion of Herbrand consistency in theories containing $\text{I}\Delta_0 + \Omega_1$; see [4]. Their paper contains some insightful ideas about the notion of Herbrand consistency.

For $\text{I}\Delta_0 + \Omega_1$ the arguments are rather smoother, in comparison to the case of $\text{I}\Delta_0$. Our proof for the main theorem on $\text{I}\Delta_0 + \Omega_1$ borrows many ideas from [2], the major difference being the coding techniques and making use of a more liberal definition of Herbrand consistency. The definition of HCon given in [3] and [2] depends on a special coding given there. For reading the present paper no familiarity with [3] is needed, but a theorem of [2] will be of critical use here (Theorem 3.6). We will even

use a modified version of it (Theorem 4.2). For $\text{I}\Delta_0$ we will see that our definition of HCon is not best suited for this theory; and we will actually tailor it for $\text{I}\Delta_0$. A hint for the obstacles in tackling Herbrand consistency in $\text{I}\Delta_0$ can be found in Chapters 3 and 4 of [13].

In Section 2 we introduce the ingredients of Herbrand's theorem from scratch, and then explain how they can be arithmetized by Gödel coding. This sets the stage for Section 3 in which we formalize the notion of Herbrand model and use it to prove our main theorem for $\text{I}\Delta_0 + \Omega_1$. Finally in Section 4 we modify our definitions and theorems to fit the $\text{I}\Delta_0$ case. After pinpointing the places where we have made an essential use of Ω_1 , we do some tailoring for $\text{I}\Delta_0$, and prove our main result for $\text{I}\Delta_0$. We finish with some conclusions and open questions.

§2. Basic definitions and arithmetizations. This section introduces the notions of Herbrand provability and Herbrand consistency, and a way of formalizing and arithmetizing these concepts. The first subsection can be read by any logician. The second subsection gets more technical with Gödel coding, for which some familiarity with [7] is presumed.

2.1. Herbrand consistency. Skolemizing a formula is usually performed on prenex normal forms (see e.g., [6]), and since prenex normalizing a formula is not necessarily done in a unique way, then one may get different Skolemized forms of a formula. For example, the tautology $F = \forall x \phi(x) \rightarrow \forall x \phi(x)$ can be prenex normalized into either $\forall x \exists y (\phi(y) \rightarrow \phi(x))$ or $\exists y \forall x (\phi(y) \rightarrow \phi(x))$. These two formulas can be Skolemized respectively as $\phi(f(x)) \rightarrow \phi(x)$ and $\phi(c) \rightarrow \phi(x)$, where f is a new unary function symbol, and c is a new constant symbol. Here we briefly describe a way of Skolemizing a (not-necessarily prenex normal) formula which results in a somehow unique (up to a variable renaming) formula.

A formula is in *negation normal* form when the implication symbol does not appear in it, and the negation symbol appears in front of atomic formulas only. A formula can be (uniquely) negation normalized by the following rewriting rules:

$$\begin{aligned} (A \rightarrow B) &\iff (\neg A \vee B); & \neg\neg A &\iff A; \\ \neg(A \vee B) &\iff (\neg A \wedge \neg B); & \neg(A \wedge B) &\iff (\neg A \vee \neg B); \\ \neg\forall x A(x) &\iff \exists x \neg A(x); & \neg\exists x A(x) &\iff \forall x \neg A(x). \end{aligned}$$

A formula is called *rectified* if no variable appears both bound and free in it, and different quantifiers refer to different variables. A formula is called *rectified negation normal* if it is both negation normalized and rectified. Again, any formula can be rectified. Indeed, any given formula is equivalent to its rectified negation normal form (RNNF) which can be obtained from the formula in a unique (up to a variable renaming) way (see e.g., [5]).

Now we introduce Skolem functions for existential formulas: for any (not necessarily RNNF) formula of the form $\exists x A(x)$, let $f_{\exists x A(x)}$ be a new m -ary function symbol where m is the number of the free variables of $\exists x A(x)$. When $m = 0$ then $f_{\exists x A(x)}$ will obviously be a new constant symbol (cf. [6]).

DEFINITION 2.1. Let φ be an RNNF formula. Define φ^S by induction:

- $\varphi^S = \varphi$ for atomic or negated-atomic φ ;
- $(\varphi \circ \psi)^S = \varphi^S \circ \psi^S$ for $\circ \in \{\wedge, \vee\}$ and RNNF formulas φ, ψ ;

- $(\forall x\varphi)^S = \forall x\varphi^S$;
- $(\exists x\varphi)^S = \varphi^S[f_{\exists x\varphi(x)}(\bar{y})/x]$ where \bar{y} are the free variables of $\exists x\varphi(x)$ and the formula $\varphi^S[f_{\exists x\varphi(x)}(\bar{y})/x]$ results from the formula φ^S by replacing all the occurrences of the variable x with the term $f_{\exists x\varphi(x)}(\bar{y})$.

The Skolemized form of any (not necessarily RNNF) formula ψ is obtained in the following way: using the above rewriting rules we negation normalize ψ and then rename the repetitive variables (if any) to get a rectified negation normal form of ψ , say φ . Then we get φ^S by the above definition, and remove all the (universal) quantifiers in it (together with the variables next to them). We denote thus resulting Skolemized form of ψ by ψ^{Sk} . \dashv

Note that ψ^{Sk} can be obtained from ψ in a unique (up to a variable renaming) way, and it is an open (quantifier-free) formula. For the above example F , assuming that ϕ is atomic, we get $F^S = (\exists y \neg\phi(y) \vee \forall x \phi(x))^S = \neg\phi(c) \vee \forall x \phi(x)$, and thus $F^{\text{Sk}} = \neg\phi(c) \vee \phi(x) \equiv \phi(c) \rightarrow \phi(x)$.

DEFINITION 2.2. A Skolem instance of the formula ψ is any formula resulting from substituting the free variables of ψ^{Sk} with some closed (variable-free) terms. So, if x_1, \dots, x_n are the free variables of ψ^{Sk} (thus written as $\psi^{\text{Sk}}(x_1, \dots, x_n)$) then a Skolem instance of ψ is $\psi^{\text{Sk}}[t_1/x_1, \dots, t_n/x_n]$ where t_1, \dots, t_n are closed terms (which could be constructed from the Skolem functions symbols).

The Skolemized form of a theory T is by definition $T^{\text{Sk}} = \{\varphi^{\text{Sk}} \mid \varphi \in T\}$. \dashv

We are now ready to state an important theorem discovered by Herbrand (probably by also Skolem and Gödel). This theorem has got some few names, and by now is a classical theorem in Mathematical Logic. Here we state a version of the theorem which we will need in the paper. The proof is omitted, though it is not too difficult to prove it directly (see e.g., [5]).

THEOREM 2.3 (Herbrand). *Any theory T is equiconsistent with its Skolemized theory T^{Sk} . Any theory T is consistent if and only if every finite set of Skolem instances of T is (propositionally) satisfiable.* \dashv

We will use the above theorem, which reduces the consistency of a first-order theory to the satisfiability of a propositional theory, for the definition of Herbrand Consistency: a theory T is Herbrand consistent when every finite set of Skolem instances of T is propositionally satisfiable. One other concept is needed for formalizing Herbrand consistency of arithmetical theories: *evaluation*.

CONVENTION 2.4. Throughout the paper we deal with closed (or ground) terms (i.e., terms with no variable) and for simplicity we call them “terms”. For this to make sense, we may assume that the language of the theory under consideration has at least one constant symbol. \dashv

DEFINITION 2.5. An *evaluation* is a function from the set of all atomic formulas constructed from a given set of terms Λ to the set $\{0, 1\}$ such that

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

The relation \sim_p on Λ is defined by $t \sim_p s \iff p[t=s] = 1$ for $t, s \in \Lambda$. \dashv

LEMMA 2.6. *The relation \sim_p defined above is an equivalence relation.*

PROOF. For $\varphi(x) \equiv (s=x)$ from $p[t=s] = 1$ one can infer $p[s=t] = p[\varphi(t)] = p[\varphi(s)] = p[s=s] = 1$. So, $t \sim_p s$ implies $s \sim_p t$. Also for $\phi(x) \equiv (t=x)$ the condition $p[s=r] = 1$ implies $p[t=s] = p[\phi(s)] = p[\phi(r)] = p[t=r]$. So, \sim_p is a symmetric and transitive (also, by definition, a reflexive) relation. \dashv

NOTATION 2.7. The \sim_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 formulas in the usual way:

- $p \models \varphi \wedge \psi$ if and only if $p \models \varphi$ and $p \models \psi$;
- $p \models \varphi \vee \psi$ if and only if $p \models \varphi$ or $p \models \psi$;
- $p \models \neg\varphi$ if and only if $p \not\models \varphi$. \dashv

Let us note that \sim_p is a congruence relation as well. That is, for any set of terms t_i and s_i ($i = 1, \dots, n$) and function symbol f , if $p \models t_1 = s_1 \wedge \dots \wedge t_n = s_n$ then $p \models f(t_1, \dots, t_n) = f(s_1, \dots, s_n)$.

DEFINITION 2.8. If all terms appearing in a 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 . \dashv

For illustrating the above concepts we now present an example.

EXAMPLE 2.9. Take the language $\mathcal{L} = \{g, P, R, S\}$ in which g is a unary function symbol, and P is a binary predicate symbol, and R, S are unary predicate symbols. Let the theory T be axiomatized by:

- $T_1: \forall x \exists y P(x, y)$;
- $T_2: \forall x (R(x) \vee S(gx))$;
- $T_3: \forall x, y (\neg P(x, y) \vee \neg S(x))$.

Let us, for the sake of simplicity, denote $\exists y P(x, y)$ by \mathfrak{f} ; then the Skolemized form of the above theory is:

- $T_1^{\text{Sk}}: P(x, \mathfrak{f}x)$; $T_2^{\text{Sk}}: R(x) \vee S(gx)$; $T_3^{\text{Sk}}: \neg P(x, y) \vee \neg S(x)$.

For a constant symbol c let $\Lambda = \{c, gc, \mathfrak{f}c\}$. Then $P(c, \mathfrak{f}c)$ and $R(c) \vee S(gc)$ are Skolem instances of T (of T_1 and T_2) available in Λ , but the Skolem instance $R(gc) \vee S(ggc)$ of T_2 is not available in Λ . Let us note also that the Skolem instance $\neg P(gc, \mathfrak{f}gc) \vee \neg S(gc)$ of T_3 is not available in Λ .

Let q be an evaluation on Λ whose set of true atomic formulas is $\{P(c, \mathfrak{f}c), R(c)\}$. Then q is a T -evaluation. On the other hand the evaluation r on Λ whose set of true atomic formulas is $\{P(c, \mathfrak{f}c), R(c), S(c)\}$, is not a T -evaluation, though it satisfies all the Skolem instances of T_1 and T_2 which are available in Λ . Note that r does not satisfy the Skolem instance $\neg P(c, \mathfrak{f}c) \vee \neg S(c)$ of T_3 . \dashv

By the above theorem of Herbrand, a theory T is consistent if and only if every finite set of its Skolem instances is satisfiable, if and only if for every finite set of terms Λ there is a T -evaluation on Λ . And for a formula φ , $T \vdash \varphi$ if and only if there exists a finite set of terms Λ such that there is no $(T + \neg\varphi)$ -evaluation on Λ . We call this notion of provability, *Herbrand Provability*; note that *Herbrand*

Consistency of a theory T means the existence of a T -evaluation on any (finite) set of terms.

