On the Herbrand Notion of Consistency for Finitely Axiomatizable Fragments of Bounded Arithmetic Theories

Leszek Aleksander Kołodziejczyk [*]

Institute of Mathematics, Warsaw University

Banacha 2, 02-097 Warszawa, Poland

lak@mimuw.edu.pl

January 23, 2006

Abstract

Modifying the methods of Z. Adamowicz's paper *Herbrand consistency and bounded arithmetic* (Fund. Math. 171 (2002)), we show that there exists a number n such that $\bigcup_m S_m$ (the union of the bounded arithmetic theories S_m) does not prove the Herbrand consistency of the finitely axiomatizable theory S_3^n.

From the point of view of bounded arithmetic, the concept of consistency based on Herbrand's theorem has at least two interesting features. Firstly, it has a more combinatorial flavour than standard consistency notions, and thus lends itself more naturally to combinatorial interpretations (cf. [Pud]). Secondly, it seems reasonably weak.

It is well-known that for stronger concepts of consistency, such as ordinary Hilbert- or Gentzen-style consistency (Cons), or even bounded consistency (consistency with respect to proofs containing only bounded formulae, BdCons), there is typically a very large gap between a given theory and the theories whose consistency it can prove. Wilkie and Paris ([WP87])

---

proved the celebrated result that $I\Delta_0 + \mathrm{Exp} \nvdash \mathrm{Cons}(Q)$. For BdCons, the gap is somewhat smaller, but still significant: Pudlák ([Pud90]) proved that $S_2 \nvdash \mathrm{BdCons}(S_2^1)$, and this was later improved by Buss and Ignjatović ([BI95]) to $S_2 \nvdash \mathrm{BdCons}(S_2^{-1})$, where S_2^{-1} is axiomatized by BASIC plus two additional open axioms. Of course, the existence of such gaps leaves us no hope of using strong consistency notions to separate bounded arithmetic theories.

For Herbrand consistency (HCons), on the other hand, the situation has seemed to be different. In fact, for quite a long time it was even open whether the analogue of Gödel's second incompleteness theorem for Herbrand consistency holds in the case of weak theories. A positive answer was finally given in a series of papers a few years ago. Gödel's theorem for HCons was proved for $I\Delta_0 + \Omega_2$ by Adamowicz and Zbierski ([AZ01]), for $I\Delta_0 + \Omega_1$ by Adamowicz ([Ada01]), and for a certain axiomatization of $I\Delta_0$ and a number of other theories by Willard ([Wil02])[1]. Another paper by Adamowicz, [Ada02], reproved the theorem for $I\Delta_0 + \Omega_2$ (and generally $I\Delta_0 + \Omega_m, m \geq 2$) using a different technique.

In the present paper, we observe that a modification of the methods of [Ada02] leads to a considerably stronger result. It turns out that the "gap phenomenon" known for Cons and BdCons occurs to some degree also for HCons. More specifically, we prove:

Main theorem. *There exists n such that the theory $\bigcup_m S_m$ (or, equivalently, $I\Delta_0 + \bigwedge_{m \in \mathbb{N}} \Omega_m$) does not prove the Herbrand consistency of S_3^n (where S_3^n is defined like S_2^n but with the language expanded by a function symbol for $\#_3$).*

The theorem can be proved with S_2^n instead of S_3^n if a symbol for the unary ω_1 function (equal to $x\#x$) is added to the language. The reason why such a symbol makes a difference is that it allows some Herbrand proofs to become exponentially shorter.

The importance of our result seems to be mainly negative: it rules out the possibility of separating some theories via a Gödel-style argument based on Herbrand consistency. It should be noted, however, that relativizations of Herbrand consistency to definable cuts might be able to prove separations

---

[1]Willard's paper actually shows that $I\Delta_0$ does not prove its own consistency w.r.t. semantic tableau deduction, which is closely related, but not necessarily equivalent (in weak theories) to Herbrand consistency. This result does carry over to Herbrand deduction for some natural axiomatizations of $I\Delta_0$. However, to the author's knowledge it is still open whether it holds for all natural axiomatizations.

where the full notion is too strong. We touch upon this issue briefly towards the end of the paper.

The paper is divided into six sections. The preliminary section 1 discusses some basic definitions and notation. In section 2, we formalize the notion of Herbrand consistency. In section 3, we explain how this formalized notion can be used to obtain new models for a theory. Section 4 contains the proof of our theorem, and section 5 outlines how to modify the proof in order to obtain the result for S_2^n. We conclude with some additional remarks in section 6.

The framework developed in sections 2 and 3, along with some of the reasoning from section 4 (in particular, the proof of Lemma 4.2), is essentially a reconstruction of parts of [Ada02]. Nevertheless, we try to provide a reasonable amount of detail. This makes the paper more self-contained, and additionally, it helps us discuss some of the subtleties involved — see sections 5 and 6.

1   Preliminaries

We assume a general familiarity with bounded arithmetic, in particular, with the meaning of the symbols $I\Delta_0$, $I\Delta_0 + \mathrm{Exp}$, Σ_n^b, Π_n^b, BASIC, S_2^n, S_2, and with notions such as that of a (possibly definable) cut or initial segment of a model. [HP93] or [Kra95] may serve as an introduction.

L_{PA} is the basic language of Peano Arithmetic (we take it to consist of 0, 1, $+$, $\cdot$, and $\leq$). We write L_{BA} to denote the standard language of the bounded arithmetic theory S_2.

The symbols Δ_0, Σ_1, and Π_1 have their usual meanings as names for certain classes of L_{PA}-formulae. In particular, Δ_0 is the class of bounded L_{PA}-formulae. $B\Sigma_1$ is the well-known collection schema for Σ_1 formulae.

For $m \geq 3$, the language L_m is formed by expanding L_{BA} with binary function symbols $\#_3, \ldots, \#_m$. The intended meaning of these symbols is defined inductively: $x\#_{m+1}y$ is $2^{|x|\#_m|y|}$ for $m \geq 2$, where $\#_2$ is just the usual smash function $\#$.

For each $m \geq 3$, the theory S_m^n is an L_m-analogue of S_2^n, with the axioms

$$|x\#_k y| = |x|\#_{k-1}|y| + 1 \quad (k = 3, \ldots, m),$$

3

added to fix the meaning of the new symbols. Like S_2^n, all of the S_m^n theories are finitely axiomatizable for $n \geq 1$. The L_m-analogue of S_2 is called S_m.

Variants of the S_m theories may also be formulated in L_{PA}. The standard way to do this is to use unary functions called ω_m, where $\omega_m(x) = x \#_{m+1} x$ for $m \geq 1$. The graph of each ω_m is Δ_0-definable in $I\Delta_0$, and Ω_m is the sentence $\forall x \, \exists y \, y = \omega_m(x)$. It is easy to see that the theory $I\Delta_0 + \Omega_m$ corresponds very closely to S_{m+1} (more precisely, the latter is a conservative extension of the former).

Some notational conventions: if f is a unary function, then a superscript, as in f^k, always denotes iteration. We sometimes write exp to denote the exponential function 2^x (especially when dealing with iterations). The symbol 2_k is an abbreviation for $\exp^{k-1}(2)$.

The symbol log represents both the obvious function and the cut of logarithmically small elements, i.e. those x for which 2^x exists. Note that if our language contains L_{BA}, the cut log can be defined by the particularly simple formula $\exists y \, x = |y|$. For $k \geq 2$, the initial segment $\log^k$ (which is a cut whenever $\#_{k-1}$ is a total function) consists of the elements for which $\exp^k$ exists. If a model $\mathbf{M}$ is given, the symbols $\log \mathbf{M}$ and $\log^k \mathbf{M}$ stand for the corresponding cuts or segments interpreted in $\mathbf{M}$.

If I is any initial segment, rI is the initial segment determined by elements of I multiplied by r. Similarly, $\omega_m(I)$ is the segment determined by elements of the form $\omega_m(i)$ for $i \in I$.

If a is an element of some model, $a^{\mathbb{N}}$ is the cut determined by the standard powers of a. Likewise, if f is an increasing function, $f^{\mathbb{N}}(a)$ is the cut determined by the standard iterations of f applied to a.

A bar, as in $\bar{x}$, always indicates a tuple. If I is a cut, then $\exists \bar{x} \in I \, \varphi$ is shorthand for a formula stating that there exists $\bar{x}$ all of whose elements are in I such that φ holds.

Finally, a word on sequence coding. Throughout the paper, we may work with any of the standard feasible coding schemes commonly used in bounded arithmetic when L_{BA} is available. However, at one point in the proof of Lemma 4.2 it will be most convenient to use the coding given by Gödel's β-function. Recall that a β-code for a sequence $s_1, \ldots, s_k$ is a pair of appropriately chosen numbers $\langle a, b \rangle$, where b has, among others, the property that all of the numbers $b(j+1)+1$ for $j = 0, \ldots, k$ are pairwise coprime. Each s_j is recovered from $\langle a, b \rangle$ as the remainder of the division of a by $b(j+1)+1$.

In other words, c is the j-th element of the sequence whose β-code is $\langle a, b \rangle$ if

$$\exists d \, (a = d(b(j+1)+1) + c \; \& \; c < b(j+1)+1). \tag{1}$$

One helpful feature of β-codes is that that the formula in (1) is very simple — once d is given, it is actually an open formula. We also note that any bounded and logarithmically long sequence (in the sense of some standard coding) has a β-code.

2 Herbrand consistency via evaluations

Let $\varphi(\bar{x})$ be a formula of the form

$$\exists x_1 \forall y_1 \ldots \exists x_k \forall y_k \, \tilde{\varphi}(x_1, y_1, \ldots, x_k, y_k),$$

where $\tilde{\varphi}$ is open (any formula can be transformed into a logically equivalent formula of this shape by taking the prenex normal form and adding superfluous quantifiers). The Herbrandization of φ, He(φ), is defined as

$$\exists x_1 \ldots \exists x_k \, \tilde{\varphi}(x_1, f_1(x_1), \ldots, x_k, f_k(x_1, \ldots, x_k)),$$

where the f_is are new function symbols, called Herbrand functions. Notice that Herbrandization is dual to Skolemization, i.e. He(φ) is the negation of the Skolemization of $\neg\varphi$ (modulo the de Morgan laws). Obviously, φ is provable (in first-order logic) if and only if He(φ) is provable. *Herbrand's theorem* states that φ is provable if and only if a finite disjunction of formulae of the form

$$(+) \; \tilde{\varphi}(t_1, f_1(t_1), \ldots, t_k, f_k(t_1, \ldots, t_k)),$$

where the t_is are terms of the language extended by adding Herbrand functions, is a *propositional* tautology. Any formula as in $(+)$ is called an *Herbrand variant* of φ, while a tautological disjunction of Herbrand variants is sometimes called an *Herbrand proof* of φ.

Now let T be a finitely axiomatizable arithmetical theory, say, $T = \{\tau_1, \ldots, \tau_r\}$. We shall call T *Herbrand inconsistent* if there is an Herbrand proof of $\neg\tau_1 \vee \ldots \vee \neg\tau_r$. T is *Herbrand consistent* otherwise.

By Herbrand's theorem, the Herbrand consistency of a theory is equivalent to consistency in the usual Hilbert-style sense. However, in weak theories

5

which do not prove Herbrand's theorem, these two concepts may well differ. In this case, the class of "weak theories" includes even $I\Delta_0 + \text{Exp}$. Indeed, there are theories, such as Robinson's Q and $I\Delta_0$, whose Herbrand consistency is provable in $I\Delta_0 + \text{Exp}$ but whose Hilbert-style consistency is not.

In the remainder of this section, we develop the formalization of Herbrand consistency used in the present paper.

Assume T is as above. We expand the language of T, L_T, by adding Herbrand functions $f_1^j, \ldots, f_{i_j}^j$ of the appropriate arity for each axiom τ_j of T (of course, the number i_j depends on the number of quantifiers in τ_j). Call this extended language L_T^H. Call the set of closed L_T^H-terms $\text{Term}(L_T^H)$.

L_T^H is a finite language, in the sense that it contains only a finite number of constant individual, predicate, and function symbols. Hence, we may gödelize L_T^H in such a way that a closed L_T^H-term of length l will have a Gödel number of about c^l, where c is a fixed standard natural number. Thus, the Gödel number of an L_T^H-term of the form $f(t_1, \ldots, t_k)$ will be polynomial in the numbers for $t_1, \ldots, t_k$, and the length $\text{lh}(\cdot)$ of a term will be linearly related to the length of its Gödel number in the sense of the function $|\cdot|$. From now on, we identify elements of $\text{Term}(L_T^H)$ with their Gödel numbers.

Consider any interval of the form $[0, a]$, where $a \in \log$. The *simple atomic sentences* over $[0, a]$ are the sentences obtained by substituting elements of $\text{Term}(L_T^H) \cap [0, a]$ for variables in formulae of the form $x_1 = x_2$, $R(x_1, \ldots, x_k)$, or $f(x_1, \ldots, x_k) = x_{k+1}$, where R, f are symbols of L_T.

An *evaluation* over $[0, a]$ is any assignment p of truth values (0 or 1) to (all) simple atomic sentences over $[0, a]$ which respects the equality axioms. We will write $p \models \psi$ for $p(\psi) = 1$.

The code of an evaluation over $[0, a]$ is roughly exponential in the number of simple atomic sentences over $[0, a]$, which in turn is polynomial in a. So, the formalization of evaluations works well in any theory which proves the totality of $\#$ (and thus the closure of log under multiplication).

An evaluation over $[0, a]$ can be extended in the obvious way to arbitrary open sentences whose terms are from $[0, a]$. This allows us to define the concept of a *T-evaluation*: p is a T-evaluation over $[0, a]$ if $p \models \psi$ for any ψ obtained by substituting terms from $[0, a]$ into the open part of the Skolemization of some τ_j (where we treat the f_i^js as Skolem functions)[2].

A weak form of the Herbrand consistency of T may now be expressed by

---

[2]Naturally, we only consider those ψ in which all the terms, not just the ones substituted for variables, belong to $[0, a]$.

the following sentence $\mathrm{HCons}^*(T)$:

> for any $a \in \log$, there exists a T-evaluation over $[0, a]$.

To see that HCons^* is implied by any adequate formulation of Herbrand consistency, consider its negation. $\neg\mathrm{HCons}^*(T)$ means that for some a, there is no T-evaluation over $[0, a]$. In other words, there is no way to assign truth values to atomic formulae over $[0, a]$ in such a way that all instances of the Skolemizations of axioms of T become true. But since an instance of the Skolemization of τ_j is precisely the negation of an Herbrand variant of $\neg\tau_j$, this means that the disjunction of all Herbrand variants of the $\neg\tau_j$s involving terms from $[0, a]$ is a propositional tautology. Thus, T is Herbrand inconsistent.

Since HCons^* only mentions Herbrand disjunctions of relatively small terms (the terms are to be from log), the above argument does not exclude the possibility that in some cases HCons^* is essentially weaker than full Herbrand consistency. In this paper, however, we are aiming for results on the *un*provability of Herbrand consistency, so this is not a problem.

Remark. Concerning the requirement that an evaluation respect the equality axioms: this corresponds to treating equality axioms as part of the theory (or of logic), and is normally needed to ensure that models obtained from T-evaluations (see the next section) are well-defined and satisfy T.

Actually, however, for the purposes of the present paper this requirement is merely convenient, not strictly necessary. For theories in the language L_{BA} (containing BASIC), the reason is that one of the BASIC axioms characterizes $x = y$ as $x \leq y \,\&\, y \leq x$, and other axioms imply that all the operations of L_{BA} respect equality in the sense of this characterization. The $\#_m$ functions for $m > 2$ might not satisfy the equality axioms exactly, as the axioms only define them up to $|\cdot|$. Nevertheless, a model obtained from an S_m^n-evaluation, $m > 2$, is a well-defined model for S_2^n and can be expanded in a canonical way to satisfy the totality of $\#_m$. This expansion also satisfies S_m^n.

3 Models from evaluations

In this section, we discuss how the existence of T-evaluations over appropriately large intervals can be used to construct models of T. The basic idea behind the construction seems to date back to Lemma 3.3 of [Pud85].

Let $a \in \log \mathbf{M}$, where $\mathbf{M}$ is a (nonstandard) model of some amount of arithmetic, and let p be a T-evaluation on $[0, b]$, where $b \in \log \mathbf{M}$, $b > a^{\mathbb{N}}$. We may use p to define a model $\mathbf{M}(p, a)$ as follows.

Define the relation $=_p$ on the set $\mathrm{Term}(L_T^H) \cap [0, b]$ by:

$$t_1 = t_2 \text{ iff } p \models t_1 = t_2.$$

Since p respects the equality axioms, this is an equivalence relation. The universe of $\mathbf{M}(p, a)$ is $(\mathrm{Term}(L_T^H) \cap a^{\mathbb{N}})/_{=_p}$. The relations and operations are defined in the natural way: $t_1 \leq t_2$ holds in $\mathbf{M}(p, a)$ iff $p \models t_1 \leq t_2$; $[t_1] + [t_2]$ is $[t_1 + t_2]$, etc. This is correct as, firstly, $=_p$ is a congruence with respect to all the operations (equality axioms again), and secondly, the closure of $a^{\mathbb{N}}$ under multiplication guarantees that all the operations can be defined (recall from section 2 that for any function symbol f of L_T and even L_T^H, the term $f(t_1, \ldots, t_k)$ is just polynomially larger than the terms $t_1, \ldots, t_k$)).

Obviously, for any open L_T-formula φ,

$$\mathbf{M}(p, a) \models \varphi([a_1], \ldots, [a_k]) \text{ iff } p \models \varphi(a_1, \ldots, a_k).$$

In particular, since p is a T-evaluation, $\mathbf{M}(p, a)$ satisfies each instance of the Skolemization of any τ_j in which terms from $\mathrm{Term}(L_T^H) \cap a^{\mathbb{N}}$ are substituted for the universally quantified variables. But since all elements of $\mathbf{M}(p, a)$ are given by terms from $\mathrm{Term}(L_T^H) \cap a^{\mathbb{N}}$, this simply means that $\mathbf{M}(p, a)$ satisfies any τ_j, and hence, $\mathbf{M}(p, a) \models T$.

We will now show that if T contains BASIC, then the model $\mathbf{M}(p, a)$ has another important property: the cut $a^{\mathbb{N}}$ in $\mathbf{M}$ is isomorphic to an initial segment of $\mathbf{M}(p, a)$, at least with respect to the basic arithmetical language of addition and multiplication.

To define the isomorphism, we need to have canonical terms naming the elements of $a^{\mathbb{N}}$. It is natural to choose the well-known *dyadic numerals*, defined inductively as follows. The numerals $\underline{0}, \underline{1}, \underline{2}$ are the terms 0, 1, and 1+1, respectively (we do not write parentheses explicitly unless it is necessary); $\underline{2k}$ is $\underline{k} \cdot \underline{2}$ for $k \geq 2$; and $\underline{2k+1}$ is $\underline{2k}+1$ for $k \geq 1$. It should be noted that $\underline{k}$ is a term of logarithmic length w.r.t. k, so it is bounded by k^c for some fixed standard c. Thus, the numerals for elements of $a^{\mathbb{N}}$ are themselves in $a^{\mathbb{N}}$.

Lemma 3.1. *If* $\mathbf{M} \models S_2^1$ *and* T *contains* BASIC*, then* $a^{\mathbb{N}}$ *is isomorphic (w.r.t. the language* L_{PA}*) to an initial segment of* $\mathbf{M}(p, a)$.

Proof. Of course, it is enough to show that the interval $[0, a]$ is isomorphic to an initial segment of $\mathbf{M}(p, a)$ (with addition and multiplication treated as relations since they are not total). The isomorphism will map each $i \leq a$ to $[\underline{i}]$. To prove that this really is an isomorphism, we will need to use induction (in $\mathbf{M}$) a number of times. In each case, it will be clear that S_2^1 suffices for the inductive arguments.

We first prove that for each $i < a$, $p \models \underline{i+1} = \underline{i} + 1$. The proof is by induction, with the inductive step split into two cases. The case where $i = 2k$ for some k is obvious, so we only deal with the case where $i = 2k + 1$. In this case $\underline{i+1}$ is, by definition, $\underline{k+1} \cdot \underline{2}$. By BASIC,

$$p \models \underline{i+1} = \underline{k+1} + \underline{k+1},$$

so by the inductive assumption,

$$p \models \underline{i+1} = (\underline{k} + 1) + (\underline{k} + 1).$$

On the other hand, $\underline{i} + 1$ is $(\underline{2k} + 1) + 1$, so by BASIC,

$$p \models \underline{i} + 1 = ((\underline{k} + \underline{k}) + 1) + 1.$$

Since the axioms of BASIC include the associativity and commutativity of addition, it follows that $p \models \underline{i+1} = \underline{i} + 1$.

The next step is to show that for $i, j \leq a$, $p \models \underline{i} + \underline{j} = \underline{i+j}$ (this actually holds even if $i + j > a$). But once we know that p identifies $\underline{j+1}$ with $\underline{j} + 1$, this is a very straightforward inductive argument which uses the two axioms $x + 0 = x$ and $x + (y + 1) = (x + y) + 1$ contained in BASIC.

Similar arguments show that an analogous fact holds for multiplication, and that the mapping given by $i \mapsto [\underline{i}]$ is injective and preserves the ordering. Altogether, we have obtained an isomorphic embedding of $[0, a]$ into $\mathbf{M}(p, a)$. It remains to prove that its range is an initial segment.

