A model-theoretic characterization of the weak pigeonhole principle

Neil Thapen*
Mathematical Institute, University of Oxford
thapen@maths.ox.ac.uk

September 26, 2002

Abstract

We bring together some facts about the weak pigeonhole principle (WPHP) from bounded arithmetic, complexity theory, cryptography and abstract model theory. We characterize the models of arithmetic in which WPHP fails as those which are determined by an initial segment and prove a conditional separation result in bounded arithmetic, that PV + (sharply bounded collection for PV formulas) lies strictly between PV and S^1_2 in strength, assuming that the cryptosystem RSA is secure.

Keywords: weak pigeonhole principle, bounded arithmetic
Classification: primary 03H15, secondary 68Q15, 03F30

The weak pigeonhole principle is a collective name for various statements of the form: if you put a large number of letters into a small number of pigeonholes, at least one pigeonhole will receive more than one letter. The version we will use most often will be the dual of this, and will have the form: there is no definable surjection from n onto n^2, for all n. WPHP was first studied by Paris, Wilkie and Woods in [17] as part of a programme to develop arithmetic without exponentiation. In particular, it was shown that $\mathrm{I}\Delta_0$ + WPHP is sufficient to prove that the set of prime numbers is unbounded,

*Research supported by EPSRC grant 98001658

and that WPHP itself is provable in $\mathrm{I}\Delta_0 + \forall x\,(x^{|x|}$ exists). The main open problem in the area is whether WPHP is provable in $\mathrm{I}\Delta_0$ (without the extra axiom). Most of the recent work on WPHP has been in propositional proof complexity, rather than bounded arithmetic, but the two are connected; the proof in [17] can be translated into a short propositional proof of WPHP. Finding a proof in $\mathrm{I}\Delta_0$ would correspond to finding a shorter propositional proof.

We characterize the weak pigeonhole principle in terms of *relative categoricity* or *categoricity over a predicate*.

Definition 0.1 (Gaifman; see [18], [19], [6]) *A structure M is relatively categorical over a definable subset R with respect to a theory T if $M \vDash T$ and for every $N \vDash T$, if there is an isomorphism between $M \restriction R$ and $N \restriction R$ then this can be extended to an isomorphism between M and N.*

Notice that this definition only really makes sense in a relational language. Our main result is that failure of WPHP in a model of arithmetic is equivalent to relative categoricity over an initial segment. Of course what this means depends upon the models, theories and versions of WPHP considered. The neatest results are in models of $\mathrm{PA}^{\mathrm{top}}$, a theory very like $\mathrm{I}\Delta_0$, but in which we do not need to worry about whether formulas are bounded.

The next section contains reminders of the basic definitions in bounded arithmetic; more detailed treatments can be found in [2] and [9]. Section 2 contains a useful method of amplifying surjections from a onto a^2 to surjections from a onto anything, and similarly for injections (lemma 2.1). In section 3 we define a theory S^i_0 which is, roughly speaking, the theory of initial segments with a greatest element of models of S^i_2, and show that failure of WPHP for a Σ^b_i function in a model of S^i_0 implies relative categoricity over an initial segment with respect to S^i_0 (theorem 3.7), and that in a model M of S^i_2 in which WPHP fails at a for a Σ^b_1 function, every subset of $M \restriction a\#a$ that is Σ^b_i definable in M is Σ^b_i definable inside $M \restriction a\#a$ (corollary 3.9). Section 4 looks at the difference in strength between the surjective and injective versions of WPHP for PV function symbols in S^1_2; we can witness proofs using surjective WPHP with a probabilistic polynomial time machine (lemma 4.2), but if we could efficiently witness injective WPHP, then we could crack RSA (lemma 4.5). This is used in section 5, together with a slightly different application of the idea of relative categoricity from that in section 3, to obtain our separation result (corollary 5.4).

The final two sections are largely independent of the preceding ones. They relate some old model theoretic theorems about relative categoricity, surjective functions and cardinality to the case of bounded arithmetic. These give some converses to our earlier results, in particular that relative categoricity of a model of S^1_0 implies failure of WPHP for some definable function (corollary 6.7) and that if a model of $\mathrm{PA}^{\mathrm{top}}$ satisfies WPHP, not even the cardinality of the model is determined by its initial segments (theorem 7.1). We also construct an uncountable model of S_2 in which the polynomial size sets are precisely the countable bounded definable sets (corollary 7.7). Most of the model theory used here is in chapter 12 of [6].

Some conventions about notation: throughout we will use δ and ε to stand for small nonstandard integers. We write sequences of numbers as $x_1, \ldots, x_n$ rather than $x_0, \ldots, x_{n-1}$. If we are working in a structure M, A is a subset of M and $r(x_1, \ldots, x_n)$ is a formula, then $r(A)$ (or just r, if A is clear from the context) is the set $\{\bar{x} \in A^n : M \vDash r(\bar{x})\}$.

1 Definitions

We define a language for arithmetic without function symbols. We use this for the reasons given above, to make it easy to work with substructures and with structures with a top element. Otherwise, the choice of language and the definition of the S^i_2 hierarchy follow Buss [2].

Definition 1.1 *L_A is the language consisting of constant symbols $0, 1$ and relation symbols $x < y$, $x = y + z$, $x = y \cdot z$ and $|x| = y$.*

$BASIC'$ is a set of axioms expressing that $<$ is a discrete linear order with least element 0, that $+$ and $\cdot$ define partial functions and $|\ |$ a total function, that the correct inductive properties hold (whenever the relevant operations are defined), that associativity, commutativity, distributivity and the relations between the operations and the ordering hold as expected, and that every number is either even or odd.