EXAMPLE 2.10. *In the previous example, let $\varphi = \forall x R(x)$. We show $T \vdash \varphi$ by Herbrand provability. We have $(\neg\varphi)^{\text{RNNF}} = \exists x \neg R(x)$; let c denote the Skolem constant symbol $\mathfrak{f}_{\exists x \neg R(x)}$; so we have $(\neg\varphi)^{\text{Sk}} = \neg R(c)$. Put $\Lambda = \{c, gc, \mathfrak{f}gc\}$, and assume (for the sake of contradiction) that there is a $(T + \neg\varphi)$ -evaluation p on Λ . Then p must satisfy the following Skolem instances of T in Λ : $P(gc, \mathfrak{f}gc)$, $R(c) \vee S(gc)$, and $\neg P(gc, \mathfrak{f}gc) \vee \neg S(gc)$. Whence p must also satisfy $\neg S(gc)$ and $R(c)$. So p cannot satisfy the Skolem instance $\neg R(c)$ of $\neg\varphi$ in Λ . Thus there cannot be any $(T + \neg\varphi)$ -evaluation on Λ ; whence $T \vdash \varphi$.*

Note that finding an appropriate Λ is as complicated as finding a formal proof. For example we could not have taken Λ as $\{c, gc, \mathfrak{f}c\}$, since the evaluation q in the previous example would be a $(T + \neg\varphi)$ -evaluation on that set. \dashv

The following couple of examples give thorough illustrations for the above ideas, and they will be actually used later in the paper.

EXAMPLE 2.11. *Let \mathbf{Q} denote Robinson's Arithmetic over the language of arithmetic $\langle 0, \mathfrak{s}, +, \cdot, \leq \rangle$, where 0 is a constant symbol, \mathfrak{s} is a unary function symbol, $+$, \cdot are binary function symbols, and \leq is a binary predicate symbol, whose axioms are:*

$$\begin{aligned} A_1: \forall x (\mathfrak{s}x \neq 0); & & A_2: \forall x \forall y (\mathfrak{s}x = \mathfrak{s}y \rightarrow x = y); \\ A_3: \forall x (x \neq 0 \rightarrow \exists y [x = \mathfrak{s}y]); & & A_4: \forall x \forall y (x \leq y \leftrightarrow \exists z [x + z = y]); \\ A_5: \forall x (x + 0 = x); & & A_6: \forall x \forall y (x + \mathfrak{s}y = \mathfrak{s}(x + y)); \\ A_7: \forall x (x \cdot 0 = 0); & & A_8: \forall x \forall y (x \cdot \mathfrak{s}y = x \cdot y + x). \end{aligned}$$

Let $\psi = \forall x (x \leq 0 \rightarrow x = 0)$ and $\varphi = \forall x \forall y (x \leq \mathfrak{s}y \rightarrow x = \mathfrak{s}y \vee x \leq y)$. We can show $\mathbf{Q} \vdash \psi$ and $\mathbf{Q} \vdash \varphi$; these will be proved below by Herbrand provability. Suppose \mathbf{Q} has been Skolemized as below:

$$\begin{aligned} A_1^{\text{Sk}}: \mathfrak{s}x \neq 0 & & A_2^{\text{Sk}}: \mathfrak{s}x \neq \mathfrak{s}y \vee x = y; \\ A_3^{\text{Sk}}: x = 0 \vee x = \mathfrak{s}px; & & A_4^{\text{Sk}}: [x \not\leq y \vee x + \mathfrak{h}(x, y) = y] \wedge [x + z \neq y \vee x \leq y]; \\ A_5^{\text{Sk}}: x + 0 = x; & & A_6^{\text{Sk}}: x + \mathfrak{s}y = \mathfrak{s}(x + y); \\ A_7^{\text{Sk}}: x \cdot 0 = 0; & & A_8^{\text{Sk}}: x \cdot \mathfrak{s}y = x \cdot y + x. \end{aligned}$$

Here \mathfrak{p} abbreviates $\mathfrak{f}_{\exists y (x=\mathfrak{s}y)}$ and \mathfrak{h} stands for $\mathfrak{f}_{\exists z (x+z=y)}$.

For a fixed term t , put Σ_t be the following set of terms:

$$\Sigma_t = \{0, t, t + 0, \mathfrak{h}(t, 0), \mathfrak{p}\mathfrak{h}(t, 0), \mathfrak{s}\mathfrak{p}\mathfrak{h}(t, 0), t + \mathfrak{s}\mathfrak{p}\mathfrak{h}(t, 0), \mathfrak{s}(t + \mathfrak{s}\mathfrak{p}\mathfrak{h}(t, 0))\},$$

and suppose that p is an \mathbf{Q} -evaluation on Σ_t . We show that $p \models t \not\leq 0 \vee t = 0$. Note that $\psi^{\text{Sk}} = (x \not\leq 0 \vee x = 0)$. If p is such an evaluation and if $p \models t \leq 0$, then by A_4^{Sk} we have $p \models t + \mathfrak{h}(t, 0) = 0$. Now, either $p \models \mathfrak{h}(t, 0) = 0$ or $p \not\models \mathfrak{h}(t, 0) = 0$. In the former case, we have $p \models t + 0 = 0$ which by A_5^{Sk} implies $p \models t = 0$. In the latter case, by A_3^{Sk} we get $p \models \mathfrak{h}(t, 0) = \mathfrak{s}\mathfrak{p}\mathfrak{h}(t, 0)$, and then $p \models 0 = t + \mathfrak{h}(t, 0) = t + \mathfrak{s}\mathfrak{p}\mathfrak{h}(t, 0) = \mathfrak{s}(t + \mathfrak{p}\mathfrak{h}(t, 0))$ by A_6^{Sk} , which is a contradiction with A_1^{Sk} . Thus we showed that if $p \models t \leq 0$ then necessarily $p \models t = 0$.

Now, for two fixed terms u, v define $\Gamma_{u,v}$ as

$$\Gamma_{u,v} = \{0, u, v, sv, h(u, sv), ph(u, sv), sph(u, sv), u + ph(u, sv), \\ u + sph(u, sv), s(u + ph(u, sv))\}.$$

We show that any Q-evaluation on $\Gamma_{u,v}$ must satisfy $u \not\leq sv \vee u = sv \vee u \leq v$. Note that the Skolemized form of φ is $\varphi^{Sk} = (x \not\leq sy \vee x = sy \vee x \leq y)$. Suppose p is an Q-evaluation on $\Gamma_{u,v}$. Then either $p \models h(u, sv) = 0$ or $p \models h(u, sv) \neq 0$. In the former case, by A_4^{Sk} , we have $p \models u \not\leq sv \vee u + 0 = sv$, and then by A_5^{Sk} , $p \models u \not\leq sv \vee u = sv$. And in the latter case $p \models h(u, sv) = sph(u, sv)$ by A_3^{Sk} , also by A_4^{Sk} we have $p \models u \not\leq sv \vee u + sph(u, sv) = sv$. On the other hand from A_6^{Sk} we get $p \models u + sph(u, sv) = s(u + ph(u, sv))$. Whence we get $p \models u \not\leq sv \vee s(u + ph(u, sv)) = sv$, then by A_2^{Sk} , $p \models u \not\leq sv \vee u + ph(u, sv) = v$, which by A_4^{Sk} implies $p \models u \not\leq sv \vee u \leq v$. Hence, in both cases we showed $p \models u \not\leq sv \vee u = sv \vee u \leq v$. Finally, let us note that one could present a Herbrand proof of $Q \vdash \psi$ and $Q \vdash \varphi$ in a very similar way. \dashv

EXAMPLE 2.12. In the language of Example 2.11, $\langle 0, s, +, \cdot, \leq \rangle$, let ind_ψ be the following induction scheme for the formula $\psi(x)$:

$$\psi(0) \wedge \forall x (\psi(x) \rightarrow \psi(sx)) \rightarrow \forall x \psi(x).$$

Assume for the moment that ψ is an atomic formula. Then the Skolemization of ind_ψ results in $\text{ind}_\psi^{Sk} : \neg\psi(0) \vee (\psi(c) \wedge \neg\psi(sc)) \vee \psi(x)$, where c is the Skolem constant symbol $\exists x (\psi(x) \wedge \neg\psi(sx))$. Then any ind_ψ -evaluation p on the set of terms $\{0, c, sc, t\}$ must satisfy one of the following:

either (1) $p \not\models \psi(0)$ or (2) $p \models \psi(c) \wedge \neg\psi(sc)$ or (3) $p \models \psi(t)$.

Now take $\psi(x)$ to be the existential formula $\exists y \varphi(x, y)$ in which φ is an atomic formula. Then the Skolemized form of ind_ψ will be as

$$\text{ind}_\psi^{Sk} : \neg\varphi(0, u) \vee (\varphi(c, qc) \wedge \neg\varphi(sc, v)) \vee \varphi(x, q(x)),$$

where q is the Skolem function symbol for the formula $\exists y \varphi(x, y)$, and c is the Skolem constant symbol for the sentence $\exists x (\exists w \varphi(x, w) \wedge \forall v \neg\varphi(sx, v))$. The (fresh) variables u, v and x are free.

We will need the case of $\varphi(x, y) = (y \leq x \cdot x \wedge y = x \cdot x)$ in the proof of Theorem 4.3 below. In this case the Skolemized form of ind_ψ is

$$(u \not\leq 0^2 \vee u \neq 0^2) \vee \\ \left((qc \leq c^2 \wedge qc = c^2) \wedge (v \not\leq (sc)^2 \vee v \neq (sc)^2) \right) \vee \\ (q(x) \leq x^2 \wedge q(x) = x^2).$$

The notation q^2 is a shorthand for $q \cdot q$. Define the set of terms Υ by

$$\Upsilon = \{0, 0 + 0, 0^2, c, c^2, c^2 + 0, sc, qc, (sc)^2, (sc)^2 + 0\}$$

and suppose p is an $(Q + \text{ind}_\psi)$ -evaluation on the set of terms $\Upsilon \cup \{t, t^2, q(t)\}$. Then p must satisfy the following Skolem instance (δ) of ind_ψ which is available in the set

$\Upsilon \cup \{t, t^2, q(t)\}$:

$$\begin{aligned} & (0 \not\leq 0^2 \vee 0 \neq 0^2) \vee \\ & \left((qc \leq c^2 \wedge qc = c^2) \wedge ((sc)^2 \not\leq (sc)^2 \vee (sc)^2 \neq (sc)^2) \right) \vee \\ & (q(t) \leq t^2 \wedge q(t) = t^2). \end{aligned}$$

Now since $p \models 0 \cdot 0 = 0 + 0 = 0$ then, by Q's axioms, $p \models 0 \leq 0^2 \wedge 0 = 0^2$, and so p cannot satisfy the first disjunct of (δ) . Similarly, since $p \models (sc)^2 + 0 = (sc)^2$ then $p \models (sc)^2 \leq (sc)^2$, thus p cannot satisfy the second disjunct of (δ) either, because $p \models (sc)^2 = (sc)^2$. Whence, p must satisfy the third disjunct of (δ) , then necessarily $p \models q(t) = t^2$ must hold. \dashv

REMARK 2.13. The arguments of the above two examples can be generalized as follows: if $T \vdash \forall \bar{x} \theta(\bar{x})$ where θ is an open (quantifier-free) RNNF formula, then $(\neg \forall \bar{x} \theta(\bar{x}))^{\text{Sk}} = \neg \theta(\bar{c})$ in which \bar{c} is a sequence of Skolem constant symbols. There exists a set of terms Γ (constructed from the Skolem function and constant symbols of T with \bar{c}) such that there exists no $(T + \neg \forall \bar{x} \theta(\bar{x}))$ -evaluation on Γ . So, for any sequence of terms \bar{t} , if $\Gamma(\bar{t})$ is the set of terms which result from the terms of Γ by substituting \bar{c} with \bar{t} , then any T -evaluation on $\Gamma(\bar{t})$ must satisfy the formula $\theta(\bar{t})$. \dashv

2.2. Arithmetization. Fix \mathcal{L}_A to be our language of arithmetic; one can set $\mathcal{L}_A = \langle 0, 1, +, \cdot, < \rangle$ as e.g., in [9] or $\mathcal{L}_A = \langle 0, s, +, \cdot, \leq \rangle$ as e.g., in [7]. Later it will be clear that choosing this fixed language is not of much importance.

Peano's arithmetic PA is the first-order theory that extends Q (see Example 2.11) by the following induction schema for any arithmetical formula $\varphi(x)$: $\varphi(0) \wedge \forall x (\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x)$. Fragments of PA are extensions of Q with the induction schema restricted to a class of formulas. A formula is called bounded if its every quantifier is bounded, i.e., is either of the form $\forall x \leq t(\dots)$ or $\exists x \leq t(\dots)$ where t is a term; they are read as $\forall x (x \leq t \rightarrow \dots)$ and $\exists x (x \leq t \wedge \dots)$ respectively. It is easy to see that bounded formulas are decidable. The theory $\text{I}\Delta_0$, also called bounded arithmetic, is axiomatized by Q plus the induction schema for bounded formulas. The exponentiation function exp is defined by $\text{exp}(x) = 2^x$; the formula Exp expresses its totality: $(\forall x \exists y [y = \text{exp}(x)])$; let us note that the formula " $y = \text{exp}(x)$ " (the graph of the exponentiation function) can be written by a Δ_0 -formula (see e.g., [7]). The converse of exp is denoted by log which is formally defined as $\text{log } x = \min\{y \mid x \leq \text{exp}(y)\}$; and the initial segment log consists of the logarithms of all elements: $\text{log} = \{x \mid \exists y [\text{exp}(x) = y]\}$. The superscripts above the function symbols indicate the iteration of the functions: $\text{exp}^2(x) = \text{exp}(\text{exp}(x))$, $\text{log}^2 x = \text{log } \text{log } x$; similarly the initial segment log^n is $\{x \mid \exists y [\text{exp}^n(x) = y]\}$. Let us recall that Exp is not provable in $\text{I}\Delta_0$. Between $\text{I}\Delta_0$ and $\text{I}\Delta_0 + \text{Exp}$ a hierarchy of theories is considered in the literature, which has close connections with computational complexity. Define the function ω_m to be $\omega_m(x) = \text{exp}^m((\text{log}^m x) \cdot (\text{log}^m x))$. It is customary to define this function by induction: $\omega_0(x) = x^2$ and $\omega_{n+1}(x) = \text{exp}(\omega_n(\text{log } x))$. Let Ω_m express the totality of ω_m (i.e., $\Omega_m \equiv \forall x \exists y [y = \omega_m(x)]$).

By Gödel's coding method, we now rest assured that the concepts introduced in the previous section all can be formalized (and arithmetized) in the language of arithmetic. But we need just a bit more; and that is an "effective" coding, suitable

for bounded arithmetic. The one we adopt here is taken from Chapter V of [7]. For convenience, and shortening the computations, we introduce the \mathcal{P} notation (for polynomially bounding).

DEFINITION 2.14. We say α is of $\mathcal{P}(\beta)$, when the code of α (denoted $\ulcorner \alpha \urcorner$) is bounded above by a polynomial of β ; and we write this as $\ulcorner \alpha \urcorner \leq \mathcal{P}(\beta)$, meaning that for some n the inequality $\ulcorner \alpha \urcorner \leq \beta^n + n$ holds. \dashv

Let us note that $\alpha \leq \mathcal{P}(\beta)$ is equivalent to the old (more familiar) \mathcal{O} -notation “ $\log \alpha \in \mathcal{O}(\log \beta)$ ”. Here we collect some very basic facts about this fixed efficient coding that will be needed later.

REMARK 2.15. Let A be a set or a sequence of terms, and let $|A|$ denote the cardinality of A , when A is a set, and the same $|A|$ denote the length of A , when A is a sequence. Then

- $\ulcorner \langle \alpha \rangle \urcorner \leq 9(\ulcorner \alpha \urcorner + 1)^2$ (Lemma 3.7.2 page 297 of [7]);
- $\ulcorner A \frown B \urcorner \leq 64 \cdot (\ulcorner A \urcorner \cdot \ulcorner B \urcorner)$ (Proposition 3.29 page 311 of [7]);
- $(|A|) \leq (\log \ulcorner A \urcorner)$ (Definition 3.27 and Section (e) pages 304–310 of [7]);

where $A \frown B$ is the concatenation of (the sequences) A and B . \dashv

If we let $\mathcal{L}_A^{\text{Sk}}$ be the closure of \mathcal{L}_A under Skolem function and constant symbols, i.e., let $\mathcal{L}_A^{\text{Sk}}$ be the smallest set that contains \mathcal{L}_A and for any $\mathcal{L}_A^{\text{Sk}}$ -formula $\exists x \phi(x)$ we have $\text{f}_{\exists x \phi(x)} \in \mathcal{L}_A^{\text{Sk}}$, then this new countable language can also be coded, and this (re-)coding can be generalized to $\mathcal{L}_A^{\text{Sk}}$ -terms and $\mathcal{L}_A^{\text{Sk}}$ -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 $t \leq s$ 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 $\langle t \leq s, i \rangle$ 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 $\mathcal{P}(\prod_{t,s \in \Lambda} \ulcorner t \urcorner \cdot \ulcorner s \urcorner)$.

LEMMA 2.16. For a set of terms Λ and evaluation p on it, $\ulcorner p \urcorner \leq \mathcal{P}(\omega_1(\ulcorner \Lambda \urcorner))$.

PROOF. It follows by the above remark and the afterward explanations, noting that $\prod_{t,s \in \Lambda} \ulcorner t \urcorner \cdot \ulcorner s \urcorner \leq \mathcal{P}(\prod_{t \in \Lambda} (\ulcorner t \urcorner)^{2|\Lambda|}) \leq \mathcal{P}(\ulcorner \Lambda \urcorner)^{2 \log \ulcorner \Lambda \urcorner} \leq \mathcal{P}((\ulcorner \Lambda \urcorner)^{\log(\ulcorner \Lambda \urcorner)})$. \dashv

Let us have another look at the above lemma, which is of great importance. For a set of terms Λ , there are $|\Lambda|$ terms in it (the cardinality of Λ). So, there are $2|\Lambda|^2$ atomic formulas constructed from the terms of Λ (atomic formulas of the form $t=s$ or $t \leq s$ for $t, s \in \Lambda$). And thus, there are $\exp(2|\Lambda|^2)$ different evaluations on the set Λ . And finally note that by $|\Lambda| \leq (\log \ulcorner \Lambda \urcorner)$ we get $\exp(2|\Lambda|^2) \leq \mathcal{P}(\exp((\log \ulcorner \Lambda \urcorner)^2)) \leq \mathcal{P}(\omega_1(\ulcorner \Lambda \urcorner))$. So, in the presence of $\omega_1(\ulcorner \Lambda \urcorner)$ we have all the evaluations on Λ in our disposal.

All these concepts can be expressed in the language of arithmetic \mathcal{L}_A by appropriate formulas. The predicate “Herbrand Consistency of the theory T ” can be arithmetized as “for every set of terms there exists an T -evaluation on it”. Let $\text{HCon}(T)$ denote the \mathcal{L}_A -formula “ T is Herbrand consistent”.

§3. Herbrand models. 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*. Here we use this notion for building up a definable inner model, which will constitute the heart of the proof of our main result for $\mathbf{I}\Delta_0 + \Omega_1$.

3.1. Arithmetically definable Herbrand models. In the sequel, we arithmetize Herbrand models.

DEFINITION 3.1. Let \mathcal{L} be a language and Λ be a set of (ground) terms¹ (constructed by the Skolem constant and function symbols of \mathcal{L}).

Put $\Lambda^{(0)} = \Lambda$, and define inductively

$$\Lambda^{(k+1)} = \Lambda^{(k)} \cup \{f(t_1, \dots, t_m) \mid f \in \mathcal{L} \ \& \ t_1, \dots, t_m \in \Lambda^{(k)}\} \\ \cup \{\mathfrak{f}_{\exists x \psi(x)}(t_1, \dots, t_m) \mid \ulcorner \psi \urcorner \leq k \ \& \ t_1, \dots, t_m \in \Lambda^{(k)}\}.$$

Let $\Lambda^{(\infty)}$ denote the union $\bigcup_{k \in \mathbb{N}} \Lambda^{(k)}$.

Suppose p is an evaluation on $\Lambda^{(\infty)}$. Define $\mathfrak{M}(\Lambda, p) = \{t/p \mid t \in \Lambda^{(\infty)}\}$ and put the \mathcal{L} -structure on it by

- $f^{\mathfrak{M}(\Lambda, p)}(t_1/p, \dots, t_m/p) = f(t_1, \dots, t_m)/p$, and
- $R^{\mathfrak{M}(\Lambda, p)} = \{(t_1/p, \dots, t_m/p) \mid p \models R(t_1, \dots, t_m)\}$;

for $f, R \in \mathcal{L}$ and $t_1, \dots, t_m \in \Lambda^{(\infty)}$.

LEMMA 3.2. *The function $f^{\mathfrak{M}(\Lambda, p)}$ and the relation $R^{\mathfrak{M}(\Lambda, p)}$ are well-defined in DEFINITION 3.1, and when p is an T -evaluation on $\Lambda^{(\infty)}$, for an \mathcal{L} -theory T , then we have $\mathfrak{M}(\Lambda, p) \models T$.*

PROOF. That the functions $f^{\mathfrak{M}(\Lambda, p)}$ and the relations $R^{\mathfrak{M}(\Lambda, p)}$ in DEFINITION 3.1 are well-defined follows directly from the definition of an evaluation (Definition 2.5). By the definition of $\Lambda^{(\infty)}$ 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 $A(t_1/p, \dots, t_m/p)$ if and only if $p \models A(t_1, \dots, t_m)$. Then it can be shown, by induction on ψ 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 $\Lambda^{(\infty)}$. \dashv

We need an upper bound on the size (cardinal) and the code of $\Lambda^{(j)}$.

LEMMA 3.3. *The following inequalities hold when $\ulcorner \Lambda \urcorner$ and $|\Lambda|$ are sufficiently larger than n (when $\exp^4(n) \leq \ulcorner \Lambda \urcorner, |\Lambda|$):*

- (1) $|\Lambda^{(n)}| \leq \mathcal{P}(|\Lambda|^{n!})$, and
- (2) $\ulcorner \Lambda^{(n)} \urcorner \leq \mathcal{P}((\ulcorner \Lambda \urcorner)^{|\Lambda|^{(n+1)!}})$.

PROOF. Denote $\ulcorner \Lambda^{(k)} \urcorner$ by λ_k (thus $\ulcorner \Lambda \urcorner = \lambda_0 = \lambda$) and $|\Lambda^{(k)}|$ by σ_k (and thus $|\Lambda| = \sigma_0 = \sigma$). We first note that $\sigma_{k+1} \leq \sigma_k + M\sigma_k^M + k\sigma_k^k$ for a fixed $M \in \mathbb{N}$. Thus for $k > M$ we have $\sigma_{k+1} \leq \sigma_k + 2k\sigma_k^k \leq \mathcal{P}(\sigma_k^{k+1})$, and then, by an inductive argument, we have $\sigma_n \leq \mathcal{P}(\sigma^{n!})$.

For the second statement, we first compute an upper bound for the code of the Cartesian power A^m for a set A . Similarly the proof of Lemma 2.16, we can show $\ulcorner A^{k+1} \urcorner \leq \mathcal{P}(\prod_{t \in A^k} \ulcorner t \urcorner \cdot \ulcorner s \urcorner) \leq \mathcal{P}(\ulcorner A^{k \cdot |A|} \urcorner \cdot \ulcorner A^{\ulcorner |A| \urcorner} \urcorner)$, and thus $\ulcorner A^m \urcorner \leq \mathcal{P}(\ulcorner A^{\ulcorner |A| \urcorner} \urcorner^m)$ can be proved by induction on m . Now we have $\lambda_{k+1} \leq \mathcal{P}(\ulcorner \Lambda^{(k)} \urcorner \cdot \ulcorner (\Lambda^{(k)})^M \urcorner \cdot \ulcorner (\Lambda^{(k)})^k \urcorner)$ for a fixed M . So, $\lambda_{k+1} \leq \mathcal{P}(\lambda_k^{\sigma_k})$ and finally the conclusion $\lambda_m \leq \mathcal{P}(\lambda^{\sigma^{(m+1)!}})$ follows by induction on m . \dashv

¹Note that we put no restriction on the cardinality of the set Λ .

Stating the above fact as a lemma, despite of the fact that it is indeed a crucial tool for our arguments, let us state the following corollary of it as a theorem, and later on we will use the theorem and will leave the lemma right here.

THEOREM 3.4. *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 $\ulcorner \Lambda^{(j)} \urcorner$ will exist.*

PROOF. There must exist a non-standard j such that $j \leq \log^4(\ulcorner \Lambda \urcorner)$. Thus $2(j+1)! \leq 2^{2^j} \leq \log^2 \ulcorner \Lambda \urcorner$. Now, by Lemma 3.3 we can write

$$\begin{aligned} \ulcorner \Lambda^{(j)} \urcorner &\leq \mathcal{P} \left((\ulcorner \Lambda \urcorner)^{|\Lambda|^{(j+1)!}} \right) &\leq \mathcal{P} \left((2^{2 \log \ulcorner \Lambda \urcorner})^{(\log \ulcorner \Lambda \urcorner)^{(j+1)!}} \right) \\ &\leq \mathcal{P} \left(\exp((\log \ulcorner \Lambda \urcorner)^{2(j+1)!}) \right) &\leq \mathcal{P} \left(\exp(\omega_1(\log \ulcorner \Lambda \urcorner)) \right), \end{aligned}$$

or in other words $\ulcorner \Lambda^{(j)} \urcorner \leq \mathcal{P}(\omega_2(\ulcorner \Lambda \urcorner))$. +

The reason that Theorem 3.4 is stated for non-standard Λ is that the set $\Lambda^{(\infty)}$, needed for constructing the model $\mathfrak{M}(\Lambda, p)$, is not definable in \mathcal{L}_A . But the existence of the definable $\Lambda^{(j)}$ for a non-standard j can guarantee the existence of $\Lambda^{(\infty)}$ and thus of $\mathfrak{M}(\Lambda, p)$. This non-standard j exists for non-standard $\ulcorner \Lambda \urcorner$.

3.2. The main theorem for $\mathbf{I}\Delta_0 + \Omega_1$. Two fundamental theorems were proved by Z. Adamowicz in [2] about the Herbrand Consistency of $\mathbf{I}\Delta_0 + \Omega_m$ for $m \geq 2$:

THEOREM 3.5 (Z. Adamowicz [2]). *For a bounded formula $\theta(\bar{x})$ and $m \geq 2$, if the theory $(\mathbf{I}\Delta_0 + \Omega_m) + \exists \bar{x} \in \log^{m+1} \theta(\bar{x}) + \text{HCon}^{\log^{m-2}}(\mathbf{I}\Delta_0 + \Omega_m)$ is consistent, then so is the theory $(\mathbf{I}\Delta_0 + \Omega_m) + \exists \bar{x} \in \log^{m+2} \theta(\bar{x})$, where $\text{HCon}^{\log^{m-2}}$ is the relativization of HCon to the initial segment \log^{m-2} .* +

THEOREM 3.6 (Z. Adamowicz [2]). *For any natural $m, n \geq 0$ there exists a bounded formula $\eta(\bar{x})$ such that $(\mathbf{I}\Delta_0 + \Omega_m) + \exists \bar{x} \in \log^n \eta(\bar{x})$ is consistent, but $(\mathbf{I}\Delta_0 + \Omega_m) + \exists \bar{x} \in \log^{n+1} \eta(\bar{x})$ is not.* +

These two theorems (by putting $n = m + 1$ for $m \geq 2$) imply together that for any $m \geq 2$: $\mathbf{I}\Delta_0 + \Omega_m \not\vdash \text{HCon}^{\log^{m-2}}(\mathbf{I}\Delta_0 + \Omega_m)$.

Here we extend Theorem 3.5 for $\mathbf{I}\Delta_0 + \Omega_1$, namely we show that

THEOREM 3.7. *For any bounded formula $\theta(x)$, the consistency of the theory*

$$(\mathbf{I}\Delta_0 + \Omega_1) + \exists x \in \log^2 \theta(x) + \text{HCon}(\mathbf{I}\Delta_0 + \Omega_1)$$

implies the consistency of the theory $(\mathbf{I}\Delta_0 + \Omega_1) + \exists x \in \log^3 \theta(x)$.

The rest of this section is devoted to proving this theorem. Let us note that Theorem 3.6 holds already for $\mathbf{I}\Delta_0 + \Omega_1$, and below we reiterate the part that we need here:

THEOREM 3.8 (Z. Adamowicz [2]). *There exists a bounded formula $\eta(\bar{x})$ such that $(\mathbf{I}\Delta_0 + \Omega_1) + \exists \bar{x} \in \log^2 \eta(\bar{x})$ is consistent, but $(\mathbf{I}\Delta_0 + \Omega_1) + \exists \bar{x} \in \log^3 \eta(\bar{x})$ is not consistent.* +

Having proved the main theorem (3.7), we can immediately infer that

$$\mathbf{I}\Delta_0 + \Omega_1 \not\vdash \text{HCon}(\mathbf{I}\Delta_0 + \Omega_1).$$

As the proof of Theorem 3.7 is long, we will break it into a few lemmas. First we note that $\alpha \in \log^3$ if and only if there exists a sequence $\langle w_0, w_1, \dots, w_\alpha \rangle$ of length $(\alpha + 1)$ such that $w_0 = \exp^3(0) = 2^2$, and for any $j < \alpha$, $w_{j+1} = \omega_1(w_j)$. Noting that $\omega_1(\exp^3(j)) = \exp^3(j+1)$ one can then see that $w_\alpha = \exp^3(\alpha)$, and so $\alpha \in \log^3$.

This can be formalized in $\mathbf{I}\Delta_0 + \Omega_1$ by an arithmetical formula. Note that the code of the above sequence is bounded by $\mathcal{P}(\prod_{j=0}^{j=\alpha} w_j) \leq \mathcal{P}(\exp(\sum_{j=0}^{j=\alpha} \exp^2(j))) \leq \mathcal{P}(\exp^3(\alpha + 1)) \leq \mathcal{P}(\omega_1(\exp^3(\alpha)))$. So, in the presence of Ω_1 , the existence of $\exp^3(\alpha)$ guarantees the existence of the above sequence of w_j 's.

For proving Theorem 3.7 let us assume that we have a model

$$\mathcal{M} \models (\mathbf{I}\Delta_0 + \Omega_1) + (\alpha \in \log^2 \wedge \theta(\alpha)) + \text{HCon}(\mathbf{I}\Delta_0 + \Omega_1),$$

for some bounded formula $\theta(x)$ and some non-standard $\alpha \in \mathcal{M}$. The proof proceeds by constructing a model

$$\mathcal{N} \models (\mathbf{I}\Delta_0 + \Omega_1) + \exists x \in \log^3 \theta(x).$$

If our language of arithmetic \mathcal{L}_A contains the successor function \mathfrak{s} , then define the terms \underline{j} 's by induction: $\underline{0} = 0$, and $\underline{j+1} = \mathfrak{s}(\underline{j})$. If \mathcal{L}_A does not contain \mathfrak{s} , then it should have the constant 1, and in this case we can put $\underline{j+1} = \underline{j} + 1$. The term \underline{j} represents the (standard or non-standard) number j . For the sake of simplicity, assume \mathfrak{w} denotes the Skolem function symbol $\mathfrak{f}_{\exists y (y=\omega_1(x))}$. Put $w_0 = \underline{4}$ and inductively $w_{j+1} = \mathfrak{w}(w_j)$. Then w_k , in the theory $\mathbf{I}\Delta_0 + \Omega_1$, is the term which represents $\exp^3(k)$. Finally, put $\Lambda = \{\underline{0}, \dots, \underline{\omega_1(\alpha)}, w_0, \dots, w_\alpha\} = \{\underline{j} \mid j \leq \omega_1(\alpha)\} \cup \{w_j \mid j \leq \alpha\}$. We can now estimate an upper bound for the code of Λ : $\ulcorner \Lambda \urcorner \leq \mathcal{P}(\prod_{j=1}^{j=\omega_1(\alpha)} 2^j) \leq \mathcal{P}(\exp(\omega_1(\alpha)^2))$. So Λ has a code in \mathcal{M} (since $\mathcal{M} \models \alpha \in \log^2$), and moreover the value of $\omega_2(\ulcorner \Lambda \urcorner)$ exists in \mathcal{M} , because $\omega_2(\ulcorner \Lambda \urcorner) \leq \mathcal{P}(\omega_2(\exp(\omega_1(\alpha)^2))) \leq \mathcal{P}(\exp(\omega_1(\omega_1(\alpha)^2))) \leq \mathcal{P}(\exp^2(4(\log \alpha)^4)) \leq \mathcal{P}(\exp^2(\alpha))$. Thus by Theorem 3.4 there exists a non-standard j such that $\Lambda^{(j)}$ has a code in \mathcal{M} . Since by the assumption above we have $\mathcal{M} \models \text{HCon}(\mathbf{I}\Delta_0 + \Omega_1)$, then there exists an $(\mathbf{I}\Delta_0 + \Omega_1)$ -evaluation p on $\Lambda^{(j)}$ (in \mathcal{M}). Now, by what was said after the proof of Theorem 3.4 one can construct the model $\mathfrak{M}(\Lambda, p) = \mathcal{N}$. By Lemma 3.2 we have $\mathcal{N} \models (\mathbf{I}\Delta_0 + \Omega_1)$, and also $\mathcal{N} \models \underline{\alpha}/p \in \log^3$ follows from the existence of w_j/p 's. It remains (only) to show that $\mathfrak{M}(\Lambda, p) \models \theta(\underline{\alpha}/p)$.

For this purpose we prove the following lemmas where we assume that \mathcal{M} is as above and there are some non-standard set of terms and evaluation Λ, p in \mathcal{M} such that $\Lambda \supseteq \{\underline{0}, \dots, \underline{\omega_1(\alpha)}\}$ for a non-standard $\alpha \in \mathcal{M}$, and p is an $\mathbf{I}\Delta_0$ -evaluation on $\Lambda^{(\infty)}$ in \mathcal{M} .

LEMMA 3.9. *If $\mathfrak{M}(\Lambda, p) \models t/p \leq \underline{i}/p$ holds for a term t and $i \leq \omega_1(\alpha)$ in \mathcal{M} , then $\mathfrak{M}(\Lambda, p) \models t/p = \underline{j}/p$ for some $j \leq i$.*

PROOF. By the assumption, $\mathcal{M} \models "p \models t \leq \underline{i}"$. We prove by induction on i that there exists some $j \leq i$ in \mathcal{M} such that $\mathcal{M} \models "p \models t = \underline{j}"$.

- For $i = 0$ by Example 2.11 the assumption $\mathcal{M} \models "p \models t \leq 0"$ implies $\mathcal{M} \models "p \models t = 0"$, noting that p is an Q-evaluation on $\Lambda^{(\infty)}$, and thus all the needed Skolem terms are in p 's disposal.

- For $i + 1$ we have $\mathcal{M} \models "p \models t \leq \underline{i} \vee t = \mathfrak{s}(\underline{i})"$ by Example 2.11 and the assumption $\mathcal{M} \models "p \models t \leq \mathfrak{s}(\underline{i})"$. Then if $\mathcal{M} \models "p \models t = \mathfrak{s}(\underline{i})"$ we are done, and if $\mathcal{M} \models "p \models t \leq \underline{i}"$ by the induction hypothesis there must exist some $j \leq i$ in \mathcal{M} such that $\mathcal{M} \models "p \models t = \underline{j}"$. \dashv

REMARK 3.10. *The proof of the above lemma does not depend on the axioms of \mathbf{Q} (and $\mathbf{I}\Delta_0$). Indeed, in some axiomatization of \mathbf{Q} in the literature, the sentences $\psi = \forall x (x \leq 0 \rightarrow x = 0)$ and $\varphi = \forall x \forall y (x \leq \mathfrak{s}y \rightarrow x = \mathfrak{s}y \vee x \leq y)$ (see Example 2.11) are accepted as axioms. In our axiomatization, the above sentences were derivable theorems. In some axiomatizations of \mathbf{Q} our axiom A_4 is replaced with $A'_4: \forall x, y (x \leq y \leftrightarrow \exists z [z + x = y])$; note the difference of $x + z$ in A_4 and $z + x$ in A'_4 (see e.g., [7]). In this new axiomatization the sentence φ is not derivable. However, since we have $\mathbf{I}\Delta_0 \vdash \psi \wedge \varphi$, then by the argument of Remark 2.13, the above Lemma 3.9 can be proved by using the fact that p is an $\mathbf{I}\Delta_0$ -evaluation on $\Lambda^{(\infty)}$. \dashv*

LEMMA 3.11. *If $t(x_1, \dots, x_m)$ is an \mathcal{L}_A -term and $i_1, \dots, i_m \leq \omega_1(\alpha)$ are elements of \mathcal{M} such that $\mathcal{M} \models x \leq t(i_1, \dots, i_m)$ for some x , then for some \mathcal{L}_A -term $t'(x_1, \dots, x_k)$ and some $j_1, \dots, j_k \leq \omega_1(\alpha)$ we have $\mathcal{M} \models x = t'(j_1, \dots, j_k)$.*

PROOF. By induction on (the complexity of) the term t .

- For $t = 0$ and $t = x_1$ the proof is straightforward.
- For $t = \mathfrak{s}u$ the assumption $\mathcal{M} \models x \leq \mathfrak{s}u(i_1, \dots, i_m)$ implies that we have either $\mathcal{M} \models x = \mathfrak{s}u(i_1, \dots, i_m)$ or $\mathcal{M} \models x \leq u(i_1, \dots, i_m)$ is true, and then the conclusion follows from the induction hypothesis.
- For $t = u + v$, and the assumption $\mathcal{M} \models x \leq u(i_1, \dots, i_m) + v(i_1, \dots, i_m)$, we consider two cases. First if $\mathcal{M} \models x \leq u(i_1, \dots, i_m)$ then we are done by the induction hypothesis. Second if $\mathcal{M} \models u(i_1, \dots, i_m) \leq x$ then there exists a y such that $\mathcal{M} \models x = u(i_1, \dots, i_m) + y$ and moreover $\mathcal{M} \models y \leq v(i_1, \dots, i_m)$. Now, by the induction hypothesis there are a term $t'(x_1, \dots, x_k)$ and some elements $j_1, \dots, j_k \leq \omega_1(\alpha)$ such that $\mathcal{M} \models y = t'(j_1, \dots, j_k)$. Whence we finally get the conclusion $\mathcal{M} \models x = u(i_1, \dots, i_m) + t'(j_1, \dots, j_k)$.
- For $t = u \cdot v$, by an argument similar to that of the previous case, we can assume $\mathcal{M} \models u(i_1, \dots, i_m) \leq x \leq u(i_1, \dots, i_m) \cdot v(i_1, \dots, i_m)$. There are some q, r such that $\mathcal{M} \models x = u(i_1, \dots, i_m) \cdot q + r$ and $\mathcal{M} \models r \leq u(i_1, \dots, i_m)$. We also have $\mathcal{M} \models q \leq v(i_1, \dots, i_m)$. By the induction hypothesis there are terms t', t'' and $j_1, \dots, j_k \leq \omega_1(\alpha)$ such that $\mathcal{M} \models q = t'(j_1, \dots, j_k) \wedge r = t''(j_1, \dots, j_k)$. Thus we finally have $\mathcal{M} \models x = u(i_1, \dots, i_m) \cdot t'(j_1, \dots, j_k) + t''(j_1, \dots, j_k)$. \dashv

LEMMA 3.12. *For $i, j, k \leq \omega_1(\alpha)$ in \mathcal{M} we have*

- (1) *if $i \leq j \leq \omega_1(\alpha)$ then $\mathfrak{M}(\Lambda, p) \models \underline{i}/p \leq \underline{j}/p$;*
- (2) *if $i + j \leq \omega_1(\alpha)$ then $\mathfrak{M}(\Lambda, p) \models \underline{i}/p + \underline{j}/p = \underline{i + j}/p$;*
- (3) *if $i \cdot j \leq \omega_1(\alpha)$ then $\mathfrak{M}(\Lambda, p) \models \underline{i}/p \cdot \underline{j}/p = \underline{i \cdot j}/p$.*

PROOF. We need to show for the $i, j \leq \omega_1(\alpha)$ that

- (1) if $\mathcal{M} \models i \leq j$ then $\mathcal{M} \models "p \models \underline{i} \leq \underline{j}"$,
- (2) if $\mathcal{M} \models i + j \leq \omega_1(\alpha)$ then $\mathcal{M} \models "p \models \underline{i} + \underline{j} = \underline{i + j}"$, and
- (3) if $\mathcal{M} \models i \cdot j \leq \omega_1(\alpha)$ then $\mathcal{M} \models "p \models \underline{i} \cdot \underline{j} = \underline{i \cdot j}"$.

First we note that the statement (2) above implies already (1), since if we have $\mathcal{M} \models i \leq j$, then for some k we should have $\mathcal{M} \models i + k = j$, and then by (2), $\mathcal{M} \models "p \models \underline{i} + \underline{k} = \underline{j}"$ which implies (by A_4 of \mathbf{Q} —see Example 2.11) that $\mathcal{M} \models "p \models \underline{i} \leq \underline{j}"$. By induction on j , very similarly to the proof of Lemma 3.9, one can prove the statements (2) and (3), noting that the evaluation p must satisfy

the following axioms of Q:

$$\begin{aligned} A_5: \quad \forall x (x + 0 = x); & \quad A_6: \forall x \forall y (x + \mathfrak{s}y = \mathfrak{s}(x + y)); \\ A_7: \quad \forall x (x \cdot 0 = 0); & \quad A_8: \forall x \forall y (x \cdot \mathfrak{s}y = x \cdot y + x). \end{aligned} \quad \dashv$$

COROLLARY 3.13. *Suppose for an \mathcal{L}_A -term $t(x_1, \dots, x_m)$ and some elements $i_1, \dots, i_m, i \leq \omega_1(\alpha)$, we have $\mathcal{M} \models t(i_1, \dots, i_m) = i$. Then we must also have $\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) = \underline{i}/p$.*

PROOF. By induction on t using Lemma 3.12. \dashv

LEMMA 3.14. *Suppose $i_1, \dots, i_m \leq \alpha^k$ are elements of \mathcal{M} for some standard number $k \in \mathbb{N}$, and $t(x_1, \dots, x_m), t'(x_1, \dots, x_m)$ are \mathcal{L}_A -terms. Then, if $\mathcal{M} \models t(i_1, \dots, i_m) = t'(i_1, \dots, i_m)$ holds, $\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) = t'(\underline{i}_1/p, \dots, \underline{i}_m/p)$ holds too.*

PROOF. By $i_1, \dots, i_m \leq \alpha^k$ we have $t(i_1, \dots, i_m) \leq \omega_1(\alpha)$. Put i be the common value $i = t(i_1, \dots, i_m) = t'(i_1, \dots, i_m)$. Then By Corollary 3.13 we must have that $\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) = \underline{i}/p = t'(\underline{i}_1/p, \dots, \underline{i}_m/p)$. \dashv

LEMMA 3.15. *Suppose $i_1, \dots, i_m \leq \alpha^k$ are elements of \mathcal{M} for some standard number $k \in \mathbb{N}$, and $t(x_1, \dots, x_m), t'(x_1, \dots, x_m)$ are two \mathcal{L}_A -terms. If we have $\mathcal{M} \models t(i_1, \dots, i_m) \leq t'(i_1, \dots, i_m)$ then we must also have the following satisfaction: $\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) \leq t'(\underline{i}_1/p, \dots, \underline{i}_m/p)$.*

PROOF. Noting that $\mathbf{Q} \vdash \forall x, y (x \leq y \leftrightarrow \exists z (x + z = y))$ by the assumption there exists some $\beta \in \mathcal{M}$ such that $\mathcal{M} \models t(i_1, \dots, i_m) + \beta = t'(i_1, \dots, i_m)$. On the other hand $\mathcal{M} \models \beta \leq t'(i_1, \dots, i_m)$, so by Lemma 3.11 there exist some term u and some elements $j_1, \dots, j_k \leq \omega_1(\alpha)$ such that $\mathcal{M} \models \beta = u(j_1, \dots, j_k)$. Thus $\mathcal{M} \models t(i_1, \dots, i_m) + u(j_1, \dots, j_k) = t'(i_1, \dots, i_m)$. Now, by Lemma 3.14,

$$\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) + u(\underline{j}_1/p, \dots, \underline{j}_k/p) = t'(\underline{i}_1/p, \dots, \underline{i}_m/p),$$

whence $\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) \leq t'(\underline{i}_1/p, \dots, \underline{i}_m/p)$. \dashv

LEMMA 3.16. *Suppose $i_1, \dots, i_m \leq \alpha^k$ are elements of \mathcal{M} for some standard number $k \in \mathbb{N}$, and $t(x_1, \dots, x_m), t'(x_1, \dots, x_m)$ are two \mathcal{L}_A -terms. If it is true that $\mathcal{M} \models t(i_1, \dots, i_m) \neq t'(i_1, \dots, i_m)$ then $\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) \neq t'(\underline{i}_1/p, \dots, \underline{i}_m/p)$ must be true too. And if $\mathcal{M} \models t(i_1, \dots, i_m) \not\leq t'(i_1, \dots, i_m)$ then $\mathfrak{M}(\Lambda, p) \models t(\underline{i}_1/p, \dots, \underline{i}_m/p) \not\leq t'(\underline{i}_1/p, \dots, \underline{i}_m/p)$.*

PROOF. It follows from Lemma 3.15 (and Remark 2.13) noting that p is an $\mathbf{I}\Delta_0$ -evaluation on $\Lambda^{(\infty)}$ and $\mathbf{I}\Delta_0 \vdash \forall x, y (x \neq y \leftrightarrow \mathfrak{s}y \leq x \vee \mathfrak{s}x \leq y)$, also $\mathbf{I}\Delta_0 \vdash \forall x, y (x \not\leq y \leftrightarrow \mathfrak{s}y \leq x)$. \dashv

THEOREM 3.17. *Suppose $\psi(x_1, \dots, x_m)$ is an open RNNF \mathcal{L}_A -formula and $i_1, \dots, i_m \leq \alpha^k$ are elements of \mathcal{M} for some standard number $k \in \mathbb{N}$. If we have $\mathcal{M} \models \psi(i_1, \dots, i_m)$ then we also have $\mathfrak{M}(\Lambda, p) \models \psi(\underline{i}_1/p, \dots, \underline{i}_m/p)$.*

PROOF. Lemmas 3.14 and 3.15 prove the theorem for atomic formulas, and Lemma 3.16 proves it for negated atomic formulas. The disjunctive and conjunctive compositions of those formulas one can be proved by a simple induction. \dashv

THEOREM 3.18. *Suppose that $\varphi(x_1, \dots, x_m)$ is a bounded \mathcal{L}_A -formula and that $i_1, \dots, i_m \leq \alpha^k$ are elements of \mathcal{M} for some standard number $k \in \mathbb{N}$. If $\mathcal{M} \models \varphi(i_1, \dots, i_m)$ then $\mathfrak{M}(\Lambda, p) \models \varphi(\underline{i}_1/p, \dots, \underline{i}_m/p)$.*

PROOF. Every bounded formula can be written as an (equivalent) RNNF formula. By Lemma 3.11 the range of bounded quantifiers of a formula whose all parameters belong to the set

$$\{t(i_1, \dots, i_m) \mid i_1, \dots, i_m \leq \alpha \ \& \ t \text{ is an } \mathcal{L}_A\text{-term}\}$$

is indeed that set again. Now the conclusion follows from Theorem 3.17. \dashv

AN ALTERNATIVE PROOF. To make this important theorem more clear, we sketch another proof, which is not really too different but has more model-theoretic flavor. Consider the above set again

$$\langle [0, \alpha] \rangle_{\mathcal{M}} = \{t(i_1, \dots, i_m) \mid i_1, \dots, i_m \leq \alpha \ \& \ t \text{ is an } \mathcal{L}_A\text{-term}\}$$

which is a subset of \mathcal{M} closed under the successor, addition, and multiplication, and thus forms a submodel of \mathcal{M} (generated by $[0, \alpha] = \{x \in \mathcal{M} \mid x \leq \alpha\}$). This submodel is an initial segment of \mathcal{M} by Lemma 3.11. Hence, whenever $\mathcal{M} \models \varphi$, for a bounded formula φ with parameters in $[0, \alpha]$, then $\langle [0, \alpha] \rangle_{\mathcal{M}} \models \varphi$.

Now, similarly, the set

$$\langle [\underline{0}/p, \underline{\alpha}/p] \rangle_{\mathcal{N}} = \{t(\underline{i}_1/p, \dots, \underline{i}_m/p) \mid i_1, \dots, i_m \leq \alpha \ \& \ t \text{ is an } \mathcal{L}_A\text{-term}\}$$

is an initial segment and a submodel of $\mathcal{N} = \mathfrak{M}(\Lambda, p)$. Thus if $\langle [\underline{0}/p, \underline{\alpha}/p] \rangle_{\mathcal{N}} \models \varphi$, where φ is a bounded formula with parameters in $[\underline{0}/p, \underline{\alpha}/p]$, then $\mathfrak{M}(\Lambda, p) \models \varphi$. Finally, we note that the mapping $t(i_1, \dots, i_m) \mapsto t(\underline{i}_1/p, \dots, \underline{i}_m/p)$ defines a bijection between $\langle [0, \alpha] \rangle_{\mathcal{M}}$ and $\langle [\underline{0}/p, \underline{\alpha}/p] \rangle_{\mathcal{N}}$ which is also an isomorphism by Lemmas 3.14, 3.15 and 3.16. So the proof of the theorem goes as follows:

If $\mathcal{M} \models \varphi(i_1, \dots, i_m)$ then $\langle [0, \alpha] \rangle_{\mathcal{M}} \models \varphi(i_1, \dots, i_m)$, so we should have that $\langle [\underline{0}/p, \underline{\alpha}/p] \rangle_{\mathcal{N}} \models \varphi(\underline{i}_1/p, \dots, \underline{i}_m/p)$ hence $\mathfrak{M}(\Lambda, p) \models \varphi(\underline{i}_1/p, \dots, \underline{i}_m/p)$. \dashv

COROLLARY 3.19. *By the above assumptions, $\mathfrak{M}(\Lambda, p) \models \theta(\underline{\alpha}/p)$.* \dashv

Let us summarize what was argued in the last few pages.

PROOF. (OF THEOREM 3.7.) By the assumption of the theorem, the theory $(\mathbf{I}\Delta_0 + \Omega_1) + \exists x \in \log^2 \theta(x) + \mathbf{HCon}(\mathbf{I}\Delta_0 + \Omega_1)$ is consistent. So there is a model

$$\mathcal{M} \models (\mathbf{I}\Delta_0 + \Omega_1) + (\alpha \in \log^2 \wedge \theta(\alpha)) + \mathbf{HCon}(\mathbf{I}\Delta_0 + \Omega_1),$$

where $\alpha \in \mathcal{M}$. We wish to show the consistency of $(\mathbf{I}\Delta_0 + \Omega_1) + \exists x \in \log^3 \theta(x)$ by constructing a model

$$\mathcal{N} \models (\mathbf{I}\Delta_0 + \Omega_1) + \exists x \in \log^3 \theta(x).$$

If α is standard (i.e., $\alpha \in \mathbb{N}$) then one can take $\mathcal{N} = \mathcal{M}$. But if $\alpha \in \mathcal{M}$ is non-standard, then we proceed as follows: Take Λ to be the following set of terms: $\Lambda = \{\underline{j} \mid j \leq \omega_1(\alpha)\} \cup \{w_j \mid j \leq \alpha\}$ in which the terms \underline{j} 's and w_j 's are defined inductively as $\underline{0} = 0$, $\underline{j+1} = \mathfrak{s}\underline{j}$; and $w_0 = \underline{4}$, $w_{j+1} = \mathfrak{w}(w_j)$. Here \mathfrak{s} is the successor function, and \mathfrak{w} denotes the Skolem function symbol $f_{\exists y} (y = \omega_1(x))$. Now $\omega_2(\ulcorner \Lambda \urcorner)$ is of order (far less than) 2^{2^α} which exists by the assumption $\mathcal{M} \models \alpha \in \log^2$. Then by Theorem 3.4, for a non-standard j the set of terms $\Lambda^{(j)}$ has a code in \mathcal{M} . Thus the assumption $\mathcal{M} \models \mathbf{HCon}(\mathbf{I}\Delta_0 + \Omega_1)$ implies that there must exist an $(\mathbf{I}\Delta_0 + \Omega_1)$ -evaluation p on $\Lambda^{(j)}$. Then one can form the model $\mathcal{N} = \mathfrak{M}(\Lambda, p)$. Now $\mathcal{N} \models \mathbf{I}\Delta_0 + \Omega_1$ by Lemma 3.2, and also $\mathcal{N} \models \underline{\alpha}/p \in \log^3$ by the definition of w_α . Finally, $\mathcal{N} \models \theta(\underline{\alpha}/p)$ by Corollary 3.19 (of Theorem 3.18). Whence \mathcal{N} is a model of $(\mathbf{I}\Delta_0 + \Omega_1) + \exists x \in \log^3 \theta(x)$; and this proves its consistency. \dashv

§4. Herbrand consistency of $\text{I}\Delta_0$. Our definition of Herbrand consistency is not best suited for $\text{I}\Delta_0$: there are $\omega_1(\ulcorner \Lambda \urcorner)$ -many evaluations on a given set of terms Λ . Though this may not seem a big problem in the first glance (one can change or modify the definition accordingly) but special care is needed for generalizing the results to the case of $\text{I}\Delta_0$. In the first subsection we pinpoint the critical usages of Ω_1 and in the second subsection we tailor the definitions and theorems in a way that we can prove our main theorem for $\text{I}\Delta_0$ finally.

4.1. Essentiality of Ω_1 . We made an essential use of Ω_1 in the following parts of our arguments:

1. The totality of the ω_1 function was needed for the upper bound of the code of an evaluation on a given set of terms Λ . Namely, the code of any evaluation on Λ is of order $\omega_1(\ulcorner \Lambda \urcorner)$, see Lemma 2.16. And indeed there is no escape from this bound since, as it was explained after Lemma 2.16, there are $\exp(2|\Lambda|^2)$ evaluations on Λ , and if $|\Lambda| \approx \log \ulcorner \Lambda \urcorner$ then there could exist as many as $\omega_1(\ulcorner \Lambda \urcorner)^2$ evaluations on Λ . So, if Ω_1 is not available, then there could be a large and non-standard set of terms Γ in a model \mathcal{M} such that \mathcal{M} cannot see all the evaluations on Γ . One of those evaluations could be a T -evaluation, that an end-extension of \mathcal{M} , say \mathcal{N} , can see. Then Γ is a Herbrand proof of contradiction in \mathcal{M} because in \mathcal{M} 's view there is no T -evaluation on Γ . But there could be indeed a very large T -evaluation on Γ which \mathcal{M} could not see, but \mathcal{N} can. Thus the definition of HCon is deficient for $\text{I}\Delta_0$ (where Ω_1 is not there) and one cannot consider all the sets of terms; those for which the ω_1 of their codes exist, should be considered instead.

2. The second critical use of Ω_1 was in the definition of w_j 's for shrinking the (double-)logarithmic witness $\mathcal{M} \models \alpha \in \log^2$ to $\mathcal{N} \models w_\alpha/p \in \log^3$. There we constructed the sequence $\langle w_0, \dots, w_\alpha \rangle$ of terms such that $w_0 = \underline{4}$ and $w_{j+1} = \mathfrak{w}(w_j)$ where \mathfrak{w} is the Skolem function symbol $f_{\exists y [y = \omega_1(x)]}$. And this was in our disposal because $\Omega_1 = \forall x \exists y [y = \omega_1(x)]$ was one of the axioms (of $\text{I}\Delta_0 + \Omega_1$) and thus every $(\text{I}\Delta_0 + \Omega_1)$ -evaluation must have satisfied $\mathfrak{w}(t) = \omega_1(t)$.

Note that we also required Λ to contain $\{\underline{j} \mid j \leq \omega_1(\alpha)\}$, but for this we did not need the existence of $\omega_1(\alpha)$; it was guaranteed by the assumption $\mathcal{M} \models \alpha \in \log^2$.

4.2. Tailoring for $\text{I}\Delta_0$. Here we introduce the necessary modifications on the above two points.

4.2.1. The definition of HCon^* . The first point can be dealt with by tailoring the definition of HCon for $\text{I}\Delta_0$:

DEFINITION 4.1. A theory T is called Herbrand Consistent*, denoted symbolically as $\text{HCon}^*(T)$, when for all set of terms Λ , if $\omega_1(\ulcorner \Lambda \urcorner)$ exists then there is an T -evaluation on Λ . –

This, obviously, can again be formalized in the language of arithmetic. The new definition cannot harm our arguments too much, because we needed HCon only for some special set of terms. And it was $\Lambda^{(j)}$ for some non-standard j where $\Lambda = \{\underline{j} \mid j \leq \omega_1(\alpha)\} \cup \{w_j \mid j \leq \alpha\}$. For constructing the model $\mathfrak{M}(\Lambda, p)$ we already needed the existence of $\omega_2(\ulcorner \Lambda \urcorner)$ (see the beginning of the proof of Theorem 3.7 before Lemma 3.9). Now, by the above definition, we need the existence of $\omega_1(\ulcorner \Lambda^{(j)} \urcorner)$ as well; and so the existence of $\omega_1(\omega_2(\ulcorner \Lambda \urcorner))$ for the above special set of terms Λ will make our arguments to go through.

4.2.2. *The initial segments \mathcal{I} and \mathcal{J} .* In the absence of Ω_1 we cannot define the above sequence $\langle w_0, \dots, w_\alpha \rangle$ satisfying $w_{j+1} = \omega_1(w_j)$. The most we can do inside $\mathbf{I}\Delta_0$ is to define a sequence like $\langle v_0, \dots, v_\beta \rangle$ where $v_0 = m$ and $v_{j+1} = (v_j)^n$ for some fixed $m, n \in \mathbb{N}$. Then $v_\beta = a^{n^{2^\beta}} \leq \mathcal{P}(\exp^2(\beta))$. Thus we cannot get anything larger than \exp^2 , and so for shortening a witness we should start from \log and remain in the realm of \log^2 . Indeed by the arguments of the beginning of the proof of Theorem 3.7 before Lemma 3.9 we did not need the existence of $\exp^2(\alpha)$ for the existence of $\omega_2(\ulcorner \Lambda \urcorner)$. We needed only $\exp^2(4(\log \alpha)^4)$, and so for the existence of $\omega_1(\omega_2(\ulcorner \Lambda \urcorner))$ we need $\omega_1(\exp^2(4(\log \alpha)^4)) = \exp^2(8(\log \alpha)^4)$. Thus it seems natural to consider the initial segment $\mathcal{I} = \{x \mid \exists y [y = \exp^2(8(\log x)^4)]\}$ and its logarithm $\mathcal{J} = \{x \mid \exists y [y = \exp^2(8x^4)]\}$. We first note that Adamowicz's theorem (Theorem 3.6) holds for $\mathbf{I}\Delta_0$ and any $n \in \mathbb{N}$; i.e., there exists a bounded formula whose \log^n -witness cannot consistently be shortened to \log^{n+1} . Indeed this theorem holds for any initial segment I and its logarithm which is, by definition, the initial segment $J = \{x \mid \exists y [y = \exp(x) \wedge y \in I]\}$. The only relation between \log^n and \log^{n+1} needed in the proof of Theorem 3.5 is that $2^x \in \log^n \iff x \in \log^{n+1}$; see [2]. As a matter of fact, the proof works for any initial segments I and J which satisfy $\forall x (2^x \in I \iff x \in J)$. The initial segments \mathcal{I} and \mathcal{J} defined above satisfy this as well ($\exp(x) \in \mathcal{I} \iff x \in \mathcal{J}$). So, we repeat Theorem 3.5 as (see also Theorem 3.6 of [15]):

THEOREM 4.2. [2] *There exists some arithmetical bounded formula $\eta(\bar{x})$ such that the theory $\mathbf{I}\Delta_0 + \exists \bar{x} \in \mathcal{I} \eta(\bar{x})$ is consistent, but $\mathbf{I}\Delta_0 + \exists \bar{x} \in \mathcal{J} \eta(\bar{x})$ is not.* \dashv

4.2.3. *The main theorem for $\mathbf{I}\Delta_0$.* Let us note that the following theorem together with Theorem 4.2 prove that $\mathbf{I}\Delta_0 \not\vdash \text{HCon}^*(\mathbf{I}\Delta_0)$. Thus we also have $\mathbf{I}\Delta_0 \not\vdash \text{HCon}(\mathbf{I}\Delta_0)$.

THEOREM 4.3. *For any Δ_0 -formula $\theta(x)$, if $\mathbf{I}\Delta_0 + \exists x \in \mathcal{I} \theta(x) + \text{HCon}^*(\mathbf{I}\Delta_0)$ is consistent then so is the theory $\mathbf{I}\Delta_0 + \exists x \in \mathcal{J} \theta(x)$.*

PROOF. Suppose the theory $\mathbf{I}\Delta_0 + \exists x \in \mathcal{I} \theta(x) + \text{HCon}^*(\mathbf{I}\Delta_0)$ is consistent. So there exists a model

$$\mathcal{M} \models \mathbf{I}\Delta_0 + (\alpha \in \mathcal{I} \wedge \theta(\alpha)) + \text{HCon}^*(\mathbf{I}\Delta_0),$$

where $\alpha \in \mathcal{M}$. We will show the consistency of $\mathbf{I}\Delta_0 + \exists x \in \mathcal{J} \theta(x)$ by constructing a model

$$\mathcal{N} \models \mathbf{I}\Delta_0 + \exists x \in \mathcal{J} \theta(x).$$

If α is standard (i.e., $\alpha \in \mathbb{N}$) then one can take $\mathcal{N} = \mathcal{M}$. But if $\alpha \in \mathcal{M}$ is non-standard, then we proceed as follows:

Let $\Upsilon = \{0, 0 + 0, 0^2, c, c^2, c^2 + 0, sc, qc, (sc)^2, (sc)^2 + 0\}$ where q is the Skolem function symbol for the formula $\exists y (y \leq x^2 \wedge y = x^2)$ and c is the Skolem constant symbol for the sentence $\exists x (\exists w (w \leq x^2 \wedge w = x^2) \wedge \forall v (v \not\leq (sx)^2 \wedge v \neq (sx)^2))$; see Example 2.12. We can use the argument of Example 2.12, since for the bounded formula $\psi(x) = \exists y \leq x^2 (y = x \cdot x)$, the sentence ind_ψ is an axiom of the theory $\mathbf{I}\Delta_0$. Take $\Lambda = \Upsilon \cup \{\underline{j} \mid j \leq \omega_1(\alpha)\} \cup \{z_j \mid j \leq 8\alpha^4\}$ in which the terms \underline{j} 's and z_j 's are defined inductively as $\underline{0} = 0$, $\underline{j+1} = sj$; and $z_0 = \underline{2}$, $z_{j+1} = q(z_j)$. Now $\omega_2(\ulcorner \Lambda \urcorner)$ is of order $\exp^2(4(\log \alpha)^4)$ which exists by the assumption $\mathcal{M} \models \alpha \in \mathcal{I}$. Then by Theorem 3.4 for a non-standard j the set of terms $\Lambda^{(j)}$ has a code in \mathcal{M} .

Moreover, the value $\omega_1(\ulcorner \Lambda^{(j)} \urcorner)$ exists in \mathcal{M} because $\omega_1(\ulcorner \Lambda^{(j)} \urcorner) \leq \exp^2(8(\log \alpha)^4)$ and $\mathcal{M} \models \alpha \in \mathcal{F}$. Thus the assumption $\mathcal{M} \models \text{HCon}^*(\text{I}\Delta_0)$ implies that there must exist an $\text{I}\Delta_0$ -evaluation p on $\Lambda^{(j)}$. Then one can form the model $\mathcal{N} = \mathfrak{M}(\Lambda, p)$. Now $\mathcal{N} \models \text{I}\Delta_0$ by Lemma 3.2, and also $\mathcal{N} \models \underline{\alpha}/p \in \mathcal{F}$ by the definition of $z_{8\alpha^4}$ (which represents $\exp^2(8\alpha^4)$). Note that $p \models z_{j+1} = z_j \cdot z_j$ by the argument of Example 2.12, and also the code of the sequence $\langle z_0, \dots, z_{8\alpha^4} \rangle$ is of order $\exp((8\alpha^4)^2) \leq \exp^2(8(\log \alpha)^4)$ which exists since $\alpha \in \mathcal{F}$. Finally, $\mathcal{N} \models \theta(\underline{\alpha}/p)$ by Corollary 3.19 (of Theorem 3.18). Whence \mathcal{N} is a model of the theory $\text{I}\Delta_0 + \exists x \in \mathcal{F} \theta(x)$; what proves its consistency. \dashv

§5. Conclusions. (A) An important property of Herbrand consistency of the theories $\text{I}\Delta_0 + \Omega_1$ and $\text{I}\Delta_0$ has been proved. That property immediately implies Gödel's second incompleteness theorem for the notion of Herbrand consistency in those theories. This version of Gödel's theorem has come a long way. The original presumption of Paris & Wilkie [10] asked for a proof of $\text{I}\Delta_0 \not\vdash \text{CFCon}(\text{I}\Delta_0)$, without specifying any variant of Cut-Free Consistency CFCon : "Presumably $\text{I}\Delta_0 \not\vdash \text{CFCon}(\text{I}\Delta_0)$ although we do not know this at present". Willard [16] solved this problem for the Tableau Consistency variant. Pudlák [11] asked a closely related question: "we know only that $T \not\vdash \text{HCon}(T)$ for T containing at least $\text{I}\Delta_0 + \text{Exp}$, for weaker theories it is an open problem". In [13] this problem was studied for the theories $\text{I}\Delta_0 + \Omega_1$ and $\text{I}\Delta_0$ (and a theory in between these two, namely $\text{I}\Delta_0$ plus the totality of the $x \mapsto x^{\log^2 x}$ function). The proof of $\text{I}\Delta_0 + \Omega_1 \not\vdash \text{HCon}(\text{I}\Delta_0 + \Omega_1)$ given here was presented for the first time in Chapter 5 of [13]. But the unprovability of $\text{HCon}(\text{I}\Delta_0)$ in $\text{I}\Delta_0$ was not as easy as it would have seemed. In Chapter 3 of [13] this unprovability was proved for a re-axiomatization of $\text{I}\Delta_0$. It was also showed in [16] for a simpler re-axiomatization of $\text{I}\Delta_0$; see page 141 of [17]. The proof of [16] is also reconsidered in [17] by showing the unprovability of $\text{HCon}(\text{I}\Delta_0 + \Omega_0)$ in $\text{I}\Delta_0$, where Ω_0 is the axiom $\forall x \exists y [y = x \cdot x]$ (see point **D**(4) below). The paper [17] finishes with "What we are trying to suggest in this appendix is that while a proof of the Herbrandized version of the $\text{I}\Sigma_0$ incompleteness theorem will be certainly quite interesting and very different from our semantic tableaux analysis, we would anticipate that there are some types of proofs of the Herbrandized version of this incompleteness theorem that partially resemble the analysis techniques used in Sections 3 and 4 for semantic tableaux deduction." But, unfortunately, it was never clear how to avoid using the Ω_0 axiom in the proof of $\text{I}\Delta_0 \not\vdash \text{HCon}(\text{I}\Delta_0 + \Omega_0)$ in [16] (and [17]).

(B) Our reason for using the induction formula ind_ψ , where $\psi(x)$ is the bounded formula $\exists y \leq x^2 (y = x^2)$, was having a Skolem function symbol for squaring $q(x) = x^2$. This way the Gödel code of $q(x)$ is $M \cdot \ulcorner x \urcorner$ for a fixed $M \in \mathbb{N}$, and thus the code of $q^n(x)$ is $M^n \cdot \ulcorner x \urcorner$ which is of order $\exp(n)$. So, we could code a term representing the number $x^{\exp(n)}$ ($=q^n(x)$) by a number of order $\exp(n)$. But if we coded the number $x^{\exp(n)}$ directly, that would be the code of $x \cdot x \cdot \dots \cdot x$ (with $2^n - \text{times } x$) which is of order $(\ulcorner x \urcorner)^{2^n}$ or $\exp^2(n)$. In that case, the code of the sequence $\langle z_0, \dots, z_{8\alpha^4} \rangle$ would be of order $\exp^2((8\alpha^4)^2)$, but we used the order $\exp((8\alpha^4)^2)$ in the proof of Theorem 4.3 (since we had at most $\exp^2(8(\log \alpha)^4)$ in our disposal—which is far less than $\exp^2((8\alpha^4)^2)$). Whence, we avoided accepting the totality of the squaring function $\Omega_0: \forall x \exists y (y = x \cdot x)$ as an (additional) axiom.

This point deserves another look: define the terms $\{z_i\}$, $\{u_i\}$, and $\{v_i\}$ inductively as $z_0 = 2$, $z_{j+1} = q(z_j)$; $u_0 = 2$, $u_{j+1} = (z_j)^2$; and $v_0 = 2$, $v_{j+1} = (v_j)^2$. Then the codes of the terms z_n 's and u_n 's are of order $\mathcal{P}(2^n)$, but the code of v_n 's are of order $\mathcal{P}(2^{2^n})$. On the other hand, the terms z_i , u_i and v_i have the same value (2^{2^i}) in any model of $\mathbf{I}\Delta_0$. In fact, for $i \leq \omega_1(\alpha)$ we have $z_i \in \Lambda$ and also $u_i \in \Lambda^{(1)}$; but v_i 's are too big to fit in small sets of terms.

(C) We can now mention a seemingly contrasting result: it was seen in the Introduction that $\mathbf{I}\Delta_0 + \text{Exp} \vdash \text{HCon}(\mathbf{I}\Delta_0)$ (also $\mathbf{I}\Delta_0 + \text{Exp} \vdash \text{HCon}(\mathbf{I}\Delta_0 + \Omega_m)$ for any $m \in \mathbb{N}$ —see Theorem 3.2 of [15]; on the other hand we have conjectured in [15] that $\mathbf{I}\Delta_0 + \text{Exp} \not\vdash \text{HCon}(\bigcup_n \{\mathbf{I}\Delta_0 + \Omega_n\})$. So, for some $k \in \mathbb{N}$, where k is the number of times that Exp has been used in the proof of the theorem $\mathbf{I}\Delta_0 + \text{Exp} \vdash \text{HCon}(\mathbf{I}\Delta_0 + \Omega_1)$, we should have

$$\mathbf{I}\Delta_0 + \Omega_1 \vdash \text{HCon}^{\log^k}(\mathbf{I}\Delta_0 + \Omega_1) \text{ and } \mathbf{I}\Delta_0 \vdash \text{HCon}^{\log^k}(\mathbf{I}\Delta_0),$$

where HCon^{\log^k} is the relativized of HCon to the initial segment \log^k (see also Theorem 5.26 of [7]). This may seem in contrast with our results

$$\mathbf{I}\Delta_0 + \Omega_1 \not\vdash \text{HCon}(\mathbf{I}\Delta_0 + \Omega_1) \text{ and } \mathbf{I}\Delta_0 \not\vdash \text{HCon}(\mathbf{I}\Delta_0).$$

To see why our proof (and in fact any other proof) cannot work for showing (the false statement) $\mathbf{I}\Delta_0 + \Omega_1 \not\vdash \text{HCon}^{\log^k}(\mathbf{I}\Delta_0 + \Omega_1)$, we note that for proving the consistency of the theory $(\mathbf{I}\Delta_0 + \Omega_1) + (\exists x \in \log \mathcal{J} \theta(x))$ by assuming the consistency of the theory $(\mathbf{I}\Delta_0 + \Omega_1) + (\alpha \in \mathcal{J} \wedge \theta(\alpha)) + \text{HCon}^{\log^k}(\mathbf{I}\Delta_0 + \Omega_1)$, for some suitable initial segment \mathcal{J} , the most we can do is to consider a set of terms Λ which contains $\{w_i \mid i \leq \beta\}$, for an appropriate β , where each w_{j+1} is (roughly of order) $\omega_1^n(w_j)$ (for a fixed n); cf. subsection 4.2.2 on “*The Initial Segments \mathcal{I} and \mathcal{J}* ”. Thus the most we can get is (roughly) $w_\beta \approx \omega_1^{n\beta}(w_0) \approx \exp^3(\beta)$, so we should have $\beta \in \log^3$; and $\log \mathcal{J}$ must be (roughly) of order \log^3 , and \mathcal{J} must be (roughly) of order \log^2 (see Theorem 3.7 above). Thus α must be (roughly) in \log^2 and so β must be (roughly) in \log^3 (or $\beta \approx 2^\alpha$), which shows that our Theorem 3.7 is (almost) optimal. On the other hand, to be able to use the assumption $\text{HCon}^{\log^k}(\mathbf{I}\Delta_0 + \Omega_1)$, we must have $\ulcorner \Lambda \urcorner \in \log^k$, which implies $\ulcorner w_\beta \urcorner \approx 2^\beta \approx \exp^2(\alpha) \in \log^k$. Thus, we can have at most $\alpha \in \log^{k+2}$. Whence, we should have $\log^2 \subseteq \log^{k+2}$, or $k + 2 \leq 2$. All this means that our proof can never work for showing the unprovability $\mathbf{I}\Delta_0 + \Omega_1 \not\vdash \text{HCon}^{\log^k}(\mathbf{I}\Delta_0 + \Omega_1)$ for any $0 < k \in \mathbb{N}$. Now, an interesting question may come to mind (see also [4]):

For which (definable) initial segments \mathfrak{J} we have $\mathbf{I}\Delta_0 + \Omega_1 \not\vdash \text{HCon}^{\mathfrak{J}}(\mathbf{I}\Delta_0 + \Omega_1)$, and for which (definable) initial segments \mathfrak{J} , $\mathbf{I}\Delta_0 + \Omega_1 \vdash \text{HCon}^{\mathfrak{J}}(\mathbf{I}\Delta_0 + \Omega_1)$ holds? and what are the boundary cases?

Let us note that for short initial segments like $\mathfrak{L} = \{x \mid \exists y [y = \omega_2(x)]\}$, our proof already implies that $\mathbf{I}\Delta_0 + \Omega_1 \not\vdash \text{HCon}^{\mathfrak{L}}(\mathbf{I}\Delta_0 + \Omega_1)$.

(D) Our treatment of Gödel's second incompleteness theorem for Herbrand consistency in weak arithmetics, can be summarized in the following improvements to the classical treatments (cf. the first paragraph of Appendix E in [17]):

(1) For Skolemizing a formula we did not transform it to a prenex normal form. This allowed a more efficient Skolemization and Herbrandization of formulas.

(2) Propositional satisfiability was achieved by evaluations, which are partial (Herbrand) models; see also [3, 2, 4, 8, 12, 13].

(3) For logarithmic shortening of bounded witnesses in $\mathbf{I}\Delta_0$, we could not go from \log to \log^2 directly. Instead we used the condition $\omega_1^2(x)^8 \in \log$ (equivalently $x \in \mathcal{F}$) to get to $8x^4 \in \log^2$ (equivalently $x \in \mathcal{F}$). For that we used the improved version of Adamowicz's theorem [2] (Theorem 4.2); see also Theorem 3.6 of [15].

(4) Finally, we used the trick of ind_ψ to get a Skolem function symbol for the squaring function. Ideally, one would not use any induction axiom for proving a formula like $\Omega_0: \forall x \exists y (y = x^2)$. This is a Q-derivable sentence, and adding it as an axiom seems much more natural than proving it by an inductive argument. But, fortunately, there was a way of avoiding the acceptance of Ω_0 as an axiom, and that was proving its Π_1 -equivalent $\forall x \exists y \leq x^2 (y = x^2)$ by induction on its bounded part $\exists y \leq x^2 (y = x^2)$ (see Example 2.12 and the proof of Theorem 4.3). That induction axiom could give us a free Skolem function symbol for the squaring operation, provided that we did not prenex normalize the induction axiom, and instead Skolemize it more effectively—see point **D**(1) above. Prenex normalizing and then Skolemizing the induction axioms can be so cumbersome that many would prefer avoiding them, but accepting new axioms instead! Trying to prenex normalize the induction axiom ind_ψ for the formula $\psi(x) = \exists y \leq x^2 (y = x^2)$ in Example 2.12 can give a hint for its difficulty.

(E) In the end, we note that by using our coding techniques and definition of Herbrand consistency, the result of L. A. Kołodziejczyk [8] has been partially generalized to show the following unprovability [15]: $\bigcup_n (\mathbf{I}\Delta_0 + \Omega_n) \not\vdash \text{HCon}(\mathbf{I}\Delta_0)$.

Also, our main result here ($\mathbf{I}\Delta_0 \not\vdash \text{HCon}(\mathbf{I}\Delta_0)$), can be generalized to show the following unprovabilities [14], which also generalize the main results of [16]:

- (1) For some finite fragment $T \subseteq \mathbf{I}\Delta_0$, we have $\mathbf{I}\Delta_0 \not\vdash \text{HCon}(T)$.
- (2) For some $\mathbf{I}\Delta_0$ -derivable Π_1 -sentence S , we have $\mathbf{I}\Delta_0 \not\vdash \text{HCon}(S)$.

We close the paper with an open question:

Can a BOOK proof (in the words of Paul Erdős) be given for Gödel's second incompleteness theorem $T \not\vdash \mathcal{H}\mathbf{Con}(T)$ for any theory $T \supseteq \mathbf{Q}$ and a canonical (standard) definition of Herbrand consistency $\mathcal{H}\mathbf{Con}$?

Acknowledgements. I highly appreciate the comments of the two anonymous referees, which have greatly improved the presentation of this article.

REFERENCES

- [1] ZOFIA ADAMOWICZ, *On tableaux consistency in weak theories*, preprint # 618, Institute of Mathematics, Polish Academy of Sciences, 34 pp., 2001.
- [2] ———, *Herbrand consistency and bounded arithmetic*, *Fundamenta Mathematicae*, vol. 171 (2002), no. 3, pp. 279–292.
- [3] ZOFIA ADAMOWICZ and PAWEŁ ZBIERSKI, *On Herbrand consistency in weak arithmetic*, *Archive for Mathematical Logic*, vol. 40 (2001), no. 6, pp. 399–413.
- [4] ZOFIA ADAMOWICZ and KONRAD ZDANOWSKI, *Lower bounds for the unprovability of Herbrand consistency in weak arithmetics*, *Fundamenta Mathematicae*, vol. 212 (2011), no. 3, pp. 191–216.
- [5] GEORGE S. BOOLOS and RICHARD C. JEFFREY, *Computability and Logic*, Cambridge University Press, 2007.

- [6] SAMUEL R. BUSS, *On Herbrand's theorem*, **Proceedings of the International Workshop on Logic and Computational Complexity, October 13–16, 1994** (D. Maurice and R. Leivant, editors), Lecture Notes in Computer Science 960, Springer-Verlag, 1995, pp. 195–209.
- [7] PETR HÁJEK and PAVEL PUDLÁK, *Metamathematics of first-order arithmetic*, Springer-Verlag, 1998.
- [8] LESZEK A. KOŁODZIEJCZYK, *On the Herbrand notion of consistency for finitely axiomatizable fragments of bounded arithmetic theories*, this JOURNAL, vol. 71 (2006), no. 2, pp. 624–638.
- [9] JAN KRAJÍČEK, *Bounded arithmetic, propositional logic and complexity theory*, Cambridge University Press, 1995.
- [10] JEFF B. PARIS and ALEX J. WILKIE, Δ_0 sets and induction, **Proceedings of Open Days in Model Theory and Set Theory** (Guzicki W., Marek W., Plec A., and Rauszer C., editors), Leeds University Press, 1981, pp. 237–248.
- [11] PAVEL PUDLÁK, *Cuts, consistency statements and interpretations*, this JOURNAL, vol. 50 (1985), no. 2, pp. 423–441.
- [12] SAEED SALEHI, *Unprovability of Herbrand consistency in weak arithmetics*, **Proceedings of the sixth ESSLLI student session, European Summer School for Logic, Language, and Information** (Striegnitz K., editor), 2001, pp. 265–274.
- [13] ———, *Herbrand consistency in arithmetics with bounded induction*, Ph.D. Dissertation, Institute of Mathematics of the Polish Academy of Sciences, 2002, 84 pages, available on the net at <http://saeedsalehi.ir/pphd.html>.
- [14] ———, *Herbrand consistency of some finite fragments of bounded arithmetical theories*, 14 pages, <http://arxiv.org/abs/1110.1848>, 2011.
- [15] ———, *Separating bounded arithmetical theories by Herbrand consistency*, **Journal of Logic and Computation**, vol. 22 (2012), no. 3, pp. 545–560.
- [16] DAN E. WILLARD, *How to extend the semantic tableaux and cut-free versions of the second incompleteness theorem almost to Robinson's arithmetic Q* , this JOURNAL, vol. 67 (2002), no. 1, pp. 465–496.
- [17] ———, *Passive induction and a solution to a Paris–Wilkie open question*, **Annals of Pure and Applied Logic**, vol. 146 (2007), no. 2–3, pp. 124–149.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TABRIZ
P.O. BOX 51666–17766
TABRIZ, IRAN
E-mail: root@saeedsalehi.ir
URL: <http://saeedsalehi.ir/>