Structures interpretable in models of bounded arithmetic

Neil Thapen*
St Hilda's College, University of Oxford
neil.thapen@st-hildas.ox.ac.uk

March 30, 2005

Abstract

We look for a converse to a result from [12] that if the weak pigeonhole principle fails in a model K of bounded arithmetic, then there is an end-extension of K interpretable inside K. We show that if a model J of an induction-free theory of arithmetic is interpretable inside K, then either J is isomorphic to an initial segment of K (J is "smaller" than K), or K is isomorphic to an initial segment of J (J is "bigger" than K) and in this case the the weak pigeonhole principle fails in K. This result is formulated in terms of a theory S^1_0 of bounded arithmetic with a greatest element.

We go on to consider structures defined by oracles, and use the probabilistic witnessing theorem for $S^1_2 +$ (dual WPHP(PV)) to give a general criterion for what can be proved about these using the weak pigeonhole principle. We also show that the injective WPHP is not provable in this theory in the relativized case.

*Part of this work was done at the Mathematical Institute of the University of Oxford supported by EPSRC grant 98001658, and part at the University of Toronto, supported by a postdoctoral fellowship

Keywords: bounded arithmetic, arithmetic with a top, weak pigeonhole principle
AMS Classification: 03H15, 68Q17, 03F30

1 Introduction

If a is a natural number and $<^*$ is any total ordering on the interval $[0,a) =_{def} \{0,\ldots,a{-}1\}$, then the standard structure $\langle [0,a), < \rangle$ is isomorphic to $\langle [0,a), <^* \rangle$. We are interested in the complexity of isomorphisms of this kind.

In this example, we will help ourselves by assuming that as well as the relation $<^*$, we also have access to the $<^*$-minimal element and to a function for $<^*$-successor. Then if we are given an element of $\langle [0,a), < \rangle$ and want to find the element of $\langle [0,a), <^* \rangle$ that it maps to, in the worst case it will take us $a-1$ steps of applying the $<^*$-successor function to find the element we want.

Now suppose we enrich the language even more and have a structure $J = \langle [0,a), 0^*, 1^*, +^*, <^* \rangle$ which we know satisfies enough of the axioms of arithmetic to be isomorphic to the standard structure on $[0,a)$. Then given an element of the standard structure we can find its counterpart in J much more quickly, in time polynomial in $\log a$, by using repeated doubling to find the powers of two in J and then expressing our element as a sum of powers of two.

Notice that in these examples we are implicitly relying on the pigeonhole principle. We construct a function $a \to a$, the properties of the structure J guarantee that it is an injection, and from the pigeonhole principle it follows that it is a bijection. Notice also the different nature of the problem if we do it in reverse: if we are given an element of J and want to find its counterpart in the standard structure, we need the ordering $<^*$ to quickly find its binary expansion with respect to J.

It is well known that the first construction above can be formalized in Peano Arithmetic, to show that if any model J of PA is interpreted in a model K of PA, then J is an end-extension of K. In the first half of this paper we formalize a version of the second construction in a theory S^1_0 of bounded arithmetic. We show that if any model J of an algebraic fragment R of S^1_0 is interpretable (in the right way) inside a model K of S^1_0, then either J is isomorphic to an initial segment of K, or J is isomorphic to an end-extension of K and the weak pigeonhole principle (WPHP) fails in K.

Here S^1_0 is closely related to Buss' theory S^1_2. The main difference is that S^1_0 is in a relational language (except for constant symbols for 0,1 and 2) and has an axiom that there is a greatest element. It is intended as an axiomatization of the important properties of initial segments of S^1_2 (here, and in general, by initial segment we mean initial segment with a greatest element). We want to be able to talk about our interpreted structures using bounded formulas, so it is natural to consider structures whose domains are bounded sets. As a result interpreted structures with no greatest element will be the pathological case rather than the normal case, which is why we work with theories of arithmetic with a top. Another advantage is that in a model with a top every quantifier is automatically bounded, so we can apply model-theoretic results that hold for general formulas to bounded formulas, although we do not use this in this paper. Lastly it makes our results slightly more general. In particular it gives us an elegant way to analyze how much quantification is needed for an argument, by moving the top element higher or lower and looking at how the properties of the structure change.

Section 2 below contains definitions of these theories, and section 3 gives our construction of an isomorphism and the connection with WPHP. Our original motivation was to find a converse to a theorem in [12], included as theorem 3.4 below, which states that if WPHP fails (in the right way) in a model K of arithmetic with a top, then there is a large end extension of K interpretable inside K. Corollary 3.10 is almost a converse; but by Friedman's theorem any countable nonstandard model of PA has an isomorphic proper initial segment (here without a top element) so we can find a model of arithmetic with a top with a similar property, and in which WPHP holds.

This is an obstacle to the moral which we would be like to be able to draw from the results of this paper, together with [12] and [8]: that the weak pigeonhole principle holds in a model M of arithmetic if and only if large initial segments of M are in some sense "more complicated" than small ones. However despite this I think that the idea of the relative complexity of initial segments, studied here in terms of what sort of "interpreted structures" or "inner models" can live in a segment, is useful and can give some insight into the open problems regarding the provability of WPHP.

At the end of section 3 we present some corollaries: that in a model of S^1_2 in which WPHP holds we can precisely count structured sets (where a structured set is one that comes with relations making it into a model of the weak theory R) and that $S^1_2 + \mathrm{WPHP}$ proves that every finite model of R is a model of S^1_0.

In the last section we look at the limits of what we can prove about an interpreted structure using the weak pigeonhole principle. Our tool here is a result of Wilkie that uses Buss' witnessing theorem to show that if a $\forall\Sigma^b_1$ formula is provable in S^1_2 plus the surjective WPHP for PV formulas, then it can be witnessed in probabilistic polynomial time. We give a new proof of this result using some of the development of this theory in [12]. We use this to show that this theory does not prove that a finite linear ordering is isomorphic to the standard linear ordering (in fact it does not prove that it has a least element) and that it does not prove the injective WPHP. In corollary 4.9 we give a general sufficient condition for unprovability from this theory in the relativized setting, intended as an analogue of Riis' elegant criterion for unprovability from T^1_2:

Theorem 1.1 (Riis [9]) *Let Φ be any sentence containing only symbols from a language α disjoint from the language of arithmetic. If Φ has an infinite model then $T^1_2(\alpha) \nvdash \forall a\,(\langle [0,a), \alpha\rangle \vDash \neg\Phi)$.*

The results from this paper are presented in more detail in my doctoral thesis [11], except for the proof of theorem 4.7. Most of the remaining material in [11] has appeared as [12].

I would like to thank my supervisor Alex Wilkie for his guidance, and also Jan Krajíček, Avner Magen, Tsuyoshi Morioka, Stephen Cook and the anonymous referee for their comments on earlier versions of this work.

2 Definitions

Our language will consist of the three place relations $x = y + z$ and $x = y \cdot z$, the two place relations $x < y$ and $|x| = y$ and the constants $0, 1, 2$. We use a relational language because $+$ and $\cdot$ do not define total functions and to ensure that initial segments of models of our theory are always substructures. The intended interpretation of the length $|x|$ of x is the number of digits in the binary expansion of x. So for example 2^i has length $i + 1$ and $2^i - 1$ has length i.

Definition 2.1 *We define a theory* BASIC$'$ *to fix the simple properties of these symbols. It consists of the following axioms:*

1. *$<$ is a discrete linear ordering;*
2. *$0, 1, 2$ are the first three elements in this ordering;*
3. *$+, \cdot$ define partial functions and $|\ |$ a total function (hence where these are defined we will use function notation);*
4. *$x + 1 = y \leftrightarrow y$ is the successor of x; $x \cdot 1 = x$;*
5. *$x + y = z \leftrightarrow y + x = z$; $x \cdot y = z \leftrightarrow y \cdot x = z$;*
6. *If $(x+y)+z = w$ then $y+z$ and $x+(y+z)$ are defined and $x+(y+z) = w$; similarly for $\cdot$;*
7. *If $x \cdot (y + z)$ is defined then $x \cdot y$, $x \cdot z$ and their sum are defined and $x \cdot (y + z) = x \cdot y + x \cdot z$; similarly for multiplication on the left;*
8. *$x + 0 = x$ and $x \cdot 0 = 0$ (the rest of the normal inductive definitions of $+$ and $\cdot$ follow from axioms 4,5,6 and 7);*

9. $|0| = 0$ *and* $|1| = 1$*;*

10. $x \neq 0 \rightarrow (|2 \cdot x| = |x| + 1 \wedge |2 \cdot x + 1| = |x| + 1)$ *and when the left hand side of either of the conjucts is defined, so is the right hand side;*

11. *If* $x < y \wedge (z + y$ *is defined* $)$ *then* $(z + x$ *is defined and* $z + x < z + y)$*; if* $x < y \wedge 0 < z \wedge (z \cdot y$ *is defined* $)$ *then* $(z \cdot x$ *is defined and* $z \cdot x < z \cdot y)$*;*

12. $x \leqslant y \rightarrow |x| \leqslant |y|$*;*

13. $x < y \leftrightarrow \exists z\, (0 < z \leqslant y \wedge x + z = y)$*;*

14. $|x| + 1 < |y| \rightarrow 2 \cdot x$ *exists;*

15. $\exists y\, (x = 2 \cdot y \vee x = 2 \cdot y + 1)$.

In summary, BASIC$'$ *says that* $+$ *and* $\cdot$ *are partial functions and* $|\ |$ *is a total function, that the inductive properties of our symbols hold, that we can do subtraction where appropriate, and that every number is even or odd. No axiom guarantees the existence of anything larger than its parameters.*

R is the theory consisting of BASIC$'$ *together with an axiom stating that there is a greatest element. A model of R is said to be of the form* $[0, e + 1)$ *if it has a greatest element e.*

To expand on what is meant by $[0, e + 1)$, it is often the case that the most natural expression to describe the form of a strucure does not name any element in the structure. For example we may want to consider models of the form $[0, 2^a)$ (for some a in the model). For a definition intrinsic to the model, we could think of it as the set of binary strings of length a, or define the top element to be the sum of 2^{a-1} and $2^{a-1} - 1$.

We will define a new class of formulas, the $\bar{\Sigma}_i^b$ formulas. These are very similar to the Σ_i^b formulas but their syntax is more appropriate for working with models with a top element, since we will bound our quantifiers with variables rather than with terms. In models with a top every quantifier is implicitly bounded by the top element.

We first need a new definition of “sharply bounded”:

Definition 2.2 *A quantifier is sharply bounded if it appears in the form* $\forall x \leqslant |y|^k$ *or* $\exists x \leqslant |y|^k$ *where* x *and* y *are variables and* $k \in \mathbb{N}$.

This definition is equivalent to the standard definition of sharply bounded, in the structures that we consider. It does not quite work as it stands, since $|\ |$ and $\cdot$ are relation rather than function symbols. So for example $\exists x \leqslant |y|^2 \, \phi(x, y)$ written out fully is

$$\exists a \, \exists b \, (a = |y| \wedge b = a \cdot a \wedge \exists x \, (x \leqslant b \wedge \phi(x, y))).$$

This will not cause any problems, since in models of our theories $|\ |$ will always define a function and multiplication will generally be a total function when restricted to lengths.

Definition 2.3 *A formula is* $\bar{\Sigma}^b_i$ *if it contains no unbounded quantifiers and* $i-1$ *alternations of blocks of variable-bounded quantifiers beginning with a bounded existential quantifier and ignoring sharply bounded quantifiers. The* $\bar{\Pi}^b_i$ *formulas are defined dually. A set in a structure is* $\bar{\Delta}^b_i$ *if it is defined by both a* $\bar{\Sigma}^b_i$ *and a* $\bar{\Pi}^b_i$ *formula, with parameters .*

We will always allow parameters in formulas defining sets in a structure, unless we say otherwise explicitly.

Definition 2.4 *For* $i \geqslant 1$, S^i_0 *is the theory consisting of* R *together with the length induction axiom*

$$[(|z|^k exists) \wedge \phi(0) \wedge \forall x < |z|^k \, (\phi(x) \rightarrow \phi(x+1))] \rightarrow \phi(|z|^k)$$

for all $\bar{\Sigma}^b_i$ *formulas* ϕ *and all* $k \in \mathbb{N}$*;* z *is a parameter and* ϕ *may possibly contain other parameters. The set of length induction axioms for* $\bar{\Sigma}^b_i$ *formulas is called* $\bar{\Sigma}^b_i$*−*LIND.

The theory S^1_0 is strong enough to prove that we can consider numbers as codes for binary sequences and define a relation $\mathrm{bit}(x, i) = 1$. In S^i_0 we can prove that any short binary sequence defined by a $\bar{\Delta}^b_i$ formula is coded by some number. The proofs are standard, we just need to be careful that we do not need to use any numbers larger than our top element - see [11] for details.

We could convert S_0^i into a theory equivalent to S_2^i by replacing the axiom that there is a greatest element with an axiom stating that the smash function is total, $\forall x\, \forall y\, \exists z\, |z| = |x|\cdot|y|$. Typical models of S_0^1 are initial segments (with a top) of models of S_2^1, although it is unlikely that every model of S_0^1 has an end-extension to a model of S_2^1.

The relativized versions of these formulas and theories are defined analogously with the relativized forms of Σ_i^b and S_2^i. The only difference is that we do not need to add axioms limiting the growth rate of any new functions we introduce, since their ranges are automatically bounded by the top element.

We will also make use of the theory PV and of a relativized version of it. Our definitions of these are slightly nonstandard and we give them here:

Definition 2.5 *The language L_{PV} of* PV *function symbols consists of a function symbol for every polynomial time function. The theory* PV *is the set of all universal consequences of $S_2^1(\mathrm{PV})$ in the language L_{PV}.*

Suppose α is a set of function and relation symbols. The language $L_{\mathrm{PV}}(\alpha)$ consists of a function symbol for every polynomial time Turing machine with oracles for the functions and relations in α. In particular $L_{\mathrm{PV}}(\alpha)$ contains all the functions in α and the characteristic functions of all the relations in α. The theory $\mathrm{PV}(\alpha)$ *consists of the universal consequences of $S_2^1(\alpha)$ in this language.*

Definition 2.6 *For $a < b$, we consider three different forms of the pigeonhole principle.*

1. $\mathrm{PHP}_a^b(f)$ *states that f is not an injection from b into a;*

2. $\mathrm{PHP}_b^a(f)$ *states that f is not a surjection from a onto b (this is sometimes called the dual* PHP*);*

3. $\mathrm{mPHP}^b_a(f,s)$ *states that* f *is not a surjection from a subset* s *of* $[0,a)$ *onto* b*.* mPHP *stands for multifunction* PHP*, since it is equivalent to the statement that the inverse of* f *is not an injective multifunction from* b *into* a*.*

We write $\mathrm{PHP}^x_y(\Gamma)$ *for the set consisting of* $\mathrm{PHP}^x_y(f)$ *for every function in the class* Γ *(or every relation or set, as appropriate). If* $b \geqslant a^2$*, we call any of these the* weak *pigeonhole principle.*

Note that the surjective and injective PHPs are weakenings of multifunction PHP. By theorem 1.1 none of these principles is provable in T^1_2, in the relativized case. In [7] it is shown that they are all provable in T^2_2.

3 Interpreted structures

For simplicity we present these results for the theory S^1_0 and for $\bar{\Sigma}^b_1$ and $\bar{\Delta}^b_1$ formulas, although versions of them hold higher in the hierarchy.

Definition 3.1 *If* S *is a set in a model* $K \vDash R$*,* $k \in \mathbb{N}$ *and* $X \subseteq S^k$*, then* X *is said to be* $\bar{\Delta}^b_1$ *in* S *if both* X *and* $S^k \setminus X$ *are definable by* $\bar{\Sigma}^b_1$ *formulas.*

Definition 3.2 *Let* $K \vDash R$*. We say that a structure* J *(in our language) is* $\bar{\Sigma}^b_1$ *interpreted in* K *if there exist a* $\bar{\Sigma}^b_1$ *subset* S *of* K *and relations* $=_J$*,* $<_J$*,* $\cdot_J$*,* $+_J$ *and* $|\ |_J$ *that are* $\bar{\Delta}^b_1$ *in* S *such that* J *consists of the domain* $S/_{=_J}$ *with the relations induced by* $<_J$*,* $\cdot_J$*,* $+_J$ *and* $|\ |_J$*. For* $a \in K$*, we say that* J *is interpreted below* a *if* $S \subseteq [0,a)$*.*

We will often write elements of such a J in the form $[b]_J$, meaning the $=_J$ equivalence class of some $b \in S$.

For a trivial example of an interpreted structure, if $K \vDash R$ and $a \in K$, then the structure $K \upharpoonright a$ is $\bar{\Sigma}^b_1$ interpreted in K below a, using the normal relation symbols.

Definition 3.3 *If K is a model of R and J is $\bar{\Sigma}^b_1$ interpreted in K, a $\bar{\Sigma}^b_1$ function from (a subset of) K to J is a function of the form*

$$x \mapsto \{y \in S : \phi(x, y)\}$$

for a $\bar{\Sigma}^b_1$ formula ϕ, which maps elements of K to $=_J$-equivalence classes.

Here is a theorem from [12] that sets out some of the consequences of WPHP failing in a structure:

Theorem 3.4 *Suppose $l \in \mathbb{N}$, $l \geqslant 2$. Let K be a model of S^1_0 of the form $[0, a^\varepsilon)$ for some $a, \varepsilon \in K$. Suppose for some $f, s \in \bar{\Sigma}^b_1$, $K \vDash \neg\mathrm{mPHP}^{a^2}_a(f, s)$. Then*

1. *K has an end extension to a model J of S^1_0 of the form $[0, a^{\varepsilon^l})$. Furthermore this end extension is $\bar{\Sigma}^b_1$ interpreted inside K below a.*

2. *If I is any end-extension of K to a model of S^1_0 of the form $[0, a^{\varepsilon^l})$, then I is* relatively categorical *over K with respect to the theory $S^1_0 + (a^{\varepsilon^l} - 1$ is the greatest element). That is, I is isomorphic over K to any end-extension of K to a model of this theory.*

In [12] we used some tools from model theory to give a partial converse to part 2 of this theorem. Here we look for a converse to part 1. We come close with corollary 3.10, but it turns out that a precise converse is impossible as a consequence of Friedman's theorem (see below).

As a warm-up for dealing with interpreted structures and definable isomorphisms, we prove a lemma that we will use later. It has the consequence that if an interpreted structure J is definably isomorphic to an initial segment of K, then that initial segment is unique.

Lemma 3.5 *Suppose $K \vDash S^1_0$ and there is a $\bar{\Sigma}^b_1$ isomorphism σ between $K \restriction a$ and $K \restriction b$. Then σ is the identity function and in particular $a = b$.*

Proof First notice that σ must be the identity at least up to $|a|$, since otherwise there would be a least $i \leqslant |a|$ for which $\sigma(i) \neq i$, which is impossible. Now suppose $x < a$ and $\sigma(x) = y$. Then $|y| \leqslant |b| = \sigma(|a|) = |a|$, and for each $i < |a|$,

$$(K \restriction a \vDash \mathrm{bit}(x, i) = 1) \leftrightarrow (K \restriction b \vDash \mathrm{bit}(y, \sigma(i)) = 1)$$

since σ is an isomorphism. But $\sigma(i) = i$ for all such i. Hence $x = y$. $\square$

The proof of our main result (theorem 3.9) is in two steps. Lemma 3.8 shows that if a small initial segment of J is isomorphic to an initial segment of K then we can extend that isomorphism to one with an exponentially larger domain. We use a similar argument to the proof of the previous lemma, using the $\mathrm{bit}(x, i) = 1$ relation to extend a small isomorphism to a large one. Lemmas 3.6 and 3.7 help with this by showing that under this assumption certain initial segments of J have well-behaved structures and define the $\mathrm{bit}(x, i) = 1$ relation correctly.

In theorem 3.9 we show that if J is defined on a domain inside some initial segment $[0, a)$ of K, and if K is big enough to be able to code sequences of elements in J, then we can find an isomorphism on a small initial segment of J to apply lemma 3.8 to. K is "big enough" if it contains $a^{|a|^{(n)}}$, where $|\ |^{(n)}$ means a nesting of $|\ |$s that is n levels deep. The element $a^{|a|^{(n)}}$ here plays a similar role in the original proof of the weak pigeonhole principle in [8].

Our argument is a more powerful version of the way of constructing an isomorphism sketched out in the introduction, where we first used repeated doubling in J to find the elements corresponding to the powers of 2, then extended this to a complete isomorphism using binary expansions. Here we use repeated adding to find the elements of J corresponding to the logarithmically sized elements of K, then prove that this tells us something about binary expansions in J, then use this to extend the isomorphism.

Lemma 3.6 *Let $K \vDash S^1_0$ be of the form $[0, a)$, and suppose $J \vDash R$ is $\bar{\Sigma}^b_1$ interpretable in K in the sense of definition 3.2 (in particular, there is a* *set S giving the domain of the interpreted structure). Suppose that for some $t \in S$ and some $\varepsilon < |a|$ there is a $\bar{\Sigma}^b_1$ isomorphism from $K \restriction \varepsilon$ onto $J \restriction |[t]_J|$. Then $J \restriction [t]_J \vDash S^1_0$.*

Proof We claim that for each $\bar{\Sigma}^b_1$ formula ϕ, there is a $\bar{\Sigma}^b_1$ formula ϕ_J such that for all $\bar{b} \in S$,

$$J \restriction [t]_J \vDash \phi([\bar{b}]_J) \Leftrightarrow K \vDash \phi_J(\bar{b}),$$

where $[\bar{b}]_J$ stands for $[b_1]_J, \ldots, [b_r]_J$.

We prove this by induction on the quantifier complexity of ϕ. From the definition of $\bar{\Sigma}^b_1$-interpretability we know how to translate open formulas into formulas that are $\bar{\Delta}^b_1$ in S, which is precisely the property required. We can translate $\exists x\, \theta(\bar{y}, x)$ as $\exists x\, (S(x) \wedge x <_J t \wedge \theta_J(\bar{y}, x))$. Lastly, suppose ϕ is of the form $\forall i {<} |[t]_J|^n\, \theta(\bar{y}, i)$, for θ a $\bar{\Sigma}^b_1$ formula and $n \in \mathbb{N}$. We extend our $\bar{\Sigma}^b_1$ isomorphism $K \restriction \varepsilon \cong J \restriction |[t]_J|$ naturally to an isomorphism $K \restriction \varepsilon^n \cong J \restriction |[t]_J|^n$, given by a $\bar{\Sigma}^b_1$ formula χ say, and define

$$\phi_J(\bar{y}) \Leftrightarrow \forall i {<} \varepsilon^n\, \exists x\, (\chi(i, x) \wedge \theta_J(\bar{y}, x)).$$

To show that $\bar{\Sigma}^b_1$−LIND holds in $J \restriction [t]_J$, suppose ϕ is a $\bar{\Sigma}^b_1$ formula, $n \in \mathbb{N}$ and $|[t]_J|^n$ exists in J. Let χ be a $\bar{\Sigma}^b_1$ formula defining the isomorphism $K \restriction \varepsilon^n \cong J \restriction |[t]_J|^n$. We will write $\phi_J(\chi(i))$ as shorthand for $\exists x\, (\chi(i, x) \wedge \phi_J(x))$. Suppose

$$J \restriction [t]_J \vDash \phi(0) \wedge \forall i {<} |[t]_J|^n\, (\phi(i) \rightarrow \phi(i+1)).$$

Then

$$K \vDash \phi_J(\chi(0)) \wedge \forall i {<} \varepsilon^n\, (\phi_J(\chi(i)) \rightarrow \phi_J(\chi(i+1))).$$

Hence $K \vDash \phi_J(\chi(\varepsilon^n))$, so $J \restriction [t]_J \vDash \phi(|[t]_J|^n)$. □

Lemma 3.7 *If $J \vDash S^1_0$ is $\bar{\Sigma}^b_1$ defined in $K \vDash S^1_0$, then the relations $(x = 2^i)_J$ and $(\mathrm{bit}(x, i) = 1)_J$ are $\bar{\Delta}^b_1$ in S.*

Proof We need to show that the formulas defining these relations do not use any sharply bounded universal quantifiers, since these would not in general translate into sharply bounded quantifiers in K. Then we can translate existential quantifiers over J as existential quantifiers over S. We use the following definitions:

1. The relation $\mathrm{parity}(x) = \delta$ is given by
$$(\delta = 1 \wedge \exists y\, (2 \cdot y + 1 = x)) \vee (\delta = 0 \wedge \exists y\, (2 \cdot y = x)).$$
2. The relation $2^i = x$ is given by
$$\exists y\,,\ |y| = i \wedge |x| = i + 1 \wedge x = y + 1.$$
3. The relation $\mathrm{decomp}(x, i) = (y, z)$ is given by
$$|y| \leqslant i \wedge x = y + 2^i \cdot z.$$
4. The relation $\mathrm{MSP}(x, i) = z$ (standing for Most Significant Part) is given by
$$\exists y\, \mathrm{decomp}(x, i) = (y, z).$$
5. The relation $\mathrm{bit}(x, i) = \delta$ is given by
$$\delta = \mathrm{parity}(\mathrm{MSP}(x, i - 1)).$$

These can be proved to have the right properties in J because $J \models S^1_0$. □

Lemma 3.8 *Suppose $K \vDash S^1_0$ is of the form $[0, a)$, $J \vDash R$ is $\bar{\Sigma}^b_1$ interpreted in K and for some $n \in \mathbb{N}$ there is a $\bar{\Sigma}^b_1$ isomorphism between $K \restriction |a|^{(n)}$ and an initial segment of J. Then there is a $\bar{\Sigma}^b_1$ isomorphism, either from all of K onto an initial segment of J, or from an initial segment of K onto all of J.*

Proof We extend a small isomorpism to a larger isomorphism in the same way as in lemma 3.5, again taking advantage of the fact that a number is uniquely described by its sequence of bits.

Let t be (a representative of) the $<_J$-greatest element of J. We will inductively construct $\bar{\Sigma}_1^b$ isomorphisms with domains $|a|^{(n-1)}, \dots, |a|, a$ stopping if at any point we reach t and exhaust J.

For the inductive step, suppose that $\phi(x, y)$ is a $\bar{\Sigma}_1^b$ formula giving an isomorphism from $K \restriction |a|^{(m)}$ onto an initial segment of J.

Let $i < |a|^{(m)}$ be greatest such that

$$2^i < |a|^{(m-1)} \wedge \exists x \, \exists y \, , \, S(x) \wedge S(y) \wedge \phi(i, x) \wedge (|y| = x + 1)_J$$

and let r be some such y. Then $J \restriction |[r]_J| \vDash S_0^1$ (since it is isomorphic to an initial segment of K) so by lemma 3.6, $J \restriction [r]_J \vDash S_0^1$. So we can choose r so that it is a power of 2 in J, and the equivalence class of r is the element of J corresponding to 2^i in K. Hence 2^i is the greatest power of 2 which exists in both $K \restriction |a|^{(m-1)}$ and J (in some sense).

Define $\theta(x, y)$ as

$$\begin{aligned} & x < 2^i \wedge y <_J r \wedge S(y) \wedge \\ & \quad \forall 1 \leqslant j \leqslant i \, \exists z \, , \, S(z) \wedge \phi(j, z) \wedge (\text{bit}(x, j) = 1 \leftrightarrow (\text{bit}(y, z) = 1)_J). \end{aligned}$$

We claim that $\sigma : x \mapsto \{y : \theta(x, y)\}$ is an isomorphism from $K \restriction 2^i$ onto $J \restriction [r]_J$.

To show well-definedness, suppose $y, y' <_J r$ with $y \neq_J y'$. Then, since $J \restriction [r]_J \vDash S_0^1$, without loss of generality we have $(\text{bit}(y, v) = 1)_J$ and $(\text{bit}(y', v) \neq 1)_J$ for some v such that $J \vDash 1 \leqslant [v]_J \leqslant |[r]_J|$. Since ϕ defines an isomorphism, $\phi(j, v)$ holds for some $1 \leqslant j \leqslant i$. Hence if for some x, x' we have $[y]_J = \sigma(x)$ and $[y']_J = \sigma(x')$, we must have $\text{bit}(x, j) \neq \text{bit}(x', j)$, so $x \neq x'$. We show that σ is injective in a similar way.

To show that σ is defined on all of $K \restriction 2^i$, let $x < 2^i$ and let $\chi(j)$ be the formula

$$\begin{aligned}
&\exists y \,, S(y) \wedge \forall 1 \leqslant k \leqslant i \, \exists z \,, \; S(z) \wedge \phi(k, z) \\
&\qquad \wedge [k \leqslant j \rightarrow (\mathrm{bit}(x, k) = 1 \leftrightarrow (\mathrm{bit}(y, z) = 1)_J)] \\
&\qquad \wedge [j < k \rightarrow (\mathrm{bit}(y, z) \neq 1)_J]
\end{aligned}$$

expressing that some $[y]_J$ is the correct image of x up to its jth bit and the remaining bits are 0. Then $\chi(0)$ holds, and if for any $j < i$ we have that $\chi(j)$ holds and is witnessed by y, we can find the element of J corresponding to 2^j and, depending on $\mathrm{bit}(x, j+1)$, let y' be either y or $(y + 2^j)_J$ (this sum exists in J and is not too big, because $J \restriction [r]_J \vDash S^1_0$). Then y' witnesses that $\chi(j+1)$ holds. Hence by $\bar{\Sigma}^b_1$−LIND in K, $\chi(i)$ holds and the set $\sigma(x)$ is not empty. Similarly we use comprehension in K to show that σ is a surjection.

Finally, since we can define all our relations bitwise, σ is an isomorphism.

To extend σ to the rest of $K \restriction |a|^{(m-1)}$, notice that by our choice of i either $2^i \geqslant |a|^{(m-1)}/2$ or $J \vDash [r]_J \geqslant [t]_J/2$. So we map $x \geqslant 2^i$ to the set

$$\{y \in S : \exists z \,, \phi(x - 2^i, z) \wedge (y = r + z)_J\}$$

if this is non-empty, which it will be until we reach the top element t of J. □

Theorem 3.9 *Suppose $K \vDash S^1_0$ is of the form $[0, b)$, and $a, a^\varepsilon \in K$ where $\varepsilon > |b|^{(n)}$ for some $n \in \mathbb{N}$. Suppose $J \vDash R$ is $\bar{\Sigma}^b_1$ defined in K below a. Then there is a $\bar{\Sigma}^b_1$ isomorphism, either from all of K onto an initial segment of J, or from an initial segment of K onto all of J.*

Proof Let $\theta(i, w)$ be the following formula, where we use a number w to code a sequence $w_0, \ldots, w_i$ of elements of $[0, a)$ via its base a expansion:

$$\begin{aligned}
\forall j, k, l \leqslant i \,, \; & (w_j + w_k = w_l)_J \leftrightarrow j + k = l \\
& \wedge (w_j \cdot w_k = w_l)_J \leftrightarrow j \cdot k = l \\
& \wedge w_j <_J w_k \leftrightarrow j < k \\
& \wedge (|w_j| = w_k)_J \leftrightarrow |j| = k \\
& \wedge w_0 =_J 0_J .
\end{aligned}$$

Let $i < \varepsilon$ be greatest such that $\exists w \leqslant a^i\, \theta(i, w)$ and let $t = w_i$. We must have $i \geqslant 1$, since we can set $w_0 = 0_J$ and $w_1 = 1_J$. Let $\phi(x, y)$ be the formula $S(y) \wedge y =_J w_x$. We claim that $\sigma : x \mapsto \{y : \phi(x, y)\}$ is an isomorphism from $K \restriction [0, i]$ onto $J \restriction [0, [t]_J]$.

It is sufficient to show that σ is surjective. Suppose it is not, and for some $s \in S$ we have $s <_J t$ and $\forall j \leqslant i\, \neg\phi(j, s)$. Let $j \leqslant i$ be greatest such that $w_j <_J s$. Then $j < i$ and $w_{j+1} \geqslant_J s$. But $J \vDash [w_{j+1}]_J = [w_j]_J + 1$, so $J \vDash [w_{j+1}]_J = [s_j]_J$ and hence $w_{j+1} =_J s$, a contradiction.

If $i < \varepsilon - 1$, then t must be the $<_J$ greatest element of J (otherwise we could add an extra element to w). Hence we have constructed an isomorphism from an initial segment of K onto all of J.

If $i = \varepsilon - 1$, then we have an isomorphism from $K \restriction |b|^{(n)}$ onto an initial segment of J and can use lemma 3.8. □

Corollary 3.10 *Suppose $K \vDash S^1_0$ is of the form $[0, a^\varepsilon)$, for $\varepsilon = |a|^{(n)}$ some $n \in \mathbb{N}$, and that K is isomorphic to a structure J that is $\bar{\Sigma}^b_1$ defined in K below a. Suppose further that K is not isomorphic to any proper initial segment of K. Then there is a $\bar{\Sigma}^b_1$ formula ϕ giving an isomorphism $J \cong K$. In particular, ϕ maps equivalence classes of $S \subseteq K \restriction a$ (where S is the set on which J is defined) bijectively onto K. Inverting this mapping we obtain a $\bar{\Sigma}^b_1$ multifunction mapping K injectively into S, hence (multifunction)* WPHP *fails in K. Furthermore, if S is all of $K \restriction a$, then we get a surjection $a \twoheadrightarrow a^\varepsilon$; if each $=_J$ equivalence class has precisely one member, then we get an injection $a^\varepsilon \hookrightarrow a$.*

We cannot do without the condition "K is not isomorphic to any proper initial segment of K" because otherwise we have the following counterexample: let M be any countable nonstandard model of PA. By Friedman's theorem [4], there exist $a, b \in M$ with $M \restriction a^{|a|} \cong M \restriction b^{|b|}$ and $a^{|a|} < b$. Hence the structure defined inside $M \restriction b$ on the set $[0, a^{|a|})$ by the normal relations is isomorphic to $M \restriction b^{|b|}$; but the weak pigeonhole principle does not fail in M.

Corollary 3.11 *If $K \vDash \mathrm{PA}^{\mathrm{top}}$ is of the form $[0,a)$ and is not isomorphic to any proper initial segment of itself, then for all $n \in \mathbb{N}$ no end-extension of K to a model of $\mathrm{PA}^{\mathrm{top}}$ of the form $[0,a^{|a|^{(n)}})$ is definable in K.*

Proof All the relevant results above go through if we use formulas of unrestricted quantifier complexity in the place of Σ_1^b formulas. Then use the fact that we can extend models of $\mathrm{PA}^{\mathrm{top}}$ to models of $\mathrm{I}\Delta_0$ and that $\mathrm{I}\Delta_0 + (a^{|a|^{(n)}}\text{exists})$ proves $\mathrm{PHP}_a^{a^2}(\Delta_0)$ [8]. □

It would be interesting to see how much of an increase in size is necessary to construct our isomorphism. For example, can we use an exponent smaller than any nesting of logs $|a|^{(n)}$ in corollary 3.11? Would an exponent of 1, and so no extra space for coding at all, be sufficient? It seems as though this should be related to the provability of WPHP in $\mathrm{I}\Delta_0$.

We can interpret theorem 3.9 together with lemma 3.5 as saying that a $\bar{\Sigma}_1^b$ set S in a model of S_0^1 is either bigger than the model or has a unique precise size in the model, provided of course that S comes with lots of structure and that we take counting statements to be about the existence of isomorphisms, rather than just bijections. In some ways this is a natural step, similar to moving from cardinal to ordinal numbers by adding an ordering relation.

If we are in a model of S_2^1, then the smash function guarantees that we have the space to code short sequences of elements of S and construct our isomorphism. If we also know that the weak pigeonhole principle holds, this guarantees that our set S is not "bigger" than the original model. We summarise this as: in a model of S_2^1 satisfying WPHP we can precisely count structured sets. We make this precise below, using the injective WPHP. There are similar results for surjective or multifunction WPHP.

We will say that a Σ_1^b set S is *structured* if it is bounded and there are relations $<_S, |\ |_S, +_S, \cdot_S$ that are Δ_1^b in S such that $\langle S, <_S, |\ |_S, +_S, \cdot_S \rangle \vDash R$.

Corollary 3.12 *Let $M \vDash S_2^1 + \forall x\, \mathrm{PHP}_x^{x^2}(\Sigma_1^b)$ and suppose S is a structured Σ_1^b subset of M, with relations $<_S, |\ |_S, +_S, \cdot_S$. Then there exists a unique*

$b \in M$ for which there is a Σ_1^b function $f : \langle S, <_S, |\ |_S, +_S, \cdot_S \rangle \cong M \restriction b$.

Proof Suppose S is bounded by a. Notice that we are using Σ_1^b sets here, where theorem 3.9 applies to $\bar{\Sigma}_1^b$ sets. However, since we are only interested in subsets of $[0, a)$ and the quantifiers in a Σ_1^b formula are bounded by terms, we can find b in M such that all of the sets considered are $\bar{\Sigma}_1^b$ definable inside $M \restriction b$. Let c be greater than both b and $a^{|a|}$. Let $K = M \restriction c$, and apply theorem 3.9. If there is a $\bar{\Sigma}_1^b$ isomorphism from S onto an initial segment of K, then we are done. If not, then there is a $\bar{\Sigma}_1^b$ isomorphism from K onto an initial segment of S, and hence there is a Σ_1^b injection $c \hookrightarrow a$, violating WPHP. □

Theorem 3.9 also holds in the relativized case, although we have to be careful about the classes of formulas for which induction holds in our different structures.

Theorem 3.13 *Let α be a set of new relation and function symbols. Suppose $\langle K, \alpha \rangle \vDash S_0^1(\alpha)$ is of the form $[0, b)$, and $a, a^\varepsilon \in K$ where $\varepsilon > |b|^{(n)}$ for some $n \in \mathbb{N}$. Suppose $J \vDash R$ is $\bar{\Sigma}_1^b(\alpha)$ defined in $\langle K, \alpha \rangle$ below a. Then there is a $\bar{\Sigma}_1^b(\alpha)$ isomorphism, either from all of K (without α) onto an initial segment of J, or from an initial segment of K onto all of J.*

Corollary 3.14 *Let α be a set $\{+^*, \cdot^*, <^*, |\ |^*, 0^*, 1^*, 2^*\}$ of relation and constant symbols of the same form as but disjoint from our normal language of arithmetic. Let R^* and S_0^{1*} be our normal theories re-written in this language. Then "every finite model of R is a model of S_0^1," or*

$$\forall a\, ,(\langle [0, a), \alpha \rangle \vDash R^*) \rightarrow (\langle [0, a), \alpha \rangle \vDash S_0^{1*}),$$

is provable in $S_2^1(\alpha) + \forall a\, \mathrm{PHP}_a^{a^2}(\Sigma_1^b(\alpha))$ but not in $S_2^2(\alpha)$.

Proof The independence from $S_2^2(\alpha)$ follows from theorem 1.1 since there is an infinite model of R that is not a model of S_0^1.

Now suppose that in a model M of $S^1_2(\alpha)+\forall a\,\mathrm{PHP}^{a^2}_a(\Sigma^b_1(\alpha))$ the structure $J=\langle[0,a),\alpha\rangle$ is a model of R^*. Then by the relativized version of corollary 3.10, J is definably isomorphic to an initial segment of M (in the normal language, without α). Hence J is a model of S^{1*}_0. □

One conclusion we can draw is that if we are looking for independence results for theories as strong as or stronger than S^3_2 (in which $\mathrm{WPHP}(\Sigma^b_1)$ is provable), then there is a class of principles (of the form $\alpha\vDash R\rightarrow\Phi(\alpha)$) where we gain nothing by considering the relativized case. This is because if we construct a model of $S^3_2(\alpha)$ in which the structure given by α is a model of R but not of Φ, then Φ must already by false in the unrelativized model (with α removed), since part of this model is isomorphic to the structure α.

It would be interesting to find other weak theories that work in the place of R to give similar results. One possibility would be a universally axiomatized theory of "discretely ordered abelian groups with a greatest element" in the language $\{0,1,e,<,+,-,\lfloor\frac{x}{2}\rfloor,\mathrm{parity}\}$ (with some sort of modulo addition); another would be a fragment of set theory, where we have set membership in place of the relation $\mathrm{bit}(x,i)=1$.

The results in the next section concern what cannot be proved about a structure given by oracles, even if we use WPHP, and may be useful for showing that certain theories are too weak to work in place of R.

4 Witnessing with a probabilistic machine

We look for some relativized independence results from theories of bounded arithmetic that include the pigeonhole principle. The strongest theory we can practically use here is (the relativized form of) $S^1_2+\forall a\,\mathrm{PHP}^a_{a^2}(\mathrm{PV})$ because we have a witnessing theorem for it, theorem 4.2 below. At the end of this section we note that we can almost extend independence from this theory to independence from $S^1_2+\forall a\,\mathrm{PHP}^a_{a^2}(\Sigma^b_1)$, which would match more closely the results in the previous section. By "almost" we mean that if $S^1_2 + \forall a\, \mathrm{PHP}^a_{a^2}(\mathrm{PV})$ does not prove $\forall \bar{x}\, \exists y\, \theta(\bar{x}, y)$, then S^1_2 together with WPHP for Σ^b_1 formulas does not prove $\forall \bar{x}\, \exists y\, \theta(\bar{x}, y)$, provided that WPHP is only applied to formulas containing only the parameters $\bar{x}$ that appear in θ.

Theorem 4.2 is due to Wilkie and was first published in [6]. We give an alternative proof here, using a technical lemma implicit in [12]:

Lemma 4.1 *For any* PV *function symbol $f(c, x)$, there is a* PV *function symbol $G(x)$ (with no other parameters) such that*

$$S^1_2 \vdash \forall b,\ \exists 1 < a < b\, \exists c < b\, \forall y < a^2\, \exists x < a\, f(c, x) = y \to \forall y < b^8\, \exists x < b^4\, G(x) = y.$$

That is, if $f(c, \_)$ violates surjective WPHP *somewhere below b then G violates surjective* WPHP *at b^4.*

Proof Suppose $f(c, x)$ is a surjection $a \twoheadrightarrow a^2$ (x here is a placeholder). Then by corollary 2.2 of [12] we have a PV surjection $F(c, b^8, a, x) : a \twoheadrightarrow b^8$, with only the parameters shown. Define G so that

$$G : (x_1, x_2, x_3, x_4) \mapsto F(x_1, (x_2 + 1)^8, x_3, x_4).$$

Since c, $b - 1$ and a are all less than b, the range of $F(c, b^8, a, x)$ on $[0, a)$ is contained in the range of $G(\bar{x})$ on $[0, b)^4$. □

Theorem 4.2 *If $S^1_2 + \forall a\, \mathrm{PHP}^a_{a^2}(\mathrm{PV}) \vdash \forall x\, \exists y\, \theta(x, y)$, for θ a Σ^b_1 formula, then there is a probabilistic polynomial time Turing machine which, for any input x, outputs with probability at least $2/3$ some y such that $\mathbb{N} \vDash \theta(x, y)$.*

Proof Let $u(e, w, t)$ be the universal PV function symbol, calculating the output of the program with code e run for time $|t|$ on input w. Suppose

$$S^1_2 + \forall a\, \forall e\, \forall t\, \mathrm{PHP}^a_{a^2}(u(e, w, t)) \vdash \forall x\, \exists y\, \theta(x, y),$$

where w is a placeholder. Moving WPHP to the right hand side and using Parikh's theorem we have that for some $k \in \mathbb{N}$,

$$S^1_2 \vdash \forall x\, ,\ (\exists a, e, t < 2^{|x|^k}\, \forall v < a^2\, \exists w < a\, u(e, w, t) = v) \vee \exists y\, \theta(x, y).$$

By lemma 4.1 there is $G \in L_{\mathrm{PV}}$ such that if the universal function symbol defines a surjection $a \twoheadrightarrow a^2$ for some $a < 2^{|x|^k}$ using parameters $e, t < 2^{|x|^k}$ then G defines a surjection $(2^{|x|^k})^4 \twoheadrightarrow (2^{|x|^k})^8$ using no parameters. So

$$S^1_2 \vdash \forall x\,,\ [\forall v < 2^{8|x|^k}\, \exists w < 2^{4|x|^k}\, G(w) = v] \vee \exists y\, \theta(x, y).$$

Hence by Buss' witnessing theorem [3] there are PV functions g_0 and g_1 such that

$$\mathbb{N} \models \forall x\, \forall v < 2^{8|x|^k}\,,\ g_0(x, v) < 2^{4|x|^k} \wedge (G(g_0(x, v)) = v \vee \theta(x, g_1(x, v))).$$

So given x, if we choose v at random in $[0, 2^{8|x|^k})$, with high probability we will have $\theta(x, g_1(x, v))$ since in the standard model very few of the elements of $[0, 2^{8|x|^k})$ will be in the range of G on the domain $[0, 2^{4|x|^k})$. □

This theorem relativizes with no significant changes to the proof.

It follows from this theorem that any set Δ^b_1-definable in $S^1_2 + \forall a\, \mathrm{PHP}^a_{a^2}(\mathrm{PV})$ is in the complexity class **RP** ∩ **coRP**. However the converse is unlikely to hold since this would imply that **RP** ∩ **coRP** has a complete language, which is not true in a relativized world [2]. This is shown in some detail in [11]. See also [5] for more connections between this theory and probabilistic complexity.

Definition 4.3 *Suppose that α is a tuple of functions, relations and constants on a domain $[0, a)$. We say that a Turing machine M is given a structure $K = \langle [0, a), \alpha \rangle$ as input if it starts with the number a written on its input tape and is given access to an oracle for the relations and functions in α, where we treat constants as constant valued functions.*

What can a probabilistic machine tell in polynomial time about a structure it is given as input? This sort of question has been studied in cryptography, where the functions are usually algebraic and the structures are called "black box" groups or fields [10, 1]; and in the design of sublinear algorithms, where a machine does not have the time to look at all of its input but only a random sample (some particular applications of theorem 4.7 are folklore in this field). We give a (weak) general result. Intuitively, it seems likely that all such a machine can do is choose a tuple of elements of the structure at random, then work on those by applying relations and functions from the structure to them. We prove that this is indeed the case.

Definition 4.4 *Suppose $a \in \mathbb{N}$, and $K = \langle [0, a), \bar{r}, \bar{f}, \bar{c} \rangle$ is a structure with a finite number of relations, functions and constants $\bar{r}, \bar{f}, \bar{c}$.*

If $\bar{x}, \bar{y}$ are tuples in $[0, a)$, and $t \in \mathbb{N}$, we say that $\bar{y}$ is derivable from $\bar{x}$ by a straight-line program in K of length t if there is a sequence $w_1, \ldots, w_t$ of elements of $[0, a)$ which contains every element in the tuple $\bar{y}$ and is such that every element of $\bar{w}$ is either an element of the tuple $\bar{x}$, or the interpretation of a constant from $\bar{c}$, or is derived from earlier elements in the sequence $\bar{w}$ by applying a function from $\bar{f}$.

We consider a probabilistic machine running for time t on a structure K, with the goal of outputting a tuple in K from some given set W of distinguished tuples. We show that the machine can do no better than choose t elements $x_1, \ldots, x_t$ at random and apply a straight-line program of length t. In practice the machine would use queries about relations to decide which program to apply to $\bar{x}$; our result here is rather weak because we effectively assume that the machine always automatically knows the best possible program (which is why relations do not play a rôle in the above definition).

Our strategy is to produce a large set of structures as permutations of some given structure; if there is a probabilistic machine that succeeds with high probability for all these permutations, there must be a deterministic machine (essentially a branching program) that succeeds with high probability on a random permutation from our set. We use this to show that if we run our deterministic machine and make sure that the only elements of the structure that it has access to are either chosen at random or by applying functions from $\bar{f}$, then it still succeeds with high probability, and this gives us our result.

Definition 4.5 *Let S_a be the set of all permutations of $[0, a)$. Let K be a structure as above. For $\sigma \in S_a$ we define K^σ, the permutation of K by σ, to be the structure $\langle [0, a), \bar{r}^\sigma, \bar{f}^\sigma, \bar{c}^\sigma \rangle$. Here, for each relation r_i and tuple $\bar{x} \subseteq [0, a)$, $r_i^\sigma(\bar{x})$ if and only if $r_i(\sigma(\bar{x}))$. Similarly $f_i^\sigma(\bar{x}) = y$ if and only if $f_i(\sigma(\bar{x})) = \sigma(y)$, and $c_i^\sigma = \sigma^{-1}(c_i)$.*

Lemma 4.6 (Birthday inequality) *If $0 \leq t < a$, then $(1 - t/a)^t \geq 1 - t^2/a$.*

Theorem 4.7 *Suppose $a \in \mathbb{N}$, and $K = \langle [0, a), \bar{r}, \bar{f}, \bar{c} \rangle$ is a structure as above. Let W be any set of tuples from $[0, a)$, and for $\sigma \in S_a$ let $W^\sigma = \{\bar{x} \subseteq [0, a) : \sigma(\bar{x}) \in W\}$. Let $t^2 < a$.*

Suppose that if we choose t elements from $[0, a)$ at random (with replacements) then with probability at least q no tuple in W can be derived from these elements by any straight-line program in K of length t. Then there is no probabilistic machine M such that for all $\sigma \in S$, M runs on K^σ for time t and ouputs a tuple in W^σ with probability at least $1 - q + t^2/a$.

Proof Suppose such a machine M does exist. By Yao's minimax principle, if for every permutation $\sigma \in S_a$ the machine succeeds for a fraction $1 - q + t^2/a$ of its possible sequences of coin tosses, then there must be some sequence c of coin tosses with which the machine succeeds for at least $(1 - q + t^2/a)|S_a|$ of the possible permutations σ. Let M_c be the branching program which simulates M with coin tosses c.

We will obtain a contradiction by constructing a large number of σs for which M_c fails.

The only part of M_c we consider is a combined oracle query and oracle reply tape, which at the end of step i of the computation will contain an element w_i of $[0, a)$. We allow M_c to do one of three things at each step $i + 1$:

1. Write down some number w_{i+1} on the tape;

2. Query $[r_k(w_{i-l+1}, \ldots, w_i)?]$ and expect w_{i+1} to be 1 (for "yes") or 0 (for "no") accordingly, where r_k is a relation symbol of arity l;

3. Query $[f_k(w_{i-l+1}, \ldots, w_i) = ?]$ and expect w_{i+1} to be the correct answer, where f_k is a function symbol of arity l. We treat constants in the language of K as constant valued functions.

We assume that everything output by the machine appears at some point on the query tape. Note that in order to treat uniformly all the information passed to and from the oracle, the oracle replies to queries about relations with a number $0/1$ rather than with an answer "yes/no", and that this reply is treated the same as any other number appearing on the tape. For our proof to work with this simplifying assumption we also assume that the first two actions of the machine are to put $w_1 = 0$ and $w_2 = 1$, which we can do without loss of generality.

We present our strategy for constructing σ as a probabilistic argument. The first step is to choose numbers $u_1, \ldots, u_t$ uniformly at random in $[0, a)$, with repetitions allowed. We will show that with high probability $u_1, \ldots, u_t$ can be used to construct a partial permutation σ^* of size t. Then we will extend this at random to a total permutation on which, with high probability, M_c will fail.

To construct σ^*, first set $\sigma_0 = \varnothing$ and begin a computation of M_c. As we go along we will define an increasing sequence σ_i of partial permutations and we will use up the numbers $u_1, \ldots, u_t$. Suppose that at the end of step i in the computation σ_i is a partial permutation of $[0, a)$, defined on $w_1, \ldots, w_i$ (this list may contain repetitions) and nowhere else.

The definition of σ_{i+1} depends on the next action M_c takes.

1. If M_c writes down an element w_{i+1}, let $\sigma_{i+1} = \sigma_i$ if w_{i+1} has already occurred on the list. If w_{i+1} is new, let y be the first element of $u_1, \ldots, u_t$ that we have not yet used, and let $\sigma_{i+1} = \sigma_i \cup \{\langle w_{i+1}, y \rangle\}$. If σ_{i+1} is no longer a partial permutation, abandon the construction.

2. If $[r_k(w_{i-l+1}, \ldots, w_i)?]$ is queried set w_{i+1} to be 0 or 1, depending on the truth of $r_k(\sigma_i(w_{i-l+1}), \ldots, \sigma_i(w_i))$. Let $\sigma_{i+1} = \sigma_i$.

3. If $[f_k(w_{i-l+1},\dots,w_i) =?]$ is queried, let $y = f_k(\sigma_i(w_{i-l+1}),\dots,\sigma_i(w_i))$. If $y = \sigma_i(x)$ for some x, set $w_{i+1} = x$ and let $\sigma_{i+1} = \sigma_i$. Otherwise let x be the first element of $u_1,\dots,u_t$ that we have not yet used, set $w_{i+1} = x$ and let $\sigma_{i+1} = \sigma_i \cup \{\langle x, y\rangle\}$. If σ_{i+1} is no longer a partial permutation, abandon the construction.

Before we estimate the probability that we can successfully complete the construction, we want to make sure that we use up all of the sequence $u_1,\dots,u_t$, and that our partial permutation has size exactly t (both of these are to make the counting argument work smoothly).

If we have so far only used elements $u_1,\dots,u_s$ then, by the construction, σ_t has size exactly s. We now take the smallest element of $[0,a) \setminus \mathrm{dom}(\sigma_t)$ and map it u_{s+1}, to get σ_{t+1}. Then we map the smallest element of $[0,a) \setminus \mathrm{dom}(\sigma_{t+1})$ to u_{s+2} to get σ_{t+2}, and so on. We abandon the construction if at any point σ_{t+i} stops being a partial permutation. This process is essentially the same thing as adding $t-s$ dummy steps of type 1. We let σ^* be the permutation we get at the end, which will have size exactly t.

We finish by extending σ^* uniformly at random to a total permutation σ.

The probability that this construction succeeded and we did not have to abandon it is bounded above by the probability that each of the (independently chosen) elements $u_1,\dots,u_t$ was outside a set of size $\leq t$ (the set was $\mathrm{ran}(\sigma_i)$ for steps of type 1, and $\mathrm{dom}(\sigma_i)$ for steps of type 2). This probability is at least$(1-t/a)^t > 1-t^2/a$.

Furthermore the set $\mathrm{ran}(\sigma^*)$ is the result of $\leq t$ random independent choices from $[0,a)$ (these were the elements of $u_1,\dots,u_t$ that were used in steps of type 1) augmented by $\leq t$ applications of functions in K (from steps of type 3). Hence with probability at least q, no tuple in W can appear in $\mathrm{ran}(\sigma^*)$.

So with probability at least $q - t/a^2$, we have both that (1) σ^* is a partial permutation and that (2) M^* run on K^σ cannot output any tuple in W^σ.

Now there were a^t different ways of choosing $u_1,\dots,u_t$ and $(a-t)!$ different ways of extending σ^* to a total permutation σ. So there are $a^t(a-t)!$ different sequences of random choices we could have made to construct σ. Suppose X and Y are two distinct such sequences, which both allow us to construct permutations, respectively σ and τ. Then $\sigma \neq \tau$, which can be seen by considering the pairs added in the constructions of σ and τ at the first step at which X and Y differ.

Hence we have constructed at least $(q - t^2/a)a^t(a-t)! > (q - t^2/a)a! = (q - t^2/a)|S_a|$ permutations σ such that M_c fails on K^σ, as required. $\square$

Corollary 4.8 *The theory* $S^1_2(<^*) + \forall x\, \mathrm{PHP}^x_{x^2}(\mathrm{PV}(<^*))$ *does not prove that every total order* $<^*$ *has a least element on every interval* $[0,a)$.

Proof Suppose

$$\forall a\,,\ (<^* \text{ is a total order on } [0,a)) \rightarrow \exists x\!<\!a\, \forall y\!<\!a\, (x \neq y \rightarrow x <^* y)$$

is provable in the theory. If we introduce a Herbrand function p to replace the universal quantifier $\forall y$, we have that

$$\begin{aligned}&\forall a\,,\ (<^* \text{ is a total order on } [0,a))\\ &\qquad\qquad\qquad \rightarrow \exists x\!<\!a\, (p(x) < a \wedge x \neq p(x) \rightarrow x <^* p(x))\end{aligned}$$

is provable in the theory. This is a $\Sigma^b_1(<^*, p)$ formula, so by theorem 4.2 there is a probabilistic polynomial time machine M which will, when equipped with oracles for $<^*$ and p and given an input a, find witnesses with high probability.

Now choose a so that it is much bigger than the running time $t = |a|^k$ of the machine. Let $K = \langle [0,a), <, P\rangle$ where P is the predecessor function. Then for any permutation σ, the relation $<^\sigma$ is a total ordering so our machine run on K^σ should output 0^σ as the the only possible witness to the sentence above.

However if $\bar{v}$ is any tuple of size $\leq t$ in K, 0 is derivable from $\bar{v}$ by a straight-line program in K of length t only if $\bar{v}$ contains an element in the range $0, \ldots, t-1$, because the most a straight-line program of length t can do is apply P t-many times. If $\bar{v}$ is chosen at random this happens with probability $\leq 1-(1-t/a)^t < t^2/a$.

So by theorem 4.7 our probabilistic machine on input K^σ, taking W as the set $\{0^\sigma\}$, cannot output a witness with high probability, a contradiction. $\square$

In more generality we have that if there is a large structure α in which a witness to a sentence $\exists \bar{x}\, \theta(\bar{x})$ is hard to find, then $S^1_2(\alpha) + \forall x\, \mathrm{PHP}^x_{x^2}(\mathrm{PV}(\alpha))$ does not prove that such a witness always exists in a finite structure. Here "hard to find" means "hard to find by choosing some elements at random and then applying functions from the structure to them". Notice this also allows us to prove independence of sentences of the form $\forall \bar{y}\, \exists \bar{x}\, \theta(\bar{x}, \bar{y})$, by treating the parameters $\bar{y}$ as constants in our structure. Formally,

Corollary 4.9 *Let $\theta(\bar{w})$ be a formula in a language α disjoint from our usual language for arithmetic. Suppose that for all $k \in \mathbb{N}$ there is $a \in \mathbb{N}$ and a structure $K = \langle [0,a), \alpha \rangle$ such that if a $|a|^k$-tuple $\bar{x}$ is chosen at random from $[0,a)$ then with probability at least $2/3$ there is no $\bar{w}$ satisfying $\theta(\bar{w})$ derivable from $\bar{x}$ by a straight-line program in K of length $|a|^k$. Then $S^1_2(\alpha) + \forall x\, \mathrm{PHP}^x_{x^2}(\mathrm{PV}(\alpha))$ does not prove $\forall a$, $\langle [0,a), \alpha \rangle \vDash \exists \bar{w}\, \theta(\bar{w})$.*

It would be interesting to match this more closely to theorem 1.1, by finding a more elegant condition or by strengthening the induction allowed from $S^1_2(\alpha)$ to $S^2_2(\alpha)$.

Notice that this result does not apply to the kinds of structures dealt with in section 3. Firstly we would have to introduce some skolem function symbols to make R into a universal theory, which we would need to do to work with it in the framework of PV functions. We would then end up with a constant symbol for 0, a function symbol for successor and a function symbol for doubling (or something equivalent to these), and these allow us to reach any element of the structure in logarithmically many steps.

In some cases we have easy ways of constructing a large number of models of a theory, so do not have to resort to the trick above of only considering isomorphic copies of a single structure. This can give apparently stronger results:

Theorem 4.10 $S^1_2(f) + \forall x\, \mathrm{PHP}^{x}_{x^2}(\mathrm{PV}(f))$ *does not prove* $\forall y\, \mathrm{PHP}^{y^2}_{y}(f)$.

Proof Suppose the theorem fails. Then there is a probabilistic machine that, given a structure $\langle [0,a), f\rangle$ as input, where f is a binary function symbol, outputs with high probability distinct pairs (x_1, x_2) and (y_1, y_2) such that $f(x_1, x_2) = f(y_1, y_2)$.

We use the same idea as in theorem 4.7, but rather than construct our counterexample structures as permutations of a given structure we simply choose a sufficiently large and consider the set of all possible binary functions on $[0, a)$. By Yao's principle again, if our machine M succeeds with high probability on every such function, there must be a sequence of coin tosses c which succeeds with high probability if we choose a function at random.

Let f be a binary function chosen at random, and let t be the running time of the machine. As it runs the machine can make at most t oracle queries $[f(z_1, z_2) =?]$. The probability that the machine finds two pairs that f maps to the same number is the probability that t elements chosen independently at random from $[0, a)$ contain a repetition, which is bounded by $1 - (1 - t/a)^t < t^2/a$.

So if we choose a sufficiently large, with high probability M_c fails on a randomly chosen function. Hence there can be no such machine M. □

Lemma 4.11 below relates provability from the surjective WPHP for PV functions and provability from the surjective WPHP for Σ^b_1 formulas, and brings the independence results from WPHP(PV) in this section a bit closer to the proofs using WPHP(Σ^b_1) in the previous section. The proof is a application of the witnessing theorem for S^1_2, along the lines of the proof of theorem 4.2.

Lemma 4.11 *Suppose $\chi(\bar{b}, u, v)$ is a Σ_1^b formula, which we will treat as a two-place formula $\chi_{\bar{b}}$ with a parameter. Suppose*

$$S_2^1 \vdash \forall a, \bar{b}\, [\mathrm{PHP}^a_{a^2}(\chi_{\bar{b}}) \to \exists y\, \theta(a, \bar{b}, y)]$$

where PHP *is of the form: $\chi_{\bar{b}}$ is not the graph of a surjective function $a \twoheadrightarrow a^2$. Then $\forall a, \bar{b}\, \exists y\, \theta(a, \bar{b}, y)$ is provable in $S_2^1 + \forall a\, \mathrm{PHP}^a_{a^2}(\mathrm{PV})$.*

Proof Re-writing our assumption slightly, we have

$$S_2^1 \vdash \forall a, \bar{b}\, [\neg\mathrm{PHP}^a_{a^2}(\chi_{\bar{b}}) \vee \exists y\, \theta(a, \bar{b}, y)].$$

Now $\neg\mathrm{PHP}^a_{a^2}(\chi_{\bar{b}})$ is the conjunction

$$\forall v < a^2\, \exists u < a\, \chi_{\bar{b}}(u, v) \wedge \forall u < a\, \exists v < a^2\, \chi_{\bar{b}}(u, v)$$
$$\wedge\, \forall v_1 < v_2 < a^2\, \forall u < a\, \neg(\chi_{\bar{b}}(u, v_1) \wedge \chi_{\bar{b}}(u, v_2))$$

so in particular, using only the middle conjunct,

$$S_2^1 \vdash \forall a, \bar{b}\, [\forall u < a\, \exists v < a^2\, \chi_{\bar{b}}(u, v) \vee \exists y\, \theta(a, \bar{b}, y)]$$

and by the witnessing theorem there is a PV function f such that

$$S_2^1 \vdash \forall a, \bar{b}\, [\forall u < a\, (f(a, b, u) < a^2 \wedge \chi_{\bar{b}}(u, f(a, \bar{b}, u))) \vee \exists y\, \theta(a, \bar{b}, y)].$$

Suppose the conclusion of the lemma fails, and there is a model M with

$$M \vDash S_2^1 + \forall x\, \mathrm{PHP}^x_{x^2}(\mathrm{PV}) + \forall y\, \neg\theta(a, \bar{b}, y)$$

for some $a, \bar{b} \in M$. Since $M \nvDash \exists y\, \theta(a, \bar{b}, y)$, three things must hold, corresponding to the three conjuncts in WPHP:

1. $M \vDash \forall v < a^2\, \exists u < a\, \chi_{\bar{b}}(u, v)$;
2. $M \vDash \forall u < a\, (f(a, \bar{b}, u) < a^2 \wedge \chi_{\bar{b}}(u, f(a, \bar{b}, u)))$;
3. $M \vDash \forall v_1 < v_2 < a^2\, \forall u < a\, \neg(\chi_{\bar{b}}(u, v_1) \wedge \chi_{\bar{b}}(u, v_2))$.

By $\mathrm{PHP}^{a}_{a^2}(f)$ in M, there exists $v_1 \in M$, $v_1 < a^2$ with $\forall u{<}a\, f(a,\bar{b},u) \neq v_1$. By (1) for some $u < a$ we have $\chi_{\bar{b}}(u, v_1)$. Now let $v_2 = f(a,\bar{b},u)$. By (2) $v_2 < a^2$ and $\chi_{\bar{b}}(u, v_2)$, and of course $v_1 \neq v_2$. But this contradicts (3). □

Corollary 4.12

1. *For any $\Sigma^b_1(f)$ formula $\chi_a(x,y)$ containing only a,x,y as free variables,*

$$S^1_2(f) \nvdash \forall a\,,\ \mathrm{PHP}^{a}_{a^2}(\chi_a(x,y)) \to \mathrm{PHP}^{a^2}_{a}(f).$$

2. *For any $\Sigma^b_1(<^*)$ formula $\chi_a(x,y)$ containing only a,x,y as free variables,*

$$S^1_2(<^*) \nvdash \forall a\,,\ \mathrm{PHP}^{a}_{a^2}(\chi_a(x,y)) \to$$
$$(\textit{if } <^* \textit{ is a total ordering on } [0,a) \textit{ then it has a least element}).$$

Proof Apply a relativized version of lemma 4.11. □

References

[1] D. Boneh and R. Lipton. Algorithms for black box fields and their application to cryptography. In *Proceedings Crypto '96*, volume 1109 of *LNCS*, pages 283–297. Springer-Verlag, 1996.

[2] D. Bovet, P. Crescenzi, and R. Silvestri. A uniform approach to define complexity classes. *Theoretical Computer Science*, 104:263–283, 1992.

[3] S. Buss. *Bounded Arithmetic.* Bibliopolis, 1986.

[4] H. Friedman. Countable models of set theories. In H. Rogers and A. Mathias, editors, *Cambridge summer school in mathematical logic*, volume 337 of *Lecture notes in mathematics*, pages 593–573. Springer-Verlag, 1973.

[5] E. Jeřábek. Dual weak pigeonhole principle, boolean complexity and derandomization. *Annals of Pure and Applied Logic*, 129:1–37, 2004.

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

[7] A. Maciel, T. Pitassi, and A. Woods. A new proof of the weak pigeonhole principle. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 368–377, 2000.

[8] J. Paris, A. Wilkie, and A. Woods. Provability of the pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic*, 53(4):1235–1244, 1988.

[9] S. Riis. Making infinite structures finite in models of second order bounded arithmetic. In P. Clote and J. Krajíček, editors, *Arithmetic, Proof Theory, and Computational Complexity*, pages 289–319. Oxford University Press, 1993.

[10] V. Shoup. Lower bounds for discrete logarithms and related problems. *Proceedings of Eurocrypt '97*, pages 246–266, 1997.

[11] N. Thapen. *The Weak Pigeonhole Principle in Models of Bounded Arithmetic.* DPhil Thesis, University of Oxford. Available from the ECCC thesis archive at `eccc.uni-trier.de/eccc-local/ECCC-Theses`.

[12] N. Thapen. A model-theoretic characterization of the weak pigeonhole principle. *Annals of Pure and Applied Logic*, 118:175–195, 2002.