

Truth definitions without exponentiation and the Σ_1 collection scheme

Zofia Adamowicz * Leszek Aleksander Kołodziejczyk[†]
Jeff Paris[‡]

December 16, 2010

Abstract

We prove that:

- If there is a model of $I\Delta_0 + \neg exp$ with cofinal Σ_1 -definable elements and a Σ_1 truth definition for Σ_1 sentences, then $I\Delta_0 + \neg exp + \neg B\Sigma_1$ is consistent,
- there is a model of $I\Delta_0 + \Omega_1 + \neg exp$ with cofinal Σ_1 -definable elements, both a Σ_2 and a Π_2 truth definition for Σ_1 sentences, and for each $n \geq 2$, a Σ_n truth definition for Σ_n sentences.

The latter result is obtained by constructing a model with a recursive truth-preserving translation of Σ_1 sentences into boolean combinations of $\exists\Sigma_0^b$ sentences.

We also present an old but previously unpublished proof of the consistency of $I\Delta_0 + \neg exp + \neg B\Sigma_1$ under the assumption that the size parameter in Lessan's Δ_0 universal formula is optimal. We then discuss a possible reason why proving the consistency of $I\Delta_0 + \neg exp + \neg B\Sigma_1$ unconditionally has turned out to be so difficult.

*Mathematical Institute of the Polish Academy of Sciences, Śniadeckich 8, 00-956 Warszawa, Poland, zosiaa@impan.gov.pl. Partially supported by grant N N201 382234 of the Polish Ministry of Science and Higher Education.

[†]Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland, lak@mimuw.edu.pl. Partially supported by grant N N201 382234 of the Polish Ministry of Science and Higher Education.

[‡]School of Mathematics, University of Manchester, Manchester M13 9PL, U.K., jeff.paris@manchester.ac.uk.

Consider the following two problems concerning weak theories of arithmetic:

- (1) Is the failure of the collection scheme $B\Sigma_1$ consistent with $I\Delta_0 + \neg exp$?
- (2) What kinds of truth definition/universal formula can there be in models of $I\Delta_0 + \neg exp$?

Problem (1) is a well-defined open question, pointed out in [WP89] and apparently quite difficult. (2) represents a whole family of questions, some also apparently hard. For example, it follows from the work of [Bus95] and [Zam96] that the question whether there is a universal Σ_1 formula which works in all models of $I\Delta_0 + \Omega_1$ is exactly equivalent to the notorious open problem whether the bounded arithmetic hierarchy collapses. However, it seems difficult to find even a single model of $I\Delta_0 + \neg exp$ with a universal Σ_1 formula.

It is well-known that problems (1) and (2) are connected: if there is a model of $I\Delta_0 + \neg exp$ with a universal Σ_1 formula, then there also is a model of $I\Delta_0 + \neg B\Sigma_1 + \neg exp$.

In this note, we prove some new results about (1) and (2). We first observe (Theorem 1.1) that to build models of $\neg exp + \neg B\Sigma_1$, it is not necessary to have a full-blown universal Σ_1 formula: a model of $\neg exp$ with a Σ_1 truth definition for Σ_1 sentences (i.e. with no parameters) is enough. To rule out trivial counterexamples, we need the model to have cofinal Σ_1 -definable elements; otherwise, it could share its Σ_1 theory with a model of exp .

This result immediately raises the question whether a model with a Σ_1 truth definition for Σ_1 sentences exists. We find a model with “almost” the right property: it has both a Σ_2 and a Π_2 truth definition for Σ_1 sentences, and moreover, a Σ_2 truth definition for Σ_2 sentences, Σ_3 truth definition for Σ_3 , etc. (Theorem 2.2). All these truth definitions are built using a recursive translation of Σ_1 sentences into boolean combinations of Σ_1 sentences with limited quantifier alternation in the bounded part (Theorem 2.1).

Unfortunately, the task of finding a model with a Σ_1 truth definition for Σ_1 sentences, and thus settling (1), remains beyond our reach. In the final part of the note (Section 3), we prove a positive answer to (1) from a very different kind of assumption than the existence of a Σ_1 truth definition, and offer a possible explanation why getting an unconditional answer seems so difficult.

We assume that the reader is familiar with basic notions and results concerning fragments of Peano Arithmetic and bounded arithmetic (all relevant information can be found in [HP93]).

When working in $I\Delta_0 + \Omega_1$, we freely speak about Buss' formula classes Σ_m^b and Π_m^b , whose definition formally requires an extended language. Recall that for each m there is a universal $\exists\Pi_m^b$ formula, which we denote $Sat_{\exists\Pi_m^b}$. Recall also that every polynomial-time property can be defined by both a Σ_1^b and a Π_1^b formula. In a strong enough language (one that contains terms related to sequence coding), a Σ_1^b formula is in particular an $\exists\Sigma_0^b$ formula.

We fix some feasible coding of sequences and syntax; $(x)_i$ is the i -th element of sequence x , \underline{x} is the numeral for x . For a model \mathcal{A} and $X \subseteq \mathcal{A}$, $\mathcal{K}_1(\mathcal{A}, X)$ denotes the submodel consisting of elements which are Σ_1 -definable using parameters from X . For $a \in \mathcal{A}$, $\omega_1^{\mathbb{N}}(a)$ denotes the cut in \mathcal{A} determined by the standard iterations of the ω_1 function applied to a .

1 How to use a Σ_1 truth definition

Theorem 1.1. *Assume that there is a model $\mathcal{A} \models I\Delta_0 + \neg exp$ in which Σ_1 definable elements are cofinal and in which there is a Σ_1 truth definition for Σ_1 sentences. Then there exists a model of $I\Delta_0 + \neg exp + \neg B\Sigma_1$.*

Proof. The proof combines the classical construction of a model of $\neg B\Sigma_1$ from [KP78] with a simple compactness argument. Let \mathcal{A} be as above and let $\text{Tr}(x)$ be the Σ_1 truth definition for Σ_1 sentences in \mathcal{A} . Since Σ_1 -definable elements are cofinal in \mathcal{A} , there is a Δ_0 formula $\eta(x)$ such that

$$\mathcal{A} \models \exists x \eta(x) \ \& \ \forall x (\eta(x) \Rightarrow \neg \exists y y = 2^x).$$

Add a new constant c to the language of arithmetic and consider the following theory T :

$$\begin{aligned} & I\Delta_0 + \exists x \eta(x) + \forall x (\eta(x) \Rightarrow \neg \exists y y = 2^x) \\ & + \{\forall x \leq c (\varphi(x, c) \Leftrightarrow \text{Tr}(\ulcorner \varphi(\underline{x}, \underline{c}) \urcorner)) : \varphi \in \Sigma_1\} \\ & + \{c > \underline{n} : n \in \mathbb{N}\}. \end{aligned}$$

Every finite fragment of T can be satisfied in \mathcal{A} , so by compactness, T is consistent. Take $\mathcal{B} \models T$. Since $\mathcal{K}_1(\mathcal{B}, \{c\}) \preceq_{\Sigma_1} \mathcal{B}$ and T is Π_2 -axiomatizable, we can assume that $\mathcal{B} = \mathcal{K}_1(\mathcal{B}, \{c\})$.

Now we essentially repeat the argument of [KP78]. Each element of \mathcal{B} satisfies some *syntactic* Σ_1 definition with parameter c , i.e. a formula of the form

$$\exists y (x = (y)_1 \ \& \ \psi(y, c) \ \& \ \forall z < y \neg\psi(z, c)),$$

where $\psi \in \Delta_0$. Obviously, no two elements satisfy the same Σ_1 definition. We have:

$$\mathcal{B} \models \forall x \leq c \exists \ulcorner \varphi \urcorner < \log c \ (\varphi \text{ is a } \Sigma_1 \text{ definition} \ \& \ \text{Tr}(\ulcorner \varphi(\underline{x}, \underline{c}) \urcorner)).$$

The formula “ φ is a Σ_1 definition & $\text{Tr}(\ulcorner \varphi(\underline{x}, \underline{c}) \urcorner)$ ” is a Σ_1 formula with just one unbounded existential quantifier, the one in $\text{Tr}(\ulcorner \varphi(\underline{x}, \underline{c}) \urcorner)$. If we had $\mathcal{B} \models B\Sigma_1$, we could bound this quantifier uniformly in $x \leq c$, which would give a Δ_0 definable injection from $[0, c]$ into $[0, \log c]$. Such injections are impossible in $I\Delta_0$, hence $\mathcal{B} \models \neg B\Sigma_1$. \square

A Σ_1 truth definition for Σ_1 sentences immediately gives a truth-value preserving translation of Σ_1 to $\exists\Pi_m^b$ sentences for some m : map Φ to $\text{Tr}(\ulcorner \Phi \urcorner)$. Actually, the existence of such a translation is also a sufficient condition for a Σ_1 truth definition:

Corollary 1.2. *Assume that there is a model $\mathcal{A} \models I\Delta_0 + \Omega_1 + \neg exp$ in which Σ_1 -definable elements are cofinal and for which there is a recursive translation $\Phi \mapsto \varphi$ of Σ_1 sentences into $\exists\Pi_m^b$ sentences such that $\mathcal{A} \models \Phi \Leftrightarrow \varphi$. Then there exists a model of $I\Delta_0 + \neg exp + \neg B\Sigma_1$.*

Proof. Let f be a recursive translation of Σ_1 sentences into $\exists\Pi_m^b$ sentences preserving truth values in \mathcal{A} . There is a Σ_1 truth definition for Σ_1 sentences in \mathcal{A} given by:

$$\text{Tr}(x) \Leftrightarrow \exists y (y = f(x) \ \& \ \text{Sat}_{\exists\Pi_m^b}(y)).$$

Now apply Theorem 1.1. \square

2 How to almost get a Σ_1 truth definition

The results of the previous section suggest looking for a model with a translation of Σ_1 sentences into $\exists\Pi_m^b$ sentences as a way of proving the consistency of $I\Delta_0 + \neg exp + \neg B\Sigma_1$.

Below we describe a translation yielding not $\exists\Pi_m^b$ sentences but their boolean combinations (for $m = 0$). The construction has much in common with an argument from [AKZ03] which produced models with Σ_1 truth Turing-reducible to $\exists\Pi_m^b$ truth; however, the technique is simpler and the result in some ways stronger. The translation also gives interesting information about truth definitions, though not quite what we need to get a model of $\neg exp + \neg B\Sigma_1$.

Theorem 2.1. *There is a model $\mathcal{A} \models I\Delta_0 + \Omega_1 + \neg exp$ in which Σ_1 -definable elements are cofinal and for which there is a recursive translation $\Phi \mapsto \varphi$ of Σ_1 sentences into $\text{bool}(\exists\Sigma_0^b)$ sentences such that $\mathcal{A} \models \Phi \Leftrightarrow \varphi$.*

Proof. Fix a recursive numbering $(\Phi_n)_{n \in \mathbb{N}}$ of Σ_1 sentences. Let $\eta(x)$ be a Σ_0^b formula for which $\exists x \eta(x)$ is false in \mathbb{N} , consistent with $I\Delta_0 + \Omega_1$ and never witnessed by more than one element (e.g., x is a power of 2 and $\log x$ is the smallest PA -proof of inconsistency).

Writing $\Phi^{(0)}$ for Φ and $\Phi^{(1)}$ for $\neg\Phi$, we define $\varepsilon \in \{0, 1\}^{\mathbb{N}}$ inductively by:

$$\varepsilon(n) = \begin{cases} 0 & \text{if } I\Delta_0 + \Omega_1 + \exists x \eta(x) + \Phi_0^{(\varepsilon(0))}, \dots, \Phi_{n-1}^{(\varepsilon(n-1))} \vdash \Phi_n, \\ 1 & \text{otherwise.} \end{cases}$$

For each $\gamma \in \{0, 1\}^{\mathbb{N}}$, let T^γ be the theory

$$I\Delta_0 + \Omega_1 + \exists x \eta(x) + \{\Phi_n^{(\gamma(n))} : n \in \mathbb{N}\}$$

and let T_n^γ be

$$I\Delta_0 + \Omega_1 + \exists x \eta(x) + \{\Phi_k^{(\gamma(k))} : k < n\}.$$

T^ε , which roughly states that “as many Π_1 sentences are true as possible”, is clearly consistent. So, consider some nonstandard model of T^ε , let a be the witness for $\exists x \eta(x)$ and let \mathcal{A} be the cut $\omega_1^{\mathbb{N}}(a)$. It is easy to show by induction on n that $\mathcal{A} \models \Phi_n^{(\varepsilon(n))}$, and hence $\mathcal{A} \models T^\varepsilon$. It follows that in \mathcal{A} , Φ_n is true exactly if it is provable from T_n^ε .

Since \mathcal{A} is of the form $\omega_1^{\mathbb{N}}(a)$ for a satisfying $\eta(x)$, \mathbb{N} is $\exists\Sigma_0^b$ -definable in \mathcal{A} . Consequently, for a recursive theory S , we can write the sentence

$$\forall x (x \in \mathbb{N} \Rightarrow x \text{ is not an } S\text{-proof of inconsistency}),$$

which we denote $Con^{\mathbb{N}}(S)$, and which holds in \mathcal{A} exactly if S is in fact consistent. This means that a Σ_1 sentence Φ_n is equivalent in \mathcal{A} to:

$$\bigvee_{\gamma \in \{0,1\}^n} \left(\bigwedge_{k < n, \gamma(k)=0} \neg Con^{\mathbb{N}}(T_k^\gamma + \neg \Phi_k) \right. \\ \left. \& \bigwedge_{k < n, \gamma(k)=1} Con^{\mathbb{N}}(T_k^\gamma + \neg \Phi_k) \& \neg Con^{\mathbb{N}}(T_n^\gamma + \neg \Phi_n) \right),$$

a boolean combination of $\exists \Sigma_0^b$ sentences. \square

Theorem 2.2. *There is a model $\mathcal{A} \models I\Delta_0 + \Omega_1 + \neg exp$ in which Σ_1 -definable elements are cofinal and in which there is:*

- (a) both a Σ_2 and a Π_2 truth definition for Σ_1 sentences,
- (b) for $n \geq 2$, a Σ_n truth definition for Σ_n sentences.

Proof. Let \mathcal{A} satisfy the thesis of Theorem 2.1. Let f be a recursive translation of Σ_1 sentences into $\text{bool}(\exists \Sigma_0^b)$ sentences preserving truth values in \mathcal{A} . Note that, using $\text{Sat}_{\exists \Sigma_0^b}$, it is easy to build both a Σ_2 and a Π_2 universal formula for $\text{bool}(\exists \Sigma_0^b)$. Call these universal formulae $\text{Sat}_{\text{bool}(\exists \Sigma_0^b)}^\Sigma$ and $\text{Sat}_{\text{bool}(\exists \Sigma_0^b)}^\Pi$, respectively. We can construct the following truth definitions for Σ_1 sentences:

$$\text{Tr}_{\Sigma_1}^\Sigma(x) \Leftrightarrow \exists y \left(y = f(x) \& \text{Sat}_{\text{bool}(\exists \Sigma_0^b)}^\Sigma(y) \right), \\ \text{Tr}_{\Sigma_1}^\Pi(x) \Leftrightarrow \forall y \left(y = f(x) \Rightarrow \text{Sat}_{\text{bool}(\exists \Sigma_0^b)}^\Pi(y) \right).$$

$\text{Tr}_{\Sigma_1}^\Sigma$ is Σ_2 and $\text{Tr}_{\Sigma_1}^\Pi(x)$ is Π_2 , so this completes the proof of (a).

To prove (b), we need to assume additionally that $\mathcal{A} = \mathcal{K}_1(\mathcal{A})$ and that \mathbb{N} is Σ_1 -definable in \mathcal{A} (we can make the former assumption by Σ_1 -elementarity and the latter by the proof of Theorem 2.1). We now explain how to build a Σ_2 truth definition for Σ_2 sentences in \mathcal{A} . The case of $n > 2$ is proved by induction on n in a similar way.

Consider a Σ_2 sentence of the form $\xi := \exists y \forall z \psi(y, z)$ with ψ bounded. This sentence is true in \mathcal{A} exactly if

$$\text{for some } \Sigma_1\text{-definable } a, \mathcal{A} \models \forall z \psi(a, z).$$

For each Σ_1 -definable a , there is a Σ_1 formula $\varphi(y)$ satisfied only by a . Hence, ξ is true in \mathcal{A} exactly if for some Σ_1 formula $\varphi(y)$, both

$$\mathcal{A} \models \exists y \varphi(y)$$

and

$$\mathcal{A} \models \forall y \forall z (\varphi(y) \Rightarrow \psi(y, z)).$$

This means that a Σ_2 truth definition for Σ_2 sentences can be written as:

$$\begin{aligned} \text{Tr}_{\Sigma_2}(x) \quad \Leftrightarrow \quad & \exists \ulcorner \varphi \urcorner \exists \ulcorner \psi \urcorner (x = \ulcorner \exists y \forall z \psi(y, z) \urcorner \& \ulcorner \varphi \urcorner \in \mathbb{N} \\ & \& \text{Tr}_{\Sigma_1}^{\Sigma}(\ulcorner \exists y \varphi(y) \urcorner) \\ & \& \neg \text{Tr}_{\Sigma_1}^{\Pi}(\ulcorner \neg \forall y \forall z (\varphi(y) \Rightarrow \psi(y, z)) \urcorner). \end{aligned}$$

□

3 Why is avoiding collection so hard?

In this section, we try to come up with an explanation (complexity-theoretic in spirit) why finding a model of $I\Delta_0 + \neg exp + \neg B\Sigma_1$ has turned out to be so difficult. This leads us to formulate a conjecture which follows from the consistency of $\neg exp + \neg B\Sigma_1$, but seems difficult to prove by current means. We begin, however, by presenting an old but previously unpublished result related to our questions (1) and (2).

It is well known that $\neg B\Sigma_1$ is consistent with $\neg exp$ assuming the collapse of the polynomial hierarchy (PH \downarrow). It turns out that the same consistency result can be proved under a very different kind of assumption. LIO (Lessan's bound Is Optimal) is the following statement:

If $\mathcal{A} \models I\Delta_0$, $\Upsilon(x, y, z)$ is a Δ_0 formula, $a \in \mathcal{A}$ and $n \in \mathbb{N}$,
then there is a Δ_0 formula $\varphi(x)$ such that $\mathcal{A} \not\models \varphi(a) \Leftrightarrow \Upsilon(\ulcorner \varphi \urcorner, a, 2^{a^n})$

In other words, LIO says that the size of the parameter in a “ Δ_0 universal formula with a parameter for elements up to a ” cannot be smaller than the $2^{a^{\mathbb{N}}}$ bound given in [Les78].

Theorem 3.1. *If LIO holds, then $I\Delta_0 + \neg exp + \neg B\Sigma_1$ is consistent.*

Proof. Choose some $\mathbb{N} < a \in \mathcal{A} \models I\Delta_0 + \Omega_1$ such that $\mathcal{A} = \omega_1^{\mathbb{N}}(a)$ and $\mathcal{A} = \mathcal{K}_1(\mathcal{A}, \{a, 2^{\log a}, 2^{\log^2 a}, \dots\})$. Under these assumptions, each element of \mathcal{A} is actually Δ_0 -definable from $\{a, 2^{\log a}, 2^{\log^2 a}, \dots\}$. LIO thus implies that the set

$$\{\ulcorner \varphi \urcorner : \varphi \in \Delta_0 \ \& \ \mathcal{A} \models \varphi(\log a)\}$$

is not coded in \mathcal{A} .

Let Sat be the usual Δ_0 universal formula with a parameter (i.e. Lessan's). For each Δ_0 formula φ , there is a natural number $n(\varphi)$ computable from φ such that

$$\mathcal{A} \models \varphi(\log a) \Leftrightarrow \text{Sat}(\ulcorner \varphi \urcorner, \log a, 2^{\log^{n(\varphi)} a}).$$

Therefore, in \mathcal{A} we have:

$$\begin{aligned} \forall x \leq a \exists \ulcorner \varphi \urcorner \exists b \exists n \\ (\varphi \in \Delta_0 \ \& \ n = n(\varphi) \ \& \ b = 2^{\log^n a} \ \& \ \neg(\ulcorner \varphi \urcorner \in x \Leftrightarrow \text{Sat}(\ulcorner \varphi \urcorner, \log a, b))) \end{aligned}$$

However, the existential quantifier $\exists b$ cannot be bounded, so $\mathcal{A} \models \neg B\Sigma_1$. \square

$\text{PH}\downarrow$ and LIO represent strongly divergent “complexity-theoretic worldviews”; in particular, they contradict each other. One might therefore hope that a consistency statement known to follow from two completely different assumptions on the complexity-theoretic world could be shown to hold without any assumptions. The problem, however, is that both $\text{PH}\downarrow$ and LIO contradict a third “worldview”, which, though not too plausible, seems difficult to disprove using known methods. Moreover, it may well be incompatible with the consistency of $\neg \text{exp} + \neg B\Sigma_1$.

AHL (“additional quantifier Alternations always Help a Lot”) says roughly that a problem which can be solved by a Σ_m machine can always be solved in less time by a Σ_{m+k} machine for some k . A very strong version of AHL might say: “for every time-constructible f such that all iterations of f are dominated by exp , $\Sigma_m\text{-TIME}(f) \subseteq \Sigma_{m+1}^{\text{poly}}$ ”.

$\text{PH}\downarrow$ obviously contradicts most formulations of AHL, including the one given above: if $\Sigma_m^{\text{poly}} = \Sigma_{m+1}^{\text{poly}}$, then e.g. $\Sigma_m\text{-TIME}(n^{\log n})$ is not in $\Sigma_{m+k}^{\text{poly}}$ for any k . LIO contradicts AHL because it implies that the problem “given a finite structure \mathcal{A} with universe $\{0, \dots, a\}$ and a sentence Φ in a relational language with $\text{lh}(\Phi) \leq \log a$, does $\mathcal{A} \models \Phi$?”, solvable in deterministic time $a^{O(\log a)}$, cannot be solved in $\Sigma_k\text{-TIME}(\text{poly}(a))$ for any k .

The model-theoretic way of comparing “more of a weaker resource” with “less of a stronger resource” seems to involve end-extensions: cf. Nepomnjaščij’s theorem that **LOGSPACE** is contained in the linear-time hierarchy and the result proved in [Zam97] that every model of $I\Delta_0$ (which corresponds to the linear time hierarchy) can be end-extended to a model of a theory for **LOGSPACE**. Thus, we are led to the following conjecture, which is a model-theoretic formulation of “AHL is false”:

Conjecture 3.2. There exists a model $\mathcal{A} \models I\Delta_0 + \neg exp$ such that for some m , \mathcal{A} has no proper end-extension to a model of T_2^m .

Note that our conjecture is implied by the consistency of $I\Delta_0 + \neg exp + \neg B\Sigma_1$. Note also, however, that it is not at all clear how to attack the conjecture. Various methods used to construct models of $I\Delta_0 + \neg exp$ without proper end-extensions to models of $I\Delta_0$ do not seem to work here.

References

- [AKZ03] Z. Adamowicz, L. A. Kołodziejczyk, and P. Zbierski, *An application of a reflection principle*, *Fundamenta Mathematicae* **180** (2003), 139–159.
- [Bus95] S. R. Buss, *Relating the bounded arithmetic and polynomial time hierarchies*, *Annals of Pure and Applied Logic* **75** (1995), 67–77.
- [HP93] P. Hájek and P. Pudlák, *Metamathematics of First-Order Arithmetic*, Springer-Verlag, 1993.
- [KP78] L. A. S. Kirby and J. B. Paris, σ_n *collection schemas in arithmetic*, *Logic Colloquium 77*, *Studies in Logic and the Foundations of Mathematics*, vol. 96, North-Holland, 1978, pp. 199–209.
- [Les78] H. Lessan, *Models of arithmetic*, Ph.D. thesis, University of Manchester, 1978.
- [WP89] A. J. Wilkie and J. B. Paris, *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.

- [Zam96] D. Zambella, *Notes on polynomially bounded arithmetic*, Journal of Symbolic Logic **61** (1996), 942–966.
- [Zam97] ———, *End extensions of models of polynomially bounded arithmetic*, Annals of Pure and Applied Logic **88** (1997), 263–277.