A model of $BASIC'$ is said to be of the form $[0, e+1)$ if it has a greatest element e. An element b of a model of $BASIC'$ is identified with the set of strictly smaller elements.

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. There is one more symbol which we will sometimes use, the smash function $x \# y$. We would like this to have the property that $2^{i \cdot j} = 2^i \# 2^j$, so we define it as $x \# y = 2^{|x-1| \cdot |y-1|}$. A useful piece of notation is $\# a$ for the cut $\bigcup\{a, a\#a, a\#(a\#a), \ldots\} = 2^{|a|^{\mathbb{N}}}$ (when it exists).

Unfortunately the standard definition of Σ_i^b formulas only makes sense if $+, \cdot, |\ |$ and $\#$ are function symbols. However in the relevant theories they will always define functions, so we will avoid the issue and pretend that they are function symbols when necessary.

Definition 1.2

1. *A formula is Δ_0 if all of its quantifiers appear bounded, that is, in the form $\forall x \leqslant y$ or $\exists x \leqslant y$. In a model with a top element we will consider every formula to be Δ_0.*

2. *(Assuming $+, \cdot, |\ |, \#$ are function symbols) A formula is sharply bounded, or Σ_0^b, if all of its quantifiers are bounded by terms of the form $|t|$. A formula is Σ_i^b if it contains sharply bounded quantifiers, which we do not count, and $i-1$ quantifier alternations of quantifiers bounded by terms in $+$, $\cdot$ and $\#$, beginning with an existential quantifier. The Π_i^b formulas are defined similarly.*

3. $\mathrm{I}\Delta_0 = BASIC' + (+$ *and* $\cdot$ *are total functions*$) + ($*induction for all* Δ_0 *formulas*$)$.

4. $S_2 = \mathrm{I}\Delta_0 + (\#$ *is a total function*$)$.

5. $S_2^i = BASIC' + (+, \cdot$ *and* $\#$ *are total functions* $) + \Sigma_i^b - LIND$, *where* $\Sigma_i^b - LIND$ *is the length induction axiom*

$$\forall \bar{c} \forall a\, , \phi(\bar{c}, 0) \wedge \forall x < |a|\, (\phi(\bar{c}, x) \rightarrow \phi(\bar{c}, x+1)) \rightarrow \phi(\bar{c}, a)$$

   for every Σ_i^b *formula* ϕ.

6. $\mathrm{PA}^{\mathrm{top}} = BASIC' + ($*there is a greatest element*$) + ($*the least number principle for all formulas*$)$.

7. *To define the theory $S_2^i(\alpha)$, we add a new predicate symbol α to the language and define the $\Sigma_i^b(\alpha)$ formulas in the same way as we did the Σ_i^b formulas, except that we now allow α to occur. Then*

$$S_2^i(\alpha) = BASIC' + \Sigma_i^b(\alpha) - LIND.$$

   $\mathrm{PA}^{\mathrm{top}}(\alpha)$, $\mathrm{I}\Delta_0(\alpha)$ *and* $S_2(\alpha)$ *are defined similarly.*

Given a model of $K \vDash \mathrm{PA}^{\mathrm{top}}$ of the form $[0, a)$, we can construct a (unique) end-extension of K to a model of $\mathrm{PA}^{\mathrm{top}}$ of the form $[0, a^2)$ by defining natural $+$ and $\cdot$ relations on $K \times K$. Repeating this construction countably many times, we get

Lemma 1.3 *Any model of* $\mathrm{PA}^{\mathrm{top}}$ *has an end-extension to a model of* $\mathrm{I}\Delta_0$*. Hence* $\mathrm{PA}^{\mathrm{top}}$ *and* $\mathrm{I}\Delta_0$ *prove the same* Π_1 *sentences.*$\square$

In particular, $\mathrm{I}\Delta_0$ proves $\mathrm{WPHP}(\Delta_0)$ if and only if $\mathrm{PA}^{\mathrm{top}}$ does. This lemma is proved in [5], where it is attributed indirectly to Paris.

We will also use a theory PV (strictly speaking the theory PV_1 defined in [2] or [9]), which is an axiomatization of the polynomial time functions. The polynomial time functions can be defined as the closure under composition, renaming of variables and "limited recursion on notation" of certain basic functions. Define PV by taking a function symbol for every such function and a set of universal axioms saying that each function is built up from the basic functions in the way described above. PV can be considered as a subtheory of S^1_2, if we add all these new symbols to the language of S^1_2.

Most of the interest in the hierarchy S^i_2 comes from the following theorem.

Theorem 1.4 (Buss [2]) *If* $i \geqslant 1$*,* ϕ *is a* Σ^b_i *formula and* $S^i_2 \vdash \forall x\, \exists y\, \phi(x, y)$*, then there is a function* f *at level* i *in the polynomial hierarchy such that* $\forall x\, \phi(x, f(x))$ *is provable in* S^i_2 *(in fact in a weaker theory). In particular, the* Σ^b_1 *functions which are provably total in* S^1_2 *are precisely the polynomial time functions.* $\square$

Definition 1.5 *In a model* M *of* $BASIC'$*, for* $a, b \in M$*,* $a < b$*, and* f *a definable function, the injective* PHP*, written* $PHP^b_a(f)$*, says that* f *is not an injection* $b \hookrightarrow a$*. The surjective* