The Journal of Symbolic Logic
Volume 81, Number 3, September 2016

END-EXTENSIONS OF MODELS OF WEAK ARITHMETIC FROM COMPLEXITY-THEORETIC CONTAINMENTS

LESZEK ALEKSANDER KOŁODZIEJCZYK

Abstract. We prove that if the linear-time and polynomial-time hierarchies coincide, then every model of $\Pi_1(\mathbb{N}) + \neg\Omega_1$ has a proper end-extension to a model of $\Pi_1(\mathbb{N})$, and so $\Pi_1(\mathbb{N}) + \neg\Omega_1 \vdash \mathrm{B}\Sigma_1$. Under an even stronger complexity-theoretic assumption which nevertheless seems hard to disprove using present-day methods, $\Pi_1(\mathbb{N}) + \neg\mathrm{Exp} \vdash \mathrm{B}\Sigma_1$. Both assumptions can be modified to versions which make it possible to replace $\Pi_1(\mathbb{N})$ by $\mathrm{I}\Delta_0$ as the base theory.

We also show that any proof that $\mathrm{I}\Delta_0 + \neg\mathrm{Exp}$ does not prove a given finite fragment of $\mathrm{B}\Sigma_1$ has to be "nonrelativizing", in the sense that it will not work in the presence of an arbitrary oracle.

The work presented below aims at a better understanding of the following notoriously hard open problem about moderately weak theories of arithmetic:

$$\text{Does } \mathrm{I}\Delta_0 + \neg\mathrm{Exp} \text{ prove } \mathrm{B}\Sigma_1? \qquad (\star)$$

Here $\mathrm{I}\Delta_0$ is induction for bounded formulas in the language of ordered rings, Exp is the axiom $\forall x\, \exists y\, (y = 2^x)$ with $y = 2^x$ expressed by an appropriate Δ_0 formula, and $\mathrm{B}\Sigma_1$ is the Σ_1 collection scheme,

$$\forall x \leq u\, \exists y\, \varphi(x, y) \Rightarrow \exists w\, \forall x \leq u\, \exists y \leq w\, \varphi(x, y),$$

where φ is Δ_0 and may contain parameters.

It is well-known and easy to show that $\mathrm{I}\Delta_0 \nvdash \mathrm{B}\Sigma_1$ [19], but all known proofs of this (e.g., [1, 5, 19, 20]) use objects of at least exponential size, often in the form of a definition of satisfaction for Σ_1 formulas.

Problem $(\star)$ was first explicitly stated in [23]. The expected answer is negative, and there have been a number of results of the form "the answer to $(\star)$ is negative under some complexity-theoretic assumptions" (e.g., [2, 9]). However, even though the assumptions used to obtain a negative answer have been varied in spirit and in some cases mutually contradictory, an unconditional negative answer remains elusive.

Below, we try to explain part of the difficulty with $(\star)$ by pointing out a complexity-theoretic assumption which implies that the answer to the question "does $\neg$Exp prove $\mathrm{B}\Sigma_1$?" is actually *positive*, over Π_1 truth if the assumption is true and over plain $\mathrm{I}\Delta_0$ if the assumption is provable in $\mathrm{I}\Delta_0$ (which has to be stated in a special way). The assumption is not at all likely to be true; what seems much more likely, however, is that disproving it might be beyond the reach of present-day methods of complexity theory.

Received December 12, 2014.

Key words and phrases. collection scheme, end-extensions, weak arithmetic, time hierarchies, linear time hierarchy, fractional-exponential growth.

© 2016, Association for Symbolic Logic
0022-4812/16/8103-0007
DOI:10.1017/jsl.2015.53

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

Earlier work had already revealed the possibility of complexity-theoretic assumptions implying the provability of fragments of $B\Sigma_1$ from $\neg$Exp, but those fragments have been rather weak. Ferreira [10] showed that some instances of Σ_1 collection are actually unconditionally provable from $\neg$Exp; his instances imply full $B\Sigma_1$ over Exp but appear to be extremely restricted in the presence of $\neg$Exp. The work of [10] is related to the correspondence between the relativized linear-time hierarchy and polynomial-size bounded-depth circuits. Recently, Cordón Franco *et al.* [7] came up with a computational condition exactly equivalent to the provability of *parameter-free* Σ_1 collection in $\Pi_1(\mathbb{N}) + \neg$Exp. Their condition concerns the complexity of search problems, and it is not obvious how it relates to the well-studied complexity hierarchies of decision problems.

Our assumption is just one among many statements of the general form "(a level of) an apparently larger time hierarchy is contained in (a higher level of) an apparently smaller time hierarchy," which tend to imply the provability of $B\Sigma_1$ from an axiom negating the totality of a suitable function f. Before discussing $f(x) = 2^x$, we first consider the simpler case of $f(x)$ equal to $\omega_1(x) = 2^{\log^2 x}$.

The connection between assumptions of this sort and Σ_1 collection involves end-extensions of models. A few specific constructions used to prove containments between computational complexity classes have also led to end-extension theorems [10, 25]. We observe that essentially any containment between levels of typical time hierarchies defined by different time bounds implies the existence of end-extensions for models of corresponding theories. Our focus is on the case where the smaller time bounds are linear, so the corresponding theory is $I\Delta_0$. As is well-known, any model of $I\Delta_0$ with a proper end-extension to another model of $I\Delta_0$ has to satisfy $B\Sigma_1$.

We also apply the technique of building end-extensions from complexity-theoretic containments to show a result about theories in a language expanded by a new "oracle" predicate. Roughly speaking, our theorem says that for every finite fragment of $B\Sigma_1$, a proof that $\neg$Exp does not imply the fragment has to be "nonrelativizing", in the sense that it will not work in the presence of an arbitrary oracle. To show this, we go through a lemma in computational complexity: for every k, there exists an oracle relative to which the k-th level of the exponential-time hierarchy is contained in the linear-time hierarchy.

The paper is structured as follows. We review the necessary definitions and background in Section 1. We prove a very simple version of our "end-extensions from containments" result in Section 2 and the version about $\neg$Exp in Section 3. The result about relativization is proved across Sections 4 and 5. The short Section 6 contains some concluding remarks.

§1. Preliminaries. We assume that the reader is well-acquainted with weak theories of arithmetic as described for instance in [16]. In particular, we assume familiarity with the theories $I\Delta_0$ and $B\Sigma_1$, the axioms Ω_1 and Exp, and with Buss' formula classes Σ^b_k and theories T_2, T^k_2, $k \in \mathbb{N}$. L_{PA} is the language of Peano Arithmetic, or in other words the language of ordered rings. L_{BA} is Buss' language for bounded arithmetic. The class of bounded formulas in L_{BA} is denoted Σ^b_∞; the symbol Δ_0 always stands for the bounded formulas of L_{PA} only.

For linearly ordered structures $\mathcal{M}, \mathcal{K}$ such that $\mathcal{M} \subseteq \mathcal{K}$, the notation $\mathcal{M} \subseteq_e \mathcal{K}$ means that $\mathcal{K}$ is an *end-extension* of $\mathcal{M}$, or in other words, that $\mathcal{M}$ is an initial

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

segment of $\mathcal{K}$. The following simple but important proposition links end-extensions with $B\Sigma_1$:

PROPOSITION 1.1. *If $\mathcal{M}, \mathcal{K}$ are models of $I\Delta_0$ and $\mathcal{M} \subsetneq_e \mathcal{K}$, then $\mathcal{M} \models B\Sigma_1$.*

The *Proof* of Proposition 1.1 (cf. e.g., [13, Lemma IV.1.22]) is as follows: given $a \in M$ and a Δ_0 formula $\varphi(x, y)$ such that $\mathcal{M} \models \forall x < a\, \exists y\, \varphi(x, y)$, we have $\mathcal{K} \models \forall x < a\, \exists y < b\, \varphi(x, y)$ for every $b \in \mathcal{K} \setminus \mathcal{M}$. By overspill, we get the same thing for some $b \in \mathcal{M}$, and thus also $\mathcal{M} \models \forall x < a\, \exists y < b\, \varphi(x, y)$; q.e.d. Note that to guarantee $\mathcal{M} \models B$ for a *finite* theory $B \subseteq B\Sigma_1$, we only need to have $\mathcal{M} \subsetneq_e \mathcal{K}$ for $\mathcal{K}$ satisfying a sufficiently large finite fragment of $I\Delta_0$.

Recall from [17] that PV_{k+1} functions are canonically defined symbols for polynomial-time functions with a Σ^p_k oracle, where Σ^p_k is the k-th level of the polynomial-time hierarchy PH. The set of PV_{k+1} functions contains a symbol for the characteristic function of every Σ^p_k relation and is closed under operations corresponding to composition and bounded recursion on notation. PV_{k+1} also stands for a theory in the language with all PV_{k+1} functions, axiomatized by the defining axioms for the function symbols and induction for all quantifier-free formulas. PV_{k+1} is an extension by definitions of T^k_2 (with some subtleties for $k = 0$, cf. [6, 14]), and can be identified with T^k_2 for all intents and purposes.

Most of the computational complexity theory we use is very old-school, and can be found for instance in [3,4]. Recall that a function f is *time-constructible* if there is a Turing machine which on inputs of length n runs for exactly $f(n)$ steps. Essentially, a time-constructible function is one which makes sense as a time bound. For $k \geq 1$, a Σ_k *machine* is one that begins its computation in an existential (nondeterministic) state, and switches between existential and universal (co-nondeterministic) states at most $k - 1$ times. Thus, for instance, Σ^p_k is Σ_k-TIME$(n^{O(1)})$. The *linear-time hierarchy* LinH is $\bigcup_k \Sigma_k$-TIME$(O(n))$. Analogously to the well-known connection between PH and Σ^b_∞ formulas, a relation on the natural numbers is in LinH exactly if it can be defined by a Δ_0 formula [24].

A *decision tree* is a finite binary tree T with leaves labelled by 0 or 1, with internal nodes labelled by boolean variables x_i, and the two outgoing edges from an internal node labelled by 0 and by 1. We also assume that on any single branch of T each variable appears as a label at most once. The height of T is the length of the longest branch in T. Any boolean assignment α determines a branch through T: from a node labelled by x_i, the branch takes the edge labelled by $\alpha(x_i)$. A decision tree T with internal nodes labelled by variables from $\{x_1, \ldots, x_n\}$ *computes* (or *decides*) a boolean function f on $\{0,1\}^n$ if for every $\langle \alpha(x_1), \ldots, \alpha(x_n) \rangle \in \{0,1\}^n$, $f(\alpha)$ equals the label of the leaf on the branch determined by α.

§2. Simplest case.

THEOREM 2.1. *If* PH = LinH, *then every model $\mathcal{M} \models \Pi_1(\mathbb{N})$ has an end-extension to $\mathcal{K} \models T_2$.*

PROOF. Assume that PH = LinH. This implies that for every Σ^b_∞ formula $\psi(x)$ there exists a Δ_0 formula $\psi^{\text{lin}}(x)$ such that ψ and ψ^{lin} are equivalent in the standard model of arithmetic.

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

Let $\mathcal{M} \models \Pi_1(\mathbb{N})$. $\mathcal{M}$ has a Δ_0-elementary extension to a model $\mathcal{N}$ satisfying the true theory of the natural numbers in L_{BA}. Define $\mathcal{K}$ to be the closure of $\mathcal{M}$ in $\mathcal{N}$ under Skolem functions for Σ^b_∞ formulas, where the Skolem function for $\exists y \leq t\, \varphi(\bar{x}, y)$ is the (Σ^b_∞-definable) function which outputs the smallest $y \leq t$ such that $\varphi(\bar{x}, y)$, or 0 if there is no such y.

Clearly, $\mathcal{K}$ is a Σ^b_∞-elementary substructure of $\mathcal{N}$ and a model of T_2. We need to show that $\mathcal{K} \supseteq_e \mathcal{M}$.

Let g be the Skolem function for a Σ^b_∞ formula $\exists y \leq t\, \varphi(\bar{x}, y)$ and assume that for $\bar{a}, b \in \mathcal{M}$, we have $g^{\mathcal{N}}(\bar{a}) \leq b$. Then $\mathcal{N} \models \exists y \leq b\, (y = g(\bar{a}))$, and thus also

$$\mathcal{N} \models \exists y \leq b\, (y = g(\bar{x}))^{\mathrm{lin}}[\bar{a}/\bar{x}].$$

By Δ_0-elementarity, the same formula holds in $\mathcal{M}$, so there is $c \in \mathcal{M}$ such that $\mathcal{M} \models (y = g(\bar{x}))^{\mathrm{lin}}[\bar{a}/\bar{x}, c/y]$. Thus $\mathcal{N} \models (y = g(\bar{x}))^{\mathrm{lin}}[\bar{a}/\bar{x}, c/y]$ and consequently $\mathcal{N} \models c = g(\bar{a})$, which means that $g^{\mathcal{N}}(\bar{a}) \in \mathcal{M}$. ⊣

If we want to end-extend an arbitrary model of $\mathrm{I}\Delta_0$, rather than just a model of $\Pi_1(\mathbb{N})$, to a structure satisfying T_2, we need a stronger assumption.

THEOREM 2.2. *Assume that there is a translation $\psi \mapsto \psi^{\mathrm{lin}}$ of Σ^b_∞ into Δ_0 formulas such that $T_2 + \{\forall x\, (\psi(x) \Leftrightarrow \psi^{\mathrm{lin}}(x)) : \psi \in \Sigma^b_\infty\}$ is Π_1-conservative over $\mathrm{I}\Delta_0$. Then every model $\mathcal{M} \models \mathrm{I}\Delta_0$ has an end-extension to $\mathcal{K} \models T_2$.*

PROOF. Consider $\mathcal{M} \models \mathrm{I}\Delta_0$. Under the assumption of the theorem, $\mathcal{M}$ has a Δ_0-elementary extension to a model $\mathcal{N}$ satisfying T_2 in which each Σ^b_∞ formula ψ is equivalent to ψ^{lin}. Define $\mathcal{K}$ to be the closure of $\mathcal{M}$ in $\mathcal{N}$ under Skolem functions for Σ^b_∞ formulas and repeat the remainder of the proof of Theorem 2.1. ⊣

COROLLARY 2.3. *If* $\mathrm{PH} = \mathrm{LinH}$, *then* $\Pi_1(\mathbb{N}) + \neg\Omega_1 \vdash \mathrm{B}\Sigma_1$. *If the assumption of Theorem* 2.2 *holds, then* $\mathrm{I}\Delta_0 + \neg\Omega_1 \vdash \mathrm{B}\Sigma_1$.

PROOF. If $\mathcal{M} \models \Pi_1(\mathbb{N}) + \neg\Omega_1$, then the end-extension given by Theorem 2.1 is of necessity proper, which by Proposition 1.1 implies $\mathcal{M} \models \mathrm{B}\Sigma_1$. Similarly for $\mathcal{M} \models \mathrm{I}\Delta_0 + \neg\Omega_1$ and Theorem 2.2. ⊣

It can be argued that the assumption of Theorem 2.2 is a reasonable way of making precise the statement "$\mathrm{I}\Delta_0 \vdash \mathrm{PH} = \mathrm{LinH}$." In essence, what the assumption says is that a model of $\mathrm{I}\Delta_0$ can be extended to another structure in which the LinH properties are preserved, the PH properties are reasonably well-behaved (in other words, T_2 is satisfied) and the PH properties are in fact in LinH.

Any sensible interpretation of "$\mathrm{I}\Delta_0 \vdash \mathrm{PH} = \mathrm{LinH}$" should imply a positive answer to the well-known problem whether T_2 (*vel* $\mathrm{I}\Delta_0 + \Omega_1$) is Π_1-conservative over $\mathrm{I}\Delta_0$, and the assumption of Theorem 2.2 does so quite explicitly. The reverse implication is open and seems worth studying:

OPEN PROBLEM 2.4. *Is it true that if T_2 is Π_1-conservative over $\mathrm{I}\Delta_0$, then in fact the assumption of Theorem* 2.2 *holds*?

§3. The case of ¬Exp. We now consider the task of formulating an assumption which would imply provability of $\mathrm{B}\Sigma_1$ from ¬Exp rather than just from $\neg\Omega_1$. As previously, we first work with $\Pi_1(\mathbb{N})$ as our base theory.

The natural idea would be to replace the polynomial time hierarchy in the statement of Theorem 2.1 by the smaller of the two well-known "exponential

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

time hierarchies," which is defined by letting Σ_k^{e} be Σ_k-TIME$(2^{O(n)})$ and EH be $\bigcup_{k\in\mathbb{N}} \Sigma_k^{\mathrm{e}}$. By an argument similar to the proof of Theorem 2.1, we could show that if EH = LinH, then every $\mathcal{M} \models \Pi_1(\mathbb{N})$ has an end-extension to $\mathcal{K} \models T_2$ such that $\mathcal{M} = \mathrm{Log}\,\mathcal{K}$, where $\mathrm{Log}\,\mathcal{K}$ equals $\{a \in \mathcal{K} : 2^a \text{ exists in } \mathcal{K}\}$. If $\mathcal{M} \models \neg\mathrm{Exp}$, the extension is proper, so $\mathcal{M} \models \mathrm{B}\Sigma_1$.

Unfortunately, the assumption EH = LinH is too strong for our purposes, because it is known to be false. The following proposition is apparently folklore.

PROPOSITION 3.1. EH $\neq$ LinH.

PROOF. For each $k \in \mathbb{N}$, Σ_k-TIME$(O(n)) \subsetneq \Sigma_k^{\mathrm{e}}$ by the nondeterministic time hierarchy theorem. So, if LinH $= \Sigma_k$-TIME$(O(n))$, then LinH $\subsetneq \Sigma_k^{\mathrm{e}}$.

On the other hand, E $=$ DTIME$(2^{O(n)})$ contains LinH and has a problem which is complete via linear-time reductions. Thus, if LinH does not collapse to a finite level, we have LinH $\subsetneq$ E $\subseteq \Sigma_1^{\mathrm{e}}$. ⊣

Proposition 3.1 means that we have to consider time bounds whose growth rate is "fractional-exponential":

DEFINITION 3.2. The function $f : \mathbb{N} \to \mathbb{N}$, $f(n) = O(2^n)$, has *fractional-exponential growth rate* if there exists $\ell \in \mathbb{N}$ such that

$$\underbrace{f \circ f \circ \cdots \circ f}_{\ell \text{ times}} \geq 2^n.$$

Functions of fractional-exponential growth rate appear naturally in some contexts in computational complexity theory [18]. There are time-constructible functions of arbitrarily slow fractional-exponential growth:

LEMMA 3.3. *For every $\ell \in \mathbb{N}$, there exists a nondecreasing time-constructible function such that*

$$\underbrace{f \circ f \circ \cdots \circ f}_{\ell \text{ times}} = o(2^n),$$

but for some $m > \ell$,

$$\underbrace{f \circ f \circ \cdots \circ f}_{m \text{ times}} \geq 2^n.$$

PROOF. We present a very straightforward construction which suffices for our purposes. For more subtle results on iterative roots of exp as a function of a real variable, see e.g., [15, 21].

Define $\exp_k(n)$ by setting $\exp_0(n) := n$, $\exp_{k+1}(n) = 2^{\exp_k(n)}$.

Let f_2 be defined as follows. For $n \leq 2$, $f_2(n)$ equals 3. For $n \geq 3$, $f_2(n)$ is the smallest number strictly greater than n appearing in the sequence

$$3, 4, 2^3, 2^4, 2^{2^3}, 2^{2^4}, \ldots$$

The sequence enumerates all numbers of the form $\exp_i(j)$ for $i \in \mathbb{N}$, $j \in \{3, 4\}$, according to the lexicographic order on the pairs $\langle i, j\rangle$. A routine calculation shows that $f_2(n) = o(2^n)$, but $f_2 \circ f_2 \circ f_2(n) > 2^n$.

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

For general $k \geq 1$, define f_k as follows. For $n \leq \exp_{k-1}(2)$, $f_k(n)$ equals $\exp_{k-1}(2)+1$. For $n \geq \exp_{k-1}(2)+1$, $f_k(n)$ is the smallest number strictly greater than n appearing in the sequence

$$\begin{aligned}&\exp_{k-1}(2)+1, \exp_{k-1}(2)+2, \ldots, \exp_k(2),\\ &\qquad \exp(\exp_{k-1}(2)+1), \exp(\exp_{k-1}(2)+2), \ldots, \exp_{k+1}(2),\\ &\qquad\qquad \exp_2(\exp_{k-1}(2)+1), \exp_2(\exp_{k-1}(2)+1), \ldots\end{aligned}$$

The sequence is constructed analogously to the one for $k = 2$, except that now $j \in \{\exp_{k-1}(2)+1, \exp_{k-1}(2)+2, \ldots, \exp_k(2)\}$. A calculation shows that

$$\underbrace{f_k \circ f_k \circ \cdots \circ f_k}_{\ell \text{ times}} = o(2^n), \qquad \underbrace{f_k \circ f_k \circ \cdots \circ f_k}_{\ell+2 \text{ times}} > 2^n$$

for $\ell := \exp_k(2) - \exp_{k-1}(2) - 1$.

It is not immediately obvious if the functions f_k are time-constructible. However, let $\tilde{f}_k(n)$ be the time needed to compute f_k on argument n with both input and output in unary (on a fixed machine). By definition, $\tilde{f}_k$ is time-constructible. Moreover, it is clear from the definition of f_k that for a reasonable choice of a machine computing f_k, $\tilde{f}_k$ will be nondecreasing and satisfy:

$$f_k \leq \tilde{f}_k = O(f_k),$$

which also implies

$$\underbrace{\tilde{f}_k \circ \tilde{f}_k \circ \cdots \circ \tilde{f}_k}_{\ell-1 \text{ times}} = o(2^n),$$

for ℓ as above. ⊣

For every time-constructible f of fractional-exponential growth, it is still the case LinH is strictly contained in $\bigcup_k \Sigma_k\text{-TIME}(f)$, the quantifier hierarchy determined by f (cf. [13, Theorem V.2.21]). This is because a straightforward padding argument shows that $\bigcup_k \Sigma_k\text{-TIME}(f) \subseteq \text{LinH}$ would imply $\bigcup_k \Sigma_k\text{-TIME}(f \circ f) \subseteq \bigcup_k \Sigma_k\text{-TIME}(f)$ and so on, eventually reaching EH $\subseteq$ LinH and contradicting Proposition 3.1.

However, the following statement, although highly implausible, appears to be consistent with present knowledge, and disproving it may well be beyond the reach of known methods:

"for every k there is some f of fractional-exponential growth
such that $\Sigma_k\text{-TIME}(f^{O(1)}) \subseteq \text{LinH}$." (♠)

By the discussion above, f, or more precisely the number ℓ of iterations of f needed to reach exp, would have to depend on k. A very slight modification of statement (♠) is the assumption we use to prove that $\Pi_1(\mathbb{N}) + \neg\text{Exp} \vdash \text{B}\Sigma_1$.

THEOREM 3.4. *If there exists a nondecreasing time-constructible function f of fractional-exponential growth such that* $\text{DTIME}(f^{O(1)})^{\Sigma_k^{\text{p}}} \subseteq \text{LinH}$, *then every model* $\mathcal{M} \models \Pi_1(\mathbb{N}) + \neg\text{Exp}$ *has a proper end-extension to* $\mathcal{K} \models T_2^k$.

PROOF. Assume that f is nondecreasing, has fractional-exponential growth, and $\text{DTIME}(f^{O(1)})^{\Sigma_k^{\text{p}}}$ is contained in LinH. Let h be the time-constructible function

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

defined by $h(x) = 2^{f(|x|)}$. Note that for an input x with $|x| = n$, the value $h(x)$ (written in binary) is computable in deterministic time $f(n)$. Because of this and the monotonicity of f, for every PV_{k+1} formula $\psi(x, y)$, the formula $\psi(x, h(y))$ defines a property in $\mathrm{DTIME}(f^{O(1)})^{\Sigma^{\mathrm{p}}_k}$. Therefore, by our assumption, there exists a Δ_0 formula $\psi^{f\to\mathrm{lin}}(x, y)$ such that $\psi(x, h(y))$ and $\psi^{f\to\mathrm{lin}}(x, y)$ are equivalent in $\mathbb{N}$.

Let $\mathcal{M} \models \Pi_1(\mathbb{N}) + \neg\mathrm{Exp}$. Since the totality of f implies the totality of the exponential function, $\mathrm{Log}\,\mathcal{M}$ is not closed under f, so there must be $d \in \mathcal{M}$ such that $h(d) \notin \mathcal{M}$. As in the proof of Theorem 2.1, $\mathcal{M}$ has a Δ_0-elementary extension to a model $\mathcal{N}$ satisfying true arithmetic in the language L_{BA}. Of course, $\mathcal{N}$ contains $h(d)$. Define $\mathcal{K}$ to be the closure of $\mathcal{M} \cup \{h(d)\}$ in $\mathcal{N}$ under PV_{k+1} functions. By construction, $\mathcal{K}$ is a proper extension of $\mathcal{M}$, a PV_{k+1}-elementary substructure of $\mathcal{N}$ and a model of T^k_2; we only need to show that $\mathcal{K} \supseteq_e \mathcal{M}$.

Let g be a PV_{k+1} function and assume that for $\bar{a}, b \in \mathcal{M}$, we have $g^{\mathcal{N}}(\bar{a}, h(d)) \leq b$. Then $\mathcal{N} \models \exists z < b\,(z = g(\bar{a}, h(d)))$, and thus also

$$\mathcal{N} \models \exists z < b\,(z = g(\bar{x}, y))^{f\to\mathrm{lin}}[\bar{a}/\bar{x}, d/y].$$

By Δ_0-elementarity, the same formula holds in $\mathcal{M}$, so there is $c \in \mathcal{M}$ such that

$$\mathcal{M} \models (z = g(\bar{x}, y))^{f\to\mathrm{lin}}[\bar{a}/\bar{x}, d/y, c/z].$$

Thus also

$$\mathcal{N} \models (z = g(\bar{x}, y))^{f\to\mathrm{lin}}[\bar{a}/\bar{x}, d/y, c/z],$$

which means that $\mathcal{N} \models c = g(\bar{a}, h(d))$. In other words, $g^{\mathcal{N}}(\bar{a}, h(d)) \in \mathcal{M}$. ⊣

As in the context of $\neg\Omega_1$, end-extending arbitrary models of $\mathrm{I}\Delta_0 + \neg\mathrm{Exp}$ rather than just models of $\Pi_1(\mathbb{N}) + \neg\mathrm{Exp}$ requires a stronger assumption. The assumption in Theorem 3.5 below can be seen as a formalization of the statement "$\mathrm{I}\Delta_0 \vdash \mathrm{DTIME}(f^{O(1)})^{\Sigma^{\mathrm{p}}_k} \subseteq \mathrm{LinH}$."

The assumption in Theorem 3.5 is more complicated than the one in Theorem 2.2, and it mentions an additional constant. Another way of formulating the assumption would involve a two-sorted language. The main reason for the complications is that if we introduce a symbol for a function of fractional-exponential growth and allow terms involving nesting of the new function symbol, we will get a theory which proves the totality of exp and thus has no chance of being Π_1-conservative over $\mathrm{I}\Delta_0$.

THEOREM 3.5. *Assume that there is a time-constructible function f which is provably nondecreasing in $\mathrm{I}\Delta_0$ and satisfies the following*:

(i) *for some $\ell \in \mathbb{N}$,*

$$\mathrm{I}\Delta_0 \vdash \forall x\,\forall y\,(y = \underbrace{f \circ f \circ \cdots \circ f}_{\ell \text{ times}}(x) \Rightarrow \exists z \leq y\,(z = 2^x)).$$

(ii) *for d a new individual constant, there is a translation $\psi \mapsto \psi^{f\to\mathrm{lin}}$ of PV_{k+1} into Δ_0 formulas such that*

$$T^k_2 + \exists y\,(y = 2^{f(|d|)}) + \{\forall x\,(\psi(x, 2^{f(|d|)}) \Leftrightarrow \psi^{f\to\mathrm{lin}}(x, d)) : \psi \in \mathrm{PV}_{k+1}\}$$

is $\Pi_1(d)$-conservative over $\mathrm{I}\Delta_0$.

Then every model $\mathcal{M} \models \mathrm{I}\Delta_0 + \neg\mathrm{Exp}$ has a proper end-extension to $\mathcal{K} \models T^k_2$.

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

PROOF. It suffices to modify the proof of Theorem 3.4 in the same way as the proof of Theorem 2.1 was modified to obtain Theorem 2.2. ⊣

COROLLARY 3.6. *If for every* $k \in \mathbb{N}$ *there is a nondecreasing time-constructible* f *of fractional-exponential growth such that* Σ_k-TIME$(f^{O(1)}) \subseteq$ LinH, *then* $\Pi_1(\mathbb{N}) + \neg\mathrm{Exp} \vdash \mathrm{B}\Sigma_1$. *If the assumption of Theorem* 3.5 *holds for every* $k \in \mathbb{N}$, *then* $\mathrm{I}\Delta_0 + \neg\mathrm{Exp} \vdash \mathrm{B}\Sigma_1$.

PROOF. To prove the first statement, notice that DTIME$(f^{O(1)})^{\Sigma_k^{\mathrm{p}}}$ is contained in Σ_{k+1}-TIME$(f^{O(1)})$. Hence, if Σ_{k+1}-TIME$(f^{O(1)}) \subseteq$ LinH, then by Theorem 3.4 every model $\mathcal{M} \models \Pi_1(\mathbb{N}) + \neg\mathrm{Exp}$ has a proper end-extension to a model of T_2^k. If this happens for every k, then $\mathcal{M}$ has a proper end-extension to a model of T_2^k for every k. This implies $\mathcal{M} \models \mathrm{B}\Sigma_1$ by the proof of Proposition 1.1.

The second statement is proved in a similar way, using Theorem 3.5 instead of Theorem 3.4. ⊣

§4. A relativized result. In this section, we take up the question whether a proof of $\mathrm{I}\Delta_0 + \neg\mathrm{Exp} \nvdash \mathrm{B}\Sigma_1$ could relativize to an arbitrary oracle. The ideal result here would be to construct an oracle relative to which statement (♠) holds, and use it to show that a proof of $\mathrm{I}\Delta_0 + \neg\mathrm{Exp} \nvdash \mathrm{B}\Sigma_1$ would have to be "nonrelativizing." Our theorem below is weaker than that; nevertheless, it is strong enough to highlight a major difference with the case without ¬Exp.

In attempting to discuss relativizations of $\mathrm{I}\Delta_0 + \neg\mathrm{Exp} \nvdash \mathrm{B}\Sigma_1$, we encounter the problem that it is less obvious how to relativize an unprovability statement than a provability statement. If α is a new predicate symbol (representing a potentially arbitrary oracle) and two theories T and S are such that it makes sense to speak of their relativized versions $T(\alpha), S(\alpha)$, then "$T \vdash S$ does not relativize" is simply "$T(\alpha) \nvdash S(\alpha)$." However, what does "$T \nvdash S$ does not relativize" mean? "$T(\alpha) \vdash S(\alpha)$" is far too strong, and in particular implies $T \vdash S$. On the other hand "there are additional axioms about α such that $T(\alpha)$ extended by the new axioms proves $S(\alpha)$" is obviously too weak, because the new axioms could include $S(\alpha)$ itself.

In the specific case of $\mathrm{I}\Delta_0 + \neg\mathrm{Exp} \nvdash \mathrm{B}\Sigma_1$, we are in the fortunate situation that the right-hand side of the $\nvdash$ symbol has higher quantifier complexity than the left-hand side. This lets us explicate "$\mathrm{I}\Delta_0 + \neg\mathrm{Exp} \nvdash \mathrm{B}\Sigma_1$ does not relativize" as "there are low-complexity axioms about α such that $\mathrm{I}\Delta_0(\alpha) + \neg\mathrm{Exp}$ is consistent and proves $\mathrm{B}\Sigma_1(\alpha)$," where the "low-complexity axioms" can for instance say that some local conditions on α are satisfied at every input; in other words, they can be $\Pi_1(\alpha)$ sentences.

We conjecture the following:

CONJECTURE 4.1. *Let* α *be a new unary relation symbol. There exists a consistent recursively axiomatized* $\Pi_1(\alpha)$ *theory* $T(\alpha) \supseteq \mathrm{I}\Delta_0(\alpha)$ *such that* $T(\alpha) + \neg\mathrm{Exp} \vdash \mathrm{B}\Sigma_1(\alpha)$.

Currently, we are only able to obtain a weaker theorem:

THEOREM 4.2. *Let* α *be a new unary relation symbol. Then for every finite fragment* $\mathrm{B}(\alpha)$ *of* $\mathrm{B}\Sigma_1(\alpha)$ *there exists a consistent recursively axiomatized* $\Pi_1(\alpha)$ *theory* $T_{\mathrm{B}}(\alpha) \supseteq \mathrm{I}\Delta_0(\alpha)$ *such that* $T_{\mathrm{B}}(\alpha) + \neg\mathrm{Exp} \vdash \mathrm{B}(\alpha)$.

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

Already this result contrasts with the situation without the ¬Exp axiom.

PROPOSITION 4.3. *Let α be a new unary relation symbol. There is a finite fragment $\mathrm{B}(\alpha)$ of $\mathrm{B}\Sigma_1(\alpha)$ such that for every set $S(\alpha)$ of $\Pi_2(\alpha)$ sentences which is consistent with $\mathrm{I}\Delta_0(\alpha) + \mathrm{Exp}$ (in particular, for every consistent extension of $\mathrm{I}\Delta_0(\alpha)$ by a set of $\Pi_1(\alpha)$ sentences), $S(\alpha) \nvdash \mathrm{B}(\alpha)$.*

PROOF. Take a nonstandard model $\mathcal{M} \models \mathrm{I}\Delta_0(\alpha) + \mathrm{Exp} + S(\alpha)$, a nonstandard element $a \in \mathcal{M}$, and let $\mathcal{K}$ consist of those elements of $\mathcal{M}$ which are $\Sigma_1(\alpha)$-definable with parameter a. By the usual Paris-Kirby argument, $\mathcal{K} \not\models \mathrm{B}\Sigma_1(\alpha)$, but $\mathcal{K}$ is a $\Sigma_1(\alpha)$-elementary substructure of $\mathcal{M}$, so in particular $\mathcal{K} \models S(\alpha)$. This completes the proof, since $\mathrm{B}\Sigma_1(\alpha)$ is finitely axiomatizable over $\mathrm{I}\Delta_0(\alpha) + \mathrm{Exp}$. ⊣

REMARK 4.4. One of the referees points out that Proposition 4.3 can be generalized by letting $S(\alpha)$ consist of $\Sigma_3(\alpha)$ rather than just $\Pi_2(\alpha)$ sentences. To see this, invoke compactness to assume that $S(\alpha)$ is finite, and allow the witnesses for the outermost existential quantifiers in $S(\alpha)$ as additional parameters in the definition of $\mathcal{K}$.

Our proof of Theorem 4.2 is based on the following computational lemma.

LEMMA 4.5. *For every k, there is an oracle α such that $\Sigma_k^{\mathrm{e}}(\alpha) \subseteq \mathrm{LinH}(\alpha)$.*

We first show how the theorem follows from the lemma and then devote the remainder of the section to a proof of the lemma.

Assume that Lemma 4.5 holds and let $\mathrm{B}(\alpha)$ be a finite fragment of $\mathrm{B}\Sigma_1(\alpha)$. By a relativized version of the proof of Proposition 1.1, there is some k be such that every model of $\mathrm{I}\Delta_0(\alpha)$ with a proper end-extension to a model of $T_2^k(\alpha)$ satisfies $\mathrm{B}(\alpha)$. Let $T_{\mathrm{B}}(\alpha)$ consist of the $\Pi_1(\alpha)$ consequences of the theory $\mathrm{I}\Delta_0(\alpha) + \mathrm{Exp} +$ "$\Sigma_k^{\mathrm{e}}(\alpha) \subseteq \mathrm{LinH}(\alpha)$." (Strictly speaking, the containment of $\Sigma_k^{\mathrm{e}}(\alpha)$ in $\mathrm{LinH}(\alpha)$ is an infinite disjunction over the levels of $\mathrm{LinH}(\alpha)$ rather than a single sentence, so it should be replaced by a sentence expressing the equivalence of some fixed formula defining a property which is $\Sigma_k^{\mathrm{e}}(\alpha)$-complete via linear-time reductions with a *fixed* $\Delta_0(\alpha)$ formula obtained from Lemma 4.5 for this specific k.) By Lemma 4.5, $T_{\mathrm{B}}(\alpha)$ is consistent, and since it has a recursively enumerable axiomatization it is also recursively axiomatized.

By relativizing the proof of Theorem 3.5, which presents no significant obstacles, we show that every model of $T_{\mathrm{B}}(\alpha) + \neg\mathrm{Exp}$ has a proper end-extension to a model of $T_2^k(\alpha)$. Thus, $T_{\mathrm{B}}(\alpha) + \neg\mathrm{Exp} \vdash \mathrm{B}(\alpha)$, which completes the proof of Theorem 4.2 from Lemma 4.5.

PROOF OF LEMMA 4.5. For $k = 0, 1$, the result is known [8, 12]. For $k \geq 2$, our strategy will be to apply a variant of "Håstad's second switching lemma" (Lemma 6.3 in [11]) k times. As a result, we will partially determine the oracle α in such a way that a given $\Sigma_k^{\mathrm{e}}(\alpha)$ property will behave roughly like a deterministic exponential-time property (more precisely, an exponential-height decision tree) relative to the undetermined part of α, whereas a well-chosen LinH property will behave like a nondeterministic linear-time property. At that point, it will remain to repeat the construction of an oracle relative to which $\mathrm{E} \subseteq \mathrm{NLIN}$ as presented in [12].

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

We describe in detail the case of $k = 2$. We choose some reasonable coding of tuples and assume that all inputs to α are quadruples of the form $\langle a, y_1, y_2, y_3\rangle$, where $y_1, y_2 < 2^{4|a|}$, $y_3 < 2^{2|a|}$. We also use the convention that whenever a appears in the proof, N stands for $2^{|a|}$, and whenever b appears in the proof, M stands for $2^{|b|}$. Note that a computation on input a is linear-time if it takes time linear in $|a|$ and exponential-time (as understood in the definition of EH) if it takes time polynomial in N.

There is a problem in Σ_2-TIME$(2^{(1/110)n})(\alpha)$ which is complete for $\Sigma_2^{\mathrm{e}}(\alpha)$ via linear-time reductions. Therefore, it suffices to construct α so that the property $\varphi^\alpha(a)$:

$$\exists x_1 < 2^{N^{1/110}}\, \forall x_2 < 2^{N^{1/110}}\, \psi^\alpha(a, x_1, x_2),$$

where ψ^α makes at most $N^{1/110}$ queries to α, is for sufficiently large a equivalent to

$$\forall y_1 < N^4\, \exists y_2 < N^4\, \forall y_3 < N^2\, \alpha(a, y_1, y_2, y_3).$$

We think of each instance $\alpha(b, y_1, y_2, y_3)$ as a propositional variable, which is set to 1 exactly if $\langle b, y_1, y_2, y_3\rangle \in \alpha$. For each fixed a, $\varphi^\alpha(a)$ can be viewed as a propositional formula in the variables $\alpha(b, y_1, y_2, y_3)$. The formula is a disjunction of $2^{N^{1/110}}$ CNF's, each of which is a conjunction of disjunctions of at most $N^{1/110}$ atoms (an $N^{1/110}$-CNF).

We will refer to a partial assignment to the propositional variables as a *restriction*. We choose a random restriction ρ by the following process:

- divide all possible inputs to α into blocks of the form $\langle b, y_1, y_2, \cdot\rangle$, where each block consists of M^2 elements corresponding to the possible values of y_3,
- for each input $\langle b, y_1, y_2, y_3\rangle$ independently, set $\alpha(b, y_1, y_2, y_3)$ to 1 with probability $1 - \frac{1}{M}$ and "set it to $*$" (in other words, leave it unassigned, or "starred") with probability $\frac{1}{M}$,
- for each block $\langle b, y_1, y_2, \cdot\rangle$ in which not all the values have been set to 1, make it a "0-block" (setting $\langle b, y_1, y_2, y_3\rangle$ to 0 for all y_3 for which this was previously set to $*$) with probability $1 - \frac{1}{M}$ and make it a "$*$-block" (keeping the variables currently set to $*$ unassigned) with probability $\frac{1}{M}$.

The restriction ρ is then extended to $g(\rho)$ which additionally sets to 1 all starred variables in a $*$-block except the one with the smallest value of y_3.

The following variant of Håstad's switching lemma will guarantee the existence of a restriction ρ with some desirable properties. The proof of the lemma is deferred until Section 5.

LEMMA 4.6. *Let $0 < \delta, \varepsilon \leq 1$ such that $12\delta < \varepsilon$. Assume N is large enough and let ψ be an N^δ-DNF. Then with probability at least $1 - 2^{-N^\varepsilon}$, $\psi|_{g(\rho)}$ can be decided by a decision tree of height less than N^ε.*

The probability that a random ρ sets all M^2 variables in a given block $\langle b, y_1, y_2, \cdot\rangle$ to 1 is $(1 - \frac{1}{M})^{M^2}$, thus exponentially small in M. Similarly, by Chernoff bounds, the probability that fewer than M^2 of the M^4 blocks $\langle b, y_1, y_2, \cdot\rangle$ for a given pair $\langle b, y_1\rangle$ become $*$-blocks under ρ is exponentially small in M. For any given M, there are only $O(M^9)$ triples $\langle b, y_1, y_2\rangle$ and pairs $\langle b, y_1\rangle$ with $2^{|b|} = M$; and for any given N,

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

there are no more than N inputs a with $2^{|a|} = N$. Therefore, by Lemma 4.6 with $\delta = 1/110, \varepsilon = 1/9$, there exists a restriction ρ such that:

- for each sufficiently large b and each $\langle b, y_1, y_2\rangle$ there is at least one y_3 such that $\alpha(b, y_1, y_2, y_3)|_\rho \neq 1$,
- for each sufficiently large b and each $\langle b, y_1\rangle$ there are at least M^2 values of y_2 such that $\langle b, y_1, y_2\rangle$ is a $*$-block under ρ,
- for each sufficiently large a and each $x_1 < 2^{N^{1/110}}$, the formula

$$(\forall x_2 < 2^{N^{1/110}}\, \psi^\alpha(a, x_1, x_2))|_{g(\rho)}$$

can be decided by a decision tree of height $N^{1/9}$.

Take some ρ with these properties and assign 0/1 values to variables $\alpha(b, y_1, y_2, y_3)$ according to $g(\rho)$. For each sufficiently large a, the formula $\varphi^\alpha(a)|_{g(\rho)}$, that is,

$$\exists x_1 < 2^{N^{1/110}}\, [\forall x_2 < 2^{N^{1/110}}\, \psi^\alpha(a, x_1, x_2)]|_{g(\rho)},$$

becomes an $N^{1/9}$-DNF. Extend $g(\rho)$ further to a restriction $h(\rho)$ by turning some $*$-blocks into 0-blocks so that for each $\langle b, y_1\rangle$ with b sufficiently large there are exactly M^2 values of y_2 for which $\langle b, y_1, y_2, \cdot\rangle$ remains a $*$-block. In this way,

$$[\forall y_1 < N^4\, \exists y_2 < N^4\, \forall y_3 < N^2\, \alpha(a, y_1, y_2, y_3)]|_{h(\rho)}$$

becomes

$$\bigwedge_{y_1 < N^4} \bigvee_{y_2 \in S(a, y_1)} \alpha(a, y_1, y_2, y_3(a, y_1, y_2)),$$

where the set $S(a, y_1)$ consists of exactly N^2 values of y_2, and for each such value, $y_3(a, y_1, y_2)$ is the unique y_3 for which $\alpha(a, y_1, y_2, y_3)$ has not been set to 1.

We think of inputs to $\alpha|_{h(\rho)}$ as triples $\langle b, y_1, y_2\rangle$ (since $\alpha(b, y_1, y_2, y_3)$ can be 0 for at most one value of y_3). We divide the tuples into blocks of the form $\langle b, y_1, \cdot\rangle$, so that a sufficiently large b corresponds to exactly M^4 blocks containing exactly M^2 variables each. We define a random distribution on restrictions $\hat{\rho}$ similarly as before, but with α replaced by $\alpha|_{h(\rho)}$, the old blocks replaced by the new blocks, and with the roles of 0 and 1 interchanged. In this setting, we can prove an analogue of Lemma 4.6 in which the requirement $12\delta < \varepsilon$ can be weakened to $8\delta < \varepsilon$ (the reason for the last change is that for every b there are now M^6 variables instead of M^{10}). Applying this lemma with $\delta = 1/9$ and $\varepsilon = 1$, we get a restriction $\hat{\rho}$ such that for each sufficiently large a the formula $\varphi^\alpha(a)|_{h(\rho)g(\hat{\rho})}$ can be decided by a decision tree of height N, and extend $g(\hat{\rho})$ to $h(\hat{\rho})$ so that

$$[\forall y_1 < N^4\, \exists y_2 < N^4\, \forall y_3 < N^2\, \alpha(a, y_1, y_2, y_3)]|_{h(\rho)h(\hat{\rho})}$$

becomes

$$\bigwedge_{y_1 \in S(a)} \alpha(a, y_1, y_2(a, y_1), y_3(a, y_1)),$$

where $S(a)$ is a set of size exactly N^2.

We then complete the construction by copying the proof of Theorem 6 of [12]. For $a = 0, 1, \ldots$, we first find the value of $\varphi^\alpha(a)$ by answering 1 to each query to α which has not been already determined by $h(\rho)h(\hat{\rho})$ and the procedure for lower

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

values of a. Assuming a is large enough, this involves answering at most N queries, because $\varphi^\alpha(a)|_{h(\rho)h(\hat{\rho})}$ can be decided by a height-N decision tree. Afterwards, of the N^2 variables $\alpha(a, y_1, y_2(a, y_1), y_3(a, y_1))$ remaining unset after $h(\rho)h(\hat{\rho})$, at most $aN + O(1) \leq N \log N + O(1)$ have been set (to 1) during the procedure for $0, 1, \ldots, a-1$, so at least one remains unset. If $\varphi^\alpha(a)$ is false, we choose one such variable and set it to 0, setting all other such variables to 1. If $\varphi^\alpha(a)$ is true, we set all such variables to 1. In this way, we guarantee that $\varphi^\alpha(a)$ is equivalent to

$$\forall y_1 < N^4 \, \exists y_2 < N^4 \, \forall y_3 < N^2 \, \alpha(a, y_1, y_2, y_3)$$

for every large enough value of a.

This completes the proof of Lemma 4.5 for the case $k = 2$. The case for general k is very similar: for a sufficiently large constant ℓ, we take a property $\varphi^\alpha(a)$ which is complete for Σ^{e}_k via linear-time reductions and can be written as

$$\exists x_1 < 2^{N^{1/\ell}} \, \forall x_2 < 2^{N^{1/\ell}} \, \ldots \, \mathrm{Q} x_k < 2^{N^{1/\ell}} \, \psi^\alpha(a, x_1, x_2, \ldots, x_k),$$

where ψ can be decided in deterministic time $N^{1/\ell}$. The aim is to construct α so that this property becomes equivalent to

$$\forall y_1 < N^4 \, \exists y_2 < N^4 \, \ldots \, \mathrm{Q} y_k < N^4 \overline{\mathrm{Q}} y_{k+1} < N^2 \, \alpha(a, y_1, y_2, \ldots, y_k, y_{k+1})$$

for large enough a. To achieve this, we assume that inputs to α have the form $\langle b, y_1, \ldots, y_k, y_{k+1} \rangle$ for $y_1, \ldots, y_k < M^4$ and $y_{k+1} < M^2$, divide the inputs into blocks of the form $\langle b, y_1, \ldots, y_k, \cdot \rangle$, and iterate restricting α using analogues of Lemma 4.6, now k times instead of 2 times. For the i-th application of the switching lemma, a block contains inputs corresponding to one fixed tuple $\langle b, y_1, \ldots, y_{k+1-i} \rangle$ and various possible values of y_{k+2-i}, and the requirement on δ and ε is $4(k + 2 - i)\delta < \varepsilon$. After k rounds of restricting α, $\varphi^\alpha(a)$ becomes computable by a decision tree of height N and we can complete the argument as in the case $k = 2$. ⊣

§5. The switching lemma. We repeat the statement of Lemma 4.6. For definitions of N, ρ etc. refer to the previous section.

LEMMA 5.1. *Let $0 < \delta, \varepsilon \leq 1$ such that $12\delta < \varepsilon$. Assume N is large enough and let ψ be an N^δ-DNF. Then with probability at least $1 - 2^{-N^\varepsilon}$, $\psi|_{g(\rho)}$ can be decided by a decision tree of height less than N^ε.*

PROOF. The argument is an essentially standard Razborov-style proof of a switching lemma; our presentation is very strongly inspired by Thapen's note [22]. One difference in comparison with the usual setting for switching lemmas is that we simultaneously consider blocks of different sizes, and the probability of assigning a value to a variable/block depends on the size of the block. We deal with this by dividing blocks into "large" and "small"; our decision tree will deal with the "small" blocks by brute-force search, and with the "large" ones in the usual way.

Note that ψ contains only variables from finitely many blocks. For this reason, we may think of our restrictions ρ as defined on finitely many blocks, so that individual restrictions have nonzero probabilities.

Consider the tree $T(\psi, \rho)$ defined in the following way. $T(\psi, \rho)$ initially queries all the $\leq N^{\frac{11}{12}\varepsilon} < N^\varepsilon/2$ variables from blocks $\langle b, y_1, y_2, \cdot \rangle$ with $M < N^{\varepsilon/12}$ that are

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

starred in ρ. Let τ be the assignment to those variables given by a specific branch of the tree. If τ is inconsistent with $g(\rho)$, the branch ends immediately in a leaf labelled (for example) 0. Otherwise, the subtree of $T(\psi, \rho)$ below τ, say $T^{\tau}(\psi, \rho)$, is defined by the following process. Let C_1 be the first conjunction in ψ such that $C_1|_{\rho\tau} \not\equiv 0$. Let β_1 be the list of blocks which contain a starred variable appearing in $C_1|_{\rho\tau}$. For each such block $\langle b, y_1, y_2, \cdot\rangle$, the tree queries α_{b,y_1,y_2,y_3} for the unique y_3 for which this is starred in $g(\rho)$ (even if this is not the variable appearing in $C_1|_{\rho\tau}$). Let π_1 be the assignment to the blocks in β_1 given by $g(\rho)$ together with the answers to the queries made by the tree. If $C_1|_{\rho\tau\pi_1} \equiv 1$, the branch ends in a leaf labelled 1. If not, let C_2 be the first conjunction in ψ such that $C_1|_{\rho\tau} \not\equiv 0$, let β_2 consist of all blocks containing a starred variable appearing in $C_2|_{\rho\tau\pi_1}$ and continue as previously. If at any point $C_i|_{\rho\tau\pi_1\ldots\pi_i} \equiv 1$, the branch ends in a leaf labelled 1. Otherwise, at some point all conjunctions in ψ have been falsified and the tree ends in a leaf labelled 0. By construction, $T(\psi, \rho)$ correctly determines the truth value of $\psi|_{g(\rho)}$.

Let S be the set of the ρ such that there exists τ for which $T^{\tau}(\psi, \rho)$ has height at least $N^{\varepsilon}/2$. To prove the lemma, it suffices to show

$$\Pr_{\rho}[\rho \in S] \leq \frac{1}{2^{N^{\varepsilon}}}.$$

Let ρ be in S, let τ be the first assignment such that $T^{\tau}(\psi, \rho)$ has height at least $N^{\varepsilon}/2$, and let π be the first branch in $T^{\tau}(\psi, \rho)$ of length at least $N^{\varepsilon}/2$. Let $C_1, \ldots, C_k$, $\beta_1, \ldots, \beta_k$, $\pi_1, \ldots, \pi_k$ be the conjunctions, sets of blocks, and assignments encountered in the first $N^{\varepsilon}/2$ queries in π. For $i = 1, \ldots, k$, let γ_i be the set of the starred variables from blocks in β_i appearing positively in C_i.

For $i = 1, \ldots, k$, the assignment σ_i is defined as follows: σ_i sets each variable appearing in γ_i to 1 and sets all other starred variables from blocks in β_i to 0. Note that σ_i sets exactly the same variables as π_i, so that the σ_i's have pairwise disjoint domains and thus $\sigma = \sigma_1 \ldots \sigma_k$ is a well-defined restriction. Note also that $C_i|_{\rho\tau\pi_1\ldots\pi_{i-1}\sigma_i} \not\equiv 0$ (in fact, $C_i|_{\rho\tau\pi_1\ldots\pi_{i-1}\sigma_i} \equiv 1$ for all $i = 1, \ldots, k-1$).

We can represent τ by a number below $3^{N^{\varepsilon}/2}$ (each of $N^{\varepsilon}/2$ variables can be set to 1, 0 or already evaluated in ρ). We code $\langle \beta_1, \ldots, \beta_k\rangle$ as a tuple $\beta' = \langle \beta'_1, \ldots, \beta'_k\rangle$, where for each block in β_i, β'_i contains the information which of the N^{δ} variables in C_i belong to the block and an additional bit to determine whether the next block listed in β' is still in β_i or already in β_{i+1}. Since there are at most $N^{\varepsilon}/2$ blocks in all the β_i's taken together, β' can be represented by a number below $(2N^{\delta})^{N^{\varepsilon}/2}$. We code $\langle \pi_1, \ldots, \pi_k\rangle$ as a number π' below $2^{N^{\varepsilon}/2}$ (a string of $N^{\varepsilon}/2$ bits listing the answers to queries given along π). Finally, we code $\langle \gamma_1, \ldots, \gamma_k\rangle$ as a tuple $\gamma' = \langle \gamma'_1, \ldots, \gamma'_k\rangle$, where each γ'_i contains the information which of the N^{δ} variables in C_i comprise γ_i. Since there are at most $N^{\varepsilon}/2$ C_i's, γ' can be represented by a number below $2^{N^{\delta+\varepsilon}/2}$.

We claim that the mapping:

$$S \ni \rho \mapsto \langle \rho\tau\sigma, \tau, \beta', \pi', \gamma'\rangle$$

is an injection. To see this, note that given $\langle \rho\tau\sigma, \tau, \beta', \pi', \gamma'\rangle$, we can find C_1 as the first conjunction C in ψ such that $C|_{\rho\tau\sigma} \neq 0$. Knowing C_1, we can decode β'_1 and γ'_1 to find β_1, γ_1. We can now identify σ_1 as the assignment which sets all variables in γ_1 to 1 and sets all the variables from blocks in β_1 which are set to 0 by $\rho\tau\sigma$.

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

Knowing σ_1, γ_1 and π'_1, we can determine π_1, identify C_2 as the first conjunction C in ψ such that $C|_{\rho\tau\pi_1\sigma_2\ldots\sigma_k} \neq 0$, identify $\beta_2, \gamma_2, \sigma_2, \pi_2$, and so on. Eventually, we are able to recover all of σ and therefore also $\rho\tau$. Since we are also given τ, this is enough to recover ρ.

For fixed values of $\tau, \beta', \pi', \gamma'$, let $S_{\tau,\beta',\pi',\gamma'}$ consist of those $\rho \in S$ to which these specific values are assigned. By the previous paragraph, the mapping $\rho \mapsto \rho\tau\sigma$ restricted to $S_{\tau,\beta',\pi',\gamma}$ is an injection. For $\rho \in S_{\tau,\beta',\pi',\gamma}$, the restriction $\rho\tau$ is obtained from ρ by turning some $*$'s into 1's and some $*$-blocks into 0-blocks, so certainly the probability of $\rho\tau$ is no smaller than that of ρ. Going from $\rho\tau$ to $\rho\tau\sigma$ changes $N^\varepsilon/2$ blocks which were previously $*$-blocks into 0-blocks, and changes some $*$'s into 1's. Since all variables from blocks with $M < N^{\varepsilon/12}$ are assigned 0/1 values in $\rho\tau$, each change of a $*$-block into a 0-block or of a $*$ into 1 increases the probability of a restriction by a factor of at least

$$\left(1 - \frac{1}{N^{\varepsilon/12}}\right) \Big/ \left(\frac{1}{N^{\varepsilon/12}}\right) = N^{\varepsilon/12} - 1.$$

If we write ℓ to denote the number of $*$'s changed into 1's, then ℓ equals the number of positively appearing variables listed in γ', so $0 \leq \ell \leq N^{\delta+\varepsilon}/2$. The ratio $\Pr(\rho\tau\sigma)/\Pr(\rho)$ is at least

$$\left(N^{\varepsilon/12} - 1\right)^{N^\varepsilon/2+\ell}$$

and the probabilities of the $\rho\tau\sigma$'s cannot add up to more than 1, so

$$\Pr_\rho[\rho \in S_{\tau,\beta',\pi',\gamma}] \leq \left(N^{\varepsilon/12} - 1\right)^{-N^\varepsilon/2-\ell}.$$

For fixed τ, β', π', write $S_{\tau,\beta',\pi'}$ for $\bigcup_{\gamma'} S_{\tau,\beta',\pi',\gamma'}$. We have

$$\begin{aligned}
\Pr_\rho[\rho \in S_{\tau,\beta',\pi'}] &= \sum_{\gamma'} \Pr_\rho[\rho \in S_{\tau,\beta',\pi',\gamma'}] \\
&= \sum_{\ell=0}^{N^{\delta+\varepsilon}/2} \sum_{|\gamma'|=\ell} \Pr_\rho[\rho \in S_{\tau,\beta',\pi',\gamma'}] \\
&\leq \sum_{\ell=0}^{N^{\delta+\varepsilon}/2} \binom{N^{\delta+\varepsilon}/2}{\ell} \left(N^{\varepsilon/12} - 1\right)^{-N^\varepsilon/2-\ell} \\
&= \left(N^{\varepsilon/12} - 1\right)^{-N^\varepsilon/2} \left(N^{\varepsilon/12} \Big/ \left(N^{\varepsilon/12} - 1\right)\right)^{N^{\delta+\varepsilon}/2} \\
&\leq \left(N^{\varepsilon/12} - 1\right)^{-N^\varepsilon/2} e^{N^\varepsilon/2}.
\end{aligned}$$

The last inequality holds for sufficiently large N by the assumption that $12\delta < \varepsilon$. Now, $S = \bigcup_{\tau,\beta',\pi'} S_{\tau,\beta',\pi'}$, so by the union bound

$$\Pr_\rho[\rho \in S] \leq \left(12eN^\delta\right)^{N^\varepsilon/2} \left(N^{\varepsilon/12} - 1\right)^{-N^\varepsilon/2},$$

which is smaller than 2^{-N^ε} for large enough N. ⊣

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

§6. **Concluding remarks.** We have formulated a complexity-theoretic statement that might be very hard to disprove and would imply that $B\Sigma_1$ is provable from $\neg$Exp. It would be interesting to give evidence for the hardness of disproving the statement (♠), for example by constructing a relativized world in which (♠) holds; this would also lead to a proof of our conjecture from Section 4. On the other hand, disproving (♠) would be no less interesting, and could lead to some progress towards a proof that $I\Delta_0 + \neg\text{Exp} \nvdash B\Sigma_1$.

The question concerning provability of $B\Sigma_1$ is just one of many classical questions about weak arithmetic which do not explicitly mention complexity theory but seem to have it lurking in the background. One other example is the problem whether every model of $I\Delta_0 + B\Sigma_1$ has a proper end-extension to a model of $I\Delta_0$ (cf. [23]). It could be interesting to find, on the one hand, "plausible" complexity-theoretic statements implying the "expected" answers to these questions, and on the other hand, "plausibly hard to disprove" statements implying the "unexpected" answers. The former task, in fact, not yet fully carried out even in the case of the question studied in this paper. The best known complexity-theoretic statement implying $I\Delta_0 + \neg\text{Exp} \nvdash B\Sigma_1$, namely PH↓, is usually conjectured to be false, while other such statements are more obscure and seem hard to make any conjecture about.

§7. **Acknowledgments.** The author is grateful to the two anonymous referees for many helpful comments and suggestions. The author is partially supported by Polish National Science Centre grant no. 2013/09/B/ST1/04390.

REFERENCES

[1] Z. Adamowicz, *A recursion-theoretic characterization of instances of $B\Sigma_n$ provable in $\Pi_{n+1}(\mathbf{N})$*. ***Fundamenta Mathematicae***, vol. 129 (1988), pp. 231–236.

[2] Z. Adamowicz, L. A. Kołodziejczyk, and J. Paris, *Truth definitions without exponentiation and the Σ_1 collection scheme*, this Journal, vol. 77 (2012), pp. 649–655.

[3] J. L. Balcázar, J. Díaz, and J. Gabarró, ***Structural Complexity II***, Springer-Verlag, Berlin, 1990.

[4] ———, ***Structural Complexity I***. Second, Revised Edition, Springer-Verlag, 1995.

[5] L. D. Beklemishev, *A proof-theoretic analysis of collection*. ***Archive for Mathematical Logic***, vol. 37 (1998), pp. 275–296.

[6] S. Boughattas and L. A. Kołodziejczyk, *The strength of sharply bounded induction requires MSP*. ***Annals of Pure and Applied Logic***, vol. 161 (2010), pp. 504–510.

[7] A. Cordón-Franco, A. Fernández-Margarit, and F. F. Lara-Martín, *On axiom schemes for T-provably Δ_1 formulas*. ***Archive for Mathematical Logic***, vol. 53 (2014), pp. 327–349.

[8] M. I. Dekhtyar, *On the relativization of deterministic and nondeterministic complexity classes*. ***Mathematical Foundations of Computer Science 1976***, vol. 45, Lecture Notes in Computer Science, Springer, Berlin, 1976, pp. 255–259.

[9] F. Ferreira, *Binary models generated by their tally part*. ***Archive for Mathematical Logic***, vol. 33 (1994), pp. 283–289.

[10] ———, *On end-extensions of models of $\neg$ exp*. ***Mathematical Logic Quarterly***, vol. 42 (1996), pp. 1–18.

[11] J. Håstad, ***Computational Limitations for Small Depth Circuits***, PhD thesis, MIT, 1986.

[12] H. Heller, *On relativized polynomial and exponential computations*. ***SIAM Journal on Computing***, vol. 13 (1984), pp. 717–725.

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

[14] E. Jeřábek, *The strength of sharply bounded induction*. ***Mathematical Logic Quarterly***, vol. 52 (2006), pp. 613–624.

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press

[15] H. Kneser, *Reelle analytische Lösungen der Gleichung* $\varphi(\varphi(x)) = e^x$ *und verwandter Funktionalgleichungen*. ***Journal für die reine und angewandte Mathematik***, vol. 187 (1950), pp. 56–67.

[16] J. Krajíček, ***Bounded Arithmetic, Propositional Logic, and Complexity Theory***, Cambridge University Press, Cambridge, 1995.

[17] J. Krajíček, P. Pudlák, and G. Takeuti, Bounded arithmetic and the polynomial hierarchy. ***Annals of Pure and Applied Logic***, vol. 52 (1991), pp. 143–154.

[18] P. B. Miltersen, N. V. Vinodchandran, and Osamu Watanabe, *Super-polynomial versus half-exponential circuit size in the exponential hierarchy*. ***Computing and Combinatorics 5th Annual International Conference, COCOON'99*** (T. Asano *et al.*, editor), Springer, Berlin, 1999, pp. 210–220.

[19] C. Parsons, *On a number-theoretic choice schema and its relation to induction*. ***Intuitionism and Proof Theory. Proceedings of the Summer Conference at Buffalo, N.Y., 1968***, vol. 57, Studies in logic and the foundations of mathematics, North-Holland, Amsterdam, 1970, pp. 459–473.

[20] J. B. Paris and L. A. S. Kirby, Σ_n *collection schemas in arithmetic*. ***Logic Colloquium '77***, vol. 96, Studies in Logic and the Foundations of Mathematics, North Holland, Amsterdam, 1978, pp. 199–209.

[21] G. Szekeres, *Fractional iteration of exponentially growing functions*. ***Journal of the Australian Mathematical Society***, vol. 2 (1961/62), pp. 301–320.

[22] N. Thapen, *Notes on switching lemmas*, unpublished, available at users.math.cas.cz/thapen, 2009.

[23] A. Wilkie and J. 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, editors), pp. 143–161, North-Holland, Amsterdam, 1989.

[24] C. Wrathall, *Rudimentary predicates and relative computation*. ***SIAM Journal on Computing***, vol. 7 (1978), pp. 194–209.

[25] D. Zambella, *End extensions of models of linearly bounded arithmetic*. ***Annals of Pure and Applied Logic***, vol. 88 (1997), pp. 263–277.

INSTITUTE OF MATHEMATICS
UNIVERSITY OF WARSAW
BANACHA 2, 02-097 WARSZAWA
POLAND
E-mail: lak@mimuw.edu.pl

https://doi.org/10.1017/jsl.2015.53 Published online by Cambridge University Press