To this end, we will show the following:

> for any $i \leq a$ and any term t, if $p \models t \leq \underline{i}$,
> then there is a $j \leq i$ such that $p \models t = \underline{j}$.

The argument is by induction on i. The base step uses the BASIC axioms $0 \leq x$ and $(x \leq y \;\&\; y \leq x) \Rightarrow x = y$. In the inductive step, assuming $p \models t \leq \underline{i+1}$, we use BASIC to get

$$p \models t = \underline{i+1} \vee t + 1 \leq \underline{i+1}.$$

If the first disjunct holds, then we are done, if it is the second, then the axiom $x + y \leq x + z \Rightarrow y \leq z$ and some work in BASIC gives $p \models t \leq \underline{i}$, so we may use the inductive assumption. $\qquad\square$

The following immediate corollary will play an important role in the next section (see the proof of Lemma 4.2).

Corollary 3.2. *If* $\mathbf{M} \models S_2^1$ *and* T *contains* BASIC, *then for any* Δ_0 *formula* θ *and any* $a_1, \ldots, a_k \leq a$,

$$\mathbf{M}(p, a) \models \theta([\underline{a_1}], \ldots, [\underline{a_k}]) \ \textit{iff} \ \mathbf{M} \models \theta(a_1, \ldots, a_k).$$

4 The results

In the present section, we prove our main theorem. We will first prove a somewhat weaker version, Theorem 4.1, and then discuss how to modify the proof in order to obtain the stronger result.

Theorem 4.1. *For every* $m \geq 3$ *there exists an* n *such that* S_m *does not prove the Herbrand consistency of* S_m^n.

We start by proving an adaptation of a crucial technical lemma from [Ada02] (the proof is also adapted from [Ada02]). The original version states, in essence, that if a bounded arithmetic theory T proved its own Herbrand consistency, then any appropriately small witness for a Δ_0 formula in a model of T could be made still logarithmically smaller in another model of T. The observation we need here is that a similar "witness reduction" phenomenon occurs also when we consider the provability of Herbrand consistency for one theory in a possibly different theory.

Lemma 4.2. *Let* $m \geq 3, n \geq 1, k \geq 0$, *and assume that* $S_m^{n+k} \vdash \mathrm{HCons}^*(S_m^n)$. *Then for any* Δ_0 *formula* $\theta(\bar{x})$ *such that* $S_m^{n+k} + \exists \bar{x} \in \log^m \theta(\bar{x})$ *is consistent,* $S_m^n + \exists \bar{x} \in \log^{m+1} \theta(\bar{x})$ *is also consistent.*

Proof. Assume $S_m^{n+k} \vdash \mathrm{HCons}^*(S_m^n)$ and consider a Δ_0 formula $\theta(\bar{x})$ such that $S_m^{n+k} + \exists \bar{x} \in \log^m \theta(\bar{x})$ is consistent. Take any nonstandard model $\mathbf{M} \models S_m^{n+k}$ which contains a tuple $\bar{a} = \langle a_1, \ldots, a_k \rangle$ such that $\mathbf{M} \models \theta(\bar{a})$ and all elements of $\bar{a}$ are in $\log^m$. We will use the fact that $\mathbf{M} \models \mathrm{HCons}^*(S_m^n)$ to construct a model of S_m^n in which $\exists \bar{x} \in \log^{m+1} \theta(\bar{x})$ holds.

Assume w.l.o.g. that a_1 is maximal among $a_1, \ldots, a_k$. Take any $b \in \log \mathbf{M}$ greater than $(2^{2^{a_1}})^{\mathbb{N}}$ (such a number b must exist, since $2^{2^{a_1}}$ is in $\log \mathbf{M}$, which is closed at least under $\#$). By $\mathrm{HCons}^*(S_m^n)$, there exists an S_m^n-evaluation p on $[0, b]$. Consider the model $\mathbf{M}(p, 2^{2^{a_1}})$.

By the previous section, $\mathbf{M}(p, 2^{2^{a_1}}) \models S_m^n$. Moreover, by Corollary 3.2,

$$\mathbf{M}(p, 2^{2^{a_1}}) \models \theta([\underline{a_1}], \ldots, [\underline{a_k}]).$$

So, all we need to show is that the elements $[\underline{a_1}], \ldots, [\underline{a_k}]$ are contained in $\log^{m+1} \mathbf{M}(p, 2^{2^{a_1}})$. Naturally, we may restrict ourselves to $[\underline{a_1}]$.

In all models of S_m^1 and for all elements a, $\exp^{m+1}(a)$ is roughly equal to

$$\underbrace{2_m \#_m \cdots \#_m 2_m}_{2^a}$$

whenever these two values exist (actually, the latter is somewhat larger, as $\#_m$ is defined in terms of the $|\cdot|$ function, whose values are a bit larger than the real binary logarithm). So, it is enough to prove that in $\mathbf{M}(p, 2^{2^{a_1}})$, there exists

$$\underbrace{2_m \#_m \cdots \#_m 2_m}_{2^{a_1}}.$$

The term

$$t := \overbrace{\underline{2_m} \#_m \cdots \#_m \underline{2_m}}^{2^{a_1}}$$

certainly is an element of $(2^{2^{a_1}})^{\mathbb{N}}$ (to see this, just note that the length of t is linear in $2^{a_1} \approx |2^{2^{a_1}}|$). Thus, we just have to check that the relation between $[\underline{a_1}]$ and $[t]$ in $\mathbf{M}(p, 2^{2^{a_1}})$ is indeed as it should be.

Working in $\mathbf{M}(p, 2^{2^{a_1}})$, consider the sequence

$$\langle 2_m, 2_m \#_m 2_m, \ldots, \underbrace{2_m \#_m \cdots \#_m 2_m}_{[i]} \rangle,$$

where i is the largest number $\leq 2^{a_1}$ such that

$$\underbrace{2_m \#_m \cdots \#_m 2_m}_{[i]} \leq [t]$$

(note that by Lemma 3.1, we can use induction in $\mathbf{M}(p, 2^{2^{a_1}})$ to find the largest element $\leq 2^{a_1}$ satisfying this condition). This sequence contains only

11

elements below $[t]$ and is at most logarithmically long, so it has a β-code in $\mathbf{M}(p, 2^{2^{a_1}})$. Let this β-code be $\langle [t_A], [t_B] \rangle$, where t_A and t_B are terms from $\mathrm{Term}(L_{S^n_m}^H) \cap (2^{2^{a_1}})^{\mathbb{N}}$ (in $\mathbf{M}$).

For any $j \leq i$, there are some t_j, t'_j in $\mathrm{Term}(L_{S^n_m}^H) \cap (2^{2^{a_1}})^{\mathbb{N}}$ such that $[t'_j]$ witnesses that $[t_j]$ is the $[\underline{j}]$-th element of the sequence whose β-code is $\langle [t_A], [t_B] \rangle$, i.e. $\mathbf{M}(p, 2^{2^{a_1}})$ satisfies

$$[t_A] = [t'_j] \cdot ([t_B] \cdot ([\underline{j}] + 1) + 1) + [t_j] \,\&\, [t_j] < [t_B] \cdot ([\underline{j}] + 1) + 1.$$

Since the above is an open formula, in $\mathbf{M}$ we have

$$p \models t_A = t'_j \cdot (t_B \cdot (\underline{j} + 1) + 1) + t_j \,\&\, t_j < t_B \cdot (\underline{j} + 1) + 1. \qquad (2)$$

For each $j \leq i$, consider the smallest pair $\langle t_j, t'_j \rangle \in \mathrm{Term}(L_{S^n_m}^H) \cap (2^{2^{a_1}})^{\mathbb{N}}$ for which (2) holds (already S_2^1 is enough to do this, as we just need to find the smallest appropriate pair $\langle t_j, t'_j \rangle$ below b, and $b \in \log \mathbf{M}$). An easy induction in $\mathbf{M}$ shows that for all $j \leq i$,

$$p \models t_j = \overbrace{\underline{2_m} \#_m \cdots \#_m \underline{2_m}}^{j}.$$

In particular, p identifies t_i with the term

$$\overbrace{\underline{2_m} \#_m \cdots \#_m \underline{2_m}}^{i}.$$

We are now ready to show that $i = 2^{a_1}$, which will complete the proof. Suppose the contrary, i.e. $i < 2^{a_1}$. Then $i + 1 \leq 2^{a_1}$, and hence it is easy to check that

$$p \models \overbrace{\underline{2_m} \#_m \cdots \#_m \underline{2_m}}^{i+1} \leq t.$$

The terms $\overbrace{\underline{2_m} \#_m \cdots \#_m \underline{2_m}}^{i+1}$ and $\overbrace{\underline{2_m} \#_m \cdots \#_m \underline{2_m}}^{i} \#_m \underline{2_m}$ are identical, therefore

$$p \models t_i \#_m \underline{2_m} \leq t,$$

so $\mathbf{M}(p, 2^{2^{a_1}})$ satisfies $[t_i] \#_m 2_m \leq [t]$. But $[t_i]$ is equal to

$$\underbrace{2_m \#_m \cdots \#_m 2_m}_{[i]}$$

for the *largest* $i \leq 2^{a_1}$ for which this is not greater than $[t]$. Contradiction. $\qquad \square$

We may now employ Lemma 4.2 to prove Theorem 4.1 in the following way.

Proof of Theorem 4.1. Assume that:

$(*)$ for some $m \geq 3$, S_m proves $\mathrm{HCons}^*(S_m^n)$ for all n.

By compactness, it follows from $(*)$ that there exists an increasing sequence $\langle n_k : k \in \mathbb{N} \rangle$ of natural numbers such that $n_0 \geq 1$ and for each k, $S_m^{n_{k+1}} \vdash \mathrm{HCons}^*(S_m^{n_k})$. Having fixed such a sequence, we will use Lemma 4.2 to derive the following:

$(**)$ for every Δ_0 formula $\theta(\bar{x})$, if $S_m + \exists \bar{x} \in \log^m \theta(\bar{x})$ is consistent, then $I\Delta_0 + \mathrm{Exp} + \exists \bar{x}\, \theta(\bar{x})$ is consistent.

The statement $(**)$ is known to be false (see the proof of theorem 1.1 in [Ada02] or theorem V.5.36 in [HP93]). Therefore, showing that $(*)$ implies $(**)$ will complete the proof of the theorem.

Thus, let us consider a Δ_0 formula $\theta(\bar{x})$ such that $S_m + \exists \bar{x} \in \log^m \theta(\bar{x})$ is consistent. We claim that:

$(***)$ for any $l \geq 1$, $S_m^1 + \exists \bar{x} \in \log^{m+l} \theta(\bar{x})$ is consistent.

Indeed, fix l and consider $S_m^{n_l}$. Since $S_m^{n_l} + \exists \bar{x} \in \log^m \theta(\bar{x})$ is consistent and $S_m^{n_l} \vdash \mathrm{HCons}^*(S_m^{n_{l-1}})$, Lemma 4.2 implies that $S_m^{n_{l-1}} + \exists \bar{x} \in \log^{m+1} \theta(\bar{x})$ is consistent. It follows that $S_m^{n_{l-1}} + \exists y \in \log^m \exists \bar{x} \leq \log y\, \theta(\bar{x})$ is also consistent. Since $S_m^{n_{l-1}} \vdash \mathrm{HCons}^*(S_m^{n_{l-2}})$ and $\exists \bar{x} \leq \log y\, \theta(\bar{x})$ is Δ_0, we may use Lemma 4.2 again to obtain the consistency of $S_m^{n_{l-2}} + \exists y \in \log^{m+1} \exists \bar{x} \leq \log y\, \theta(\bar{x})$, or, equivalently, of $S_m^{n_{l-2}} + \exists \bar{x} \in \log^{m+2} \theta(\bar{x})$. Iterating this procedure l times, we will eventually show that $S_m^{n_0} + \exists \bar{x} \in \log^{m+l} \theta(\bar{x})$ is consistent. This completes the proof of claim $(***)$, as n_0 was assumed to be at least 1.

Applying compactness to $(***)$, we see that the theory

$$T := S_m^1 + \theta(\bar{d}) + \{\bar{d} \in \log^l : l \in \mathbb{N}\}$$

is consistent (where $\bar{d}$ is a tuple of new constants of the appropriate length). Consider any $\mathbf{M} \models T$. $\mathbf{M}$ is a model for S_m^1, and it contains a tuple $\bar{d}$ satisfying θ such that all standard iterations of the exponential function on elements of $\bar{d}$ exist. Let $\mathbf{M}'$ be the submodel of $\mathbf{M}$ determined by the cut $\exp^{\mathbb{N}}(\max(\bar{d}))$. Clearly, $\mathbf{M}' \models S_m^1 + \mathrm{Exp} + \theta(\bar{d})$. But the theory $S_m^1 + \mathrm{Exp}$ is equal to $I\Delta_0 + \mathrm{Exp}$, except for the difference in language (this is essentially established by the standard proof of the finite axiomatizability of $I\Delta_0 + \mathrm{Exp}$),

so $\mathbf{M}$ is a model for $I\Delta_0 + \mathrm{Exp} + \exists\bar{x}\,\theta(\bar{x})$. This concludes the proof of $(*)$, and hence also of Theorem 4.1. $\square$

The proof of the main theorem is quite similar to the one given above, but it requires two additional observations. The first of these is that an analogue of Lemma 4.2 still holds if the two theories involved differ not only in the amount of induction they prove, but also in the growth-rate of functions they make total[3]. This can be shown in the same way as Lemma 4.2.

Lemma 4.3. *Let $m \geq 3, n \geq 1, k, l \geq 0$, and assume $S_{m+l}^{n+k} \vdash \mathrm{HCons}^*(S_m^n)$. Then for any Δ_0 formula $\theta(\bar{x})$ such that $S_{m+l}^{n+k} + \exists\bar{x} \in \log^3 \theta(\bar{x})$ is consistent, $S_m^n + \exists\bar{x} \in \log^4 \theta(\bar{x})$ is also consistent.*

Proof idea. The proof is entirely analogous to that of Lemma 4.2 for the case of $m = 3$. In particular, the term

$$\overbrace{\underline{16}\#_3 \ldots \#_3 \underline{16}}^{2^{a_1}}$$

is used (where a_1 has the same meaning as in the proof of Lemma 4.2). $\square$

Now the second observation: if we are able to make a witness for some formula exponentially smaller in a model of S_3^n, then we can also make it almost exponentially smaller in a model of S_m^n for some higher m:

Lemma 4.4. *The family of theories $\{S_m^n\}_{n,m \geq 1}$ has the following properties:*

(i) *Let $m \geq 3, n \geq 1, k \geq 0$, and assume $S_m^{n+k} \vdash \mathrm{HCons}^*(S_3^n)$. Then for any Δ_0 formula $\theta(\bar{x})$ such that $S_m^{n+k} + \exists\bar{x} \in \log^3 \theta(\bar{x})$ is consistent, $S_m^n + \exists\bar{x} \in \omega_{m-4}(\log^4)\,\theta(\bar{x})$ is also consistent[4].*

(ii) *Let $m \geq 3, n \geq 1, k, l \geq 0$, and assume that $S_m^{n+k+l} \vdash \mathrm{HCons}^*(S_3^{n+k})$ and $S_m^{n+k} \vdash \mathrm{HCons}^*(S_3^n)$. Then for any Δ_0 formula $\theta(\bar{x})$ such that $S_m^{n+k+l} + \exists\bar{x} \in \log^3 \theta(\bar{x})$ is consistent, $S_m^n + \exists\bar{x} \in \log^4 \theta(\bar{x})$ is also consistent.*

---

[3] I owe the idea that a version of Lemma 4.2 should be true in this situation to Konrad Zdanowski.

[4] In both the lemma and the proof, we are using the convention that $\omega_0(x) = x^2$, $\omega_{-1}(x) = 2x$, and $\omega_{-2}(x) = x + 1$.

Proof. Assume the hypothesis of part (i) and consider $\theta(\bar{x}) \in \Delta_0$ such that $S_m^{n+k} + \exists \bar{x} \in \log^3 \theta(\bar{x})$ is consistent. By Lemma 4.3, there exists a (nonstandard) model $\mathbf{M} \models S_3^n$ containing a tuple $\bar{d}$ such that $\max(\bar{d}) \in \log^4 \mathbf{M}$ and $\mathbf{M} \models \theta(\bar{d})$. Let $d := \max(\bar{d})$. W.l.o.g. we may consider only the case where d is nonstandard.

In that case, there is some $c \in \mathbf{M}$ such that $\omega_{m-4}(c) \geq d$, but $\omega_{m-5}^{\mathbb{N}}(c) < d$. Indeed: already S_2^1 proves that there exists a smallest number c such that $\omega_{m-4}(c) \geq d$. It can also be proved that there is a value of ω_{m-4} between d and $\omega_{m-5}(\omega_{m-5}((d))$. Thus, $\omega_{m-4}(c) \leq \omega_{m-5}(\omega_{m-5}((d)))$, and since ω_{m-4} dominates all standard iterations of ω_{m-5} for all nonstandard arguments, it follows that $\omega_{m-5}^{\mathbb{N}}(c) < d$.

We have $\omega_{m-1}^{\mathbb{N}}(\exp^4(c)) < \exp^4(d)$, as $\omega_{m-1}(\exp^4(x))$ is approximately $\exp^4(\omega_{m-5}(x))$. Let $\mathbf{M}'$ be the submodel of $\mathbf{M}$ determined by $\omega_{m-1}^{\mathbb{N}}(\exp^4(c))$. Clearly, $\mathbf{M}' \models S_3^n$, and since ω_{m-1}, and thus $\#_m$, is total in $\mathbf{M}'$, we also get $\mathbf{M}' \models S_m^n$. Furthermore, $\mathbf{M}' \models \theta(\bar{d})$. Finally, $d \in \omega_{m-4}(\log^4 \mathbf{M}')$, because $d \leq \omega_{m-4}(c)$, and c is obviously in $\log^4 \mathbf{M}'$. Hence, part (i) of the lemma is proved.

Now let m, n, k, l be as required for part (ii) and let $\theta(\bar{x})$ be such that $S_m^{n+k+l} + \exists \bar{x} \in \log^3 \theta(\bar{x})$ is consistent. We will apply part (i) twice. The first application, with $n := n+k$ and $k := l$, proves that S_m^{n+k} is consistent with $\exists \bar{x} \in \omega_{m-4}(\log^4) \theta(\bar{x})$. In the second application, taking $\theta(y)$ to be $\exists \bar{x} \leq \omega_{m-4}(\log y) \theta(\bar{x})$, we get the consistency of

$$\exists y \in \omega_{m-4}(\log^4) \exists \bar{x} \leq \omega_{m-4}(\log y) \theta(\bar{x})$$

with S_m^n. This means that in some model of S_m^n, a tuple of witnesses for $\theta(\bar{x})$ may be found in $\omega_{m-4}(\log(\omega_{m-4}(\log^4)))$, or equivalently in $\omega_{m-4}(\omega_{m-5}(\log^5))$. But in any model of S_m^n, $\omega_{m-4}(\omega_{m-5}(\log^5))$ is contained in $\log^4$, so in fact we may find the tuple of witnesses in $\log^4$. $\qquad\square$

Proof of Main theorem. Assume that $\bigcup_m S_m$ does prove $\mathrm{HCons}^*(S_3^n)$ for all n. In that case, there exists an increasing sequence $\langle n_k : k \in \mathbb{N} \rangle$ such that $n_0 \geq 3$ and for each k, there exists a number $n \in [n_k, n_{k+1}]$ such that $S_{n_{k+1}}^{n_{k+1}} \vdash \mathrm{HCons}^*(S_3^n)$ and $S_{n_{k+1}}^n \vdash \mathrm{HCons}^*(S_3^{n_k})$.[5] Fix such a sequence.

We now emulate the proof of Theorem 4.1, using Lemma 4.4(ii) instead of Lemma 4.2, to get

---

[5]The sequence $\langle n_k : k \in \mathbb{N} \rangle$ can be defined inductively. Let n_0 be any number greater or equal to 3. Given n_k, let $\tilde{n}_k \geq n_k$ be any number such that $S_{\tilde{n}_k}^{\tilde{n}_k} \vdash \mathrm{HCons}^*(S_3^{n_k})$, and let $n_{k+1} \geq \tilde{n}_k$ be such that $S_{n_{k+1}}^{n_{k+1}} \vdash \mathrm{HCons}^*(S_3^{\tilde{n}_k})$. Obviously, $\tilde{n}_k$ may play the role of n.

($\circ$) for any Δ_0 formula $\theta(\bar{x})$, if $\bigcup_m S_m + \exists \bar{x} \in \log^3 \theta(\bar{x})$ is consistent, then $I\Delta_0 + \mathrm{Exp} + \exists \bar{x} \, \theta(\bar{x})$ is consistent.

All that remains is to show that ($\circ$) is false. To see this, one may apply an argument analogous to the one used to disprove the statement ($**$) (from the proof of Theorem 4.1) in the proof of theorem 1.1 in [Ada02]. One small modification is needed: we have to show that any model $\mathbf{M}$ of $I\Delta_0 + \mathrm{Exp} + B\Sigma_1$ has a proper end-extension to a model of $I\Delta_0 + \bigwedge_{m \in \mathbb{N}} \Omega_m$. But this is easy: take the proper end-extension of $\mathbf{M}$ to a model $\mathbf{M}' \models I\Delta_0$ guaranteed to exist by [WP89], and let $\mathbf{M}''$ be the submodel of $\mathbf{M}'$ determined by the cut $\{a \in \mathbf{M}' : \omega_m^{(m)}(a) \text{ exists for all standard } m\}$. $\mathbf{M}''$ is clearly an end-extension of $\mathbf{M}$ to a model of $I\Delta_0 + \bigwedge_{m \in \mathbb{N}} \Omega_m$, and it is straightforward to show that the extension is proper[6]. $\qquad \square$

5 The case of S_2^n

One would like to extend our main theorem, or at least Theorem 4.1, to the most important case of $m = 2$. Unfortunately, there is a technical problem here. An extension of Lemma 4.2 to $m = 2$ would involve "pushing witnesses down" from $\log^2$ to $\log^3$. But the term required to do this,

$$\overbrace{\underline{4}\# \ldots \#\underline{4}}^{2^a},$$

is polynomially related to 2^{2^a}. Thus, we would need the model $\mathbf{M}(p, 2^{2^a})$, which may be unavailable as 2^{2^a} will in general fall outside $\log$.

Interestingly, however, even our main result does extend to $m = 2$ if the vocabulary is modified by adding a function symbol for ω_1. The benefit of having ω_1 is that it is a unary function, so it can be used to define fast-growing functions using shorter terms. In particular, for any x,

$$\omega_1^a(x) \approx \underbrace{x \# \ldots \# x}_{2^a},$$

so $\exp^3(a)$ is roughly equal to $\omega_1^a(4)$. The size of the term $\omega_1^a(\underline{4})$ is about 2^a, which is in $\log$ for $a \in \log^2$. With this in mind, one may reconstruct the argument of section 4 to prove:

---

[6]This particularly simple way of getting end-extensions to models of $I\Delta_0 + \bigwedge_{m \in \mathbb{N}} \Omega_m$ was suggested to me by Zofia Adamowicz.

Theorem 5.1. *Let $S_2^{n,*}$ be defined by adding the axiom*

$$\omega_1(x) = x \# x$$

to S_2^n. Then there exists n such that $\bigcup_m S_m$ does not prove $\mathrm{HCons}^*(S_2^{n,*})$.

In what follows, we will briefly explain the details of the reconstruction. The reader uninterested in those details may prefer to skip the rest of the present section, save perhaps the remark at the very end.

We want to proceed as in the proof of our main theorem. Therefore, we need appropriately modified versions of Lemmata 4.3 and 4.4. The analogue of Lemma 4.3 is as follows:

Lemma 5.2. *Let $m \geq 2, n \geq 1, k \geq 0$, and assume $S_m^{n+k} \vdash \mathrm{HCons}^*(S_2^{n,*})$. Then for any Δ_0 formula $\theta(\bar{x})$ such that $S_m^{n+k} + \exists \bar{x} \in \log^2 \theta(\bar{x})$ is consistent, $S_2^n + \exists \bar{x} \in \log^3 \theta(\bar{x})$ is also consistent.*

Proof idea. The reasoning is again based on the idea used in the proof of Lemma 4.2. This time, we need to move a tuple $\bar{a}$ of witnesses for $\theta(\bar{x})$ down from $\log^2$ in a nonstandard model $\mathbf{M}$ of S_m^{n+k} into $\log^3$ in a model of S_2^n. To achieve this, we work with the term $\omega_1^{a_1}(\underline{4})$ (where the significance of a_1 is as before). As noted above, this term is roughly of the size 2^{a_1}, and $(2^{a_1})^{\mathbb{N}}$ is contained in $\log \mathbf{M}$, which allows the proof to go through as previously if $m \geq 3$.

In the case of $m = 2$ (which will not actually be needed for the proof of Theorem 5.1), there is an additional technicality. Here, the cut $\log$ is in general only closed under multiplication. Thus, it might happen that $(2^{a_1})^{\mathbb{N}}$ is cofinal in $\log \mathbf{M}$, leaving no room to find some $\log \mathbf{M} \ni b > (2^{a_1})^{\mathbb{N}}$ and an evaluation p over $[0, b]$ which could be used to define $\mathbf{M}(p, 2^{a_1})$. Fortunately, there is an easy way out: a standard compactness argument shows that we may consider only models in which $\log$ is closed under taking some nonstandard power. If $\mathbf{M}$ is such a model, then obviously $(2^{a_1})^{\mathbb{N}}$ can no longer be cofinal in $\log \mathbf{M}$. $\square$

The modified version of Lemma 4.4 is:

Lemma 5.3. *The family of theories $\{S_m^n\}_{n,m \geq 1}$ has the following properties:*

(i) *Let $m \geq 2, n \geq 1, k \geq 0$, and assume $S_m^{n+k} \vdash \mathrm{HCons}^*(S_2^{n,*})$. Then for any Δ_0 formula $\theta(\bar{x})$ such that $S_m^{n+k} + \exists \bar{x} \in \log^2 \theta(\bar{x})$ is consistent, $S_2^n + \exists \bar{x} \in \omega_{m-3}(\log^3) \theta(\bar{x})$ is also consistent.*

17

(ii) *Let $m \geq 2, n \geq 1, k, l \geq 0$, and assume that $S_m^{n+k+l} \vdash \mathrm{HCons}^*(S_2^{n+k,*})$ and $S_m^{n+k} \vdash \mathrm{HCons}^*(S_2^{n,*})$. Then for any Δ_0 formula $\theta(\bar{x})$ such that $S_m^{n+k+l} + \exists \bar{x} \in \log^2 \theta(\bar{x})$ is consistent, $S_m^n + \exists \bar{x} \in \log^3 \theta(\bar{x})$ is also consistent.*

Proof idea. Little changes in comparison with the proof of Lemma 4.4, except that we now use Lemma 5.2 instead of Lemma 4.3. For this reason, each $\log^r$ appearing in the proof of Lemma 4.4 now becomes $\log^{r-1}$, ω_{m-4} becomes ω_{m-3}, and ω_{m-5} becomes ω_{m-4} (so there is no need to define ω_{-3}). $\qquad \square$

Given Lemma 5.3(ii), we conclude the proof of Theorem 5.1 just like the proof of the main theorem in section 4. Of course, the statement ($\circ$) has to be suitably modified.

Remark. Note that instead of expanding the language by adding ω_1, we could have extended our theories by adding the seemingly irrelevant axiom

$$\forall x \, \exists y \, y = x \# x.$$

Indeed, this axiom gives rise to a unary Herbrand function[7] which behaves as ω_1, so for some n, $\bigcup_m S_m$ does not prove $\mathrm{HCons}^*(S_2^n + \forall x \, \exists y \, y = x \# x)$. Thus, the notion of Herbrand consistency, or at least our way of proving its unprovability, seems to be somewhat less robust than one could wish.

6 Concluding remarks

We conclude the paper with two further remarks. One of these relates our results to an important open problem in the area of weak arithmetics, while the other concerns the notion of cut-free consistency.

I. There is a famous open problem whether $I\Delta_0 + \Omega_k$ is Π_1-conservative over $I\Delta_0 + \Omega_m$ for $k > m$. Our results indicate that the generally expected negative answer will not be shown by proving

$$I\Delta_0 + \Omega_k \vdash \mathrm{HCons}(I\Delta_0 + \Omega_m),$$

---

[7] If one defines the Herbrand consistency of a finitely axiomatizable theory T as the non-existence of an Herbrand proof for the prenex normal form of $\neg \bigwedge T$ (note that our approach, described in section 2, is slightly different), then the question whether this Herbrand function is unary or not will depend on the details of how exactly the normal form is built; more precisely, it will depend on the order of quantifiers in the prefix.

at least not if $m \geq 1$.

However, there can still be hope that the study of variants of Herbrand consistency might have some relevance for the Π_1-conservativity problem, if a more careful quantitative analysis is involved. For any cut I, define the statement $\mathrm{HCons}^{\mathrm{dp}(I)}(T)$ to be:

For any set of terms $A \subseteq \log$ closed under subterms and containing only terms whose depth is in I, there exists a T-evaluation over A.

(Note that we are now allowing evaluations over sets which are not intervals of the form $[0, a]$; the depth of an L_T^H-term is defined inductively in the natural way.) $\mathrm{HCons}^{\mathrm{dp}(I)}$ is Π_1 whenever I has a Σ_1 definition, and it is well-known ([Pud85]) that any reasonable finitely axiomatizable T proves $\mathrm{HCons}^{\mathrm{dp}(I)}(T)$ for some definable cut I. One possible way to solve the Π_1-conservativity problem is to gain more insight into the question of exactly how large I can be for various bounded arithmetic theories.

To give a specific example, let us focus on the case of $I\Delta_0 + \Omega_3$ versus $I\Delta_0 + \Omega_2$, or equivalently, S_4 versus S_3. We know from Theorem 4.1 that for some n, S_3 does not prove $\mathrm{HCons}^*(S_3^n)$. On the other hand, it can be shown that for any n and any fixed standard r, S_4 does prove

$$\mathrm{HCons}^{\mathrm{dp}(r \log^4)}(S_3^n)$$

([Ada][8]). So, if Theorem 4.1 for $m = 3$ could be improved to the result that for some n and r,

$$S_3 \nvdash \mathrm{HCons}^{\mathrm{dp}(r \log^4)}(S_3^n),$$

it would follow that $I\Delta_0 + \Omega_3$ is not Π_1 conservative over $I\Delta_0 + \Omega_2$.

In this context, it is quite interesting that the methods of [AZ01] can be refined to show that for each sufficiently large n, $S_3^n \nvdash \mathrm{HCons}^{\mathrm{dp}(3 \log^4)}(S_3^n)$. So, the question to ask is whether the unprovability of $\mathrm{HCons}(S_3^n)$, for large n, is as "deep" in S_3 as it is in S_3^n. A positive answer would give a partial solution to the Π_1-conservativity problem, while a negative answer would separate S_3 from its finitely axiomatizable fragments.

---

[8]The basic idea of the argument is that in a model $\mathbf{M}$ in which ω_3 is total, ω_2 can be iterated at least $\log^4$ times. Therefore, an evaluation on terms of depth from $\log^4$ can be defined by interpreting them as elements of $\mathbf{M}$. This idea can be elaborated to show that S_4 proves $\mathrm{HCons}^{\mathrm{dp}(\log^4)}(S_3^n)$ for every n. Finally, Ω_3 implies that $\log^4$ is closed under addition.

II. There are many notions of consistency similar in strength to Herbrand consistency. Most widely known among these is probably the concept of cut-free consistency (CFCons), i.e. consistency relative to cut-free proofs in the sequent calculus. Hence, it is natural to ask whether our theorems can be translated into results about the unprovability of CFCons.

Herbrand consistency and cut-free consistency are known to be equivalent in $I\Delta_0+\text{Exp}$. Moreover, analysis of the proof of this equivalence (as presented in [HP93]) reveals that already in S_2, any formula φ which has an Herbrand proof contained in log also has a cut-free proof.

We have shown that for some n, models of $\bigcup_m S_m$ may contain Herbrand proofs of the inconsistency of S_3^n. Unfortunately, these Herbrand proofs are comparable in size to evaluations over an interval $[0, b]$, where b is larger than 2^{2^a} for $a \in \log^3$. Such evaluations are in general too large to fit inside log. Thus, our reasoning does not suffice to conclude the unprovability of $\text{CFCons}(S_3^n)$ in $\bigcup_m S_m$.

The situation changes once we move to fragments of S_4. A proof that $\bigcup_m S_m$ does not prove $\text{HCons}(S_4^n)$ for some n can be based on moving witnesses for Δ_0 formulae not from $\log^3$ to $\log^4$, but from $\log^4$ to $\log^5$. The evaluations needed to do this only have to cover an interval of the form $[0, b]$, where $b > (2^{2^a})^{\mathbb{N}}$ for some $a \in \log^4$ (cf. the proof of Lemma 4.2 for $m = 4$). By a compactness argument, b itself can be pushed into $\log^2$, which means that evaluations over $[0, b]$ will be in log. Using this, one can show that a model of $\bigcup_m S_m$ may contain an Herbrand proof of the inconsistency of S_4^n inside log. Consequently, a model of $\bigcup_m S_m$ may contain a cut-free proof of the inconsistency of S_4^n.

We have thus sketched a justification of the following theorem:

Theorem 6.1. *There exists n such that $\bigcup_m S_m \nvdash \text{CFCons}(S_4^n)$.*

Although the general algorithm for constructing a cut-free proof from an Herbrand proof results in an exponential blow-up in size, Herbrand proofs typically tend to be larger than cut-free proofs. For this reason, we conjecture that Theorem 6.1 may be improved to (at least) the unprovability of $\text{CFCons}(S_2^n)$ for some n in $\bigcup_m S_m$.

[Wil02], and the anonymous referee for comments which allowed me to improve the structure of the paper and notice a small error in one of the proofs in the original draft version.

References

[Ada]   Z. Adamowicz, private communication.

[Ada01] _____, *On tableaux consistency in weak theories*, Preprint 618, Institute of Mathematics of the Polish Academy of Sciences, 2001.

[Ada02] _____, *Herbrand consistency and bounded arithmetic*, Fundamenta Mathematicae **171** (2002), 279–292.

[AZ01]   Z. Adamowicz and P. Zbierski, *On Herbrand consistency in weak arithmetic*, Archive for Mathematical Logic **40** (2001), 399–413.

[BI95]   S. R. Buss and A. Ignjatović, *Unprovability of consistency statements in fragments of bounded arithmetic*, Annals of Pure and Applied Logic **74** (1995), 221–244.

[HP93]   P. Hájek and P. Pudlák, *Metamathematics of first-order arithmetic*, Springer-Verlag, 1993.

[Kra95] J. Krajíček, *Bounded arithmetic, propositional logic, and complexity theory*, Cambridge University Press, 1995.

[Pud]   P. Pudlák, *Consistency and games*, to appear in Proceedings of Logic Colloquium 2003.

[Pud85] _____, *Cuts, consistency statements, and interpretations*, Journal of Symbolic Logic **50** (1985), 423–441.

[Pud90] _____, *A note on bounded arithmetic*, Fundamenta Mathematicae **136** (1990), 85–89.

[Wil02] D. E. Willard, *How to extend the semantic tableaux and cut-free versions of the second incompleteness theorem almost to Robinson's arithmetic Q*, Journal of Symbolic Logic **67** (2002), 465–496.

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

[WP89] _____, *On the existence of end-extensions of models of bounded induction*, Logic, Methodology, and Philosophy of Science VIII (Moscow 1987) (J.E. Fenstad, I.T. Frolov, and R. Hilpinen, eds.), North-Holland, 1989, pp. 143–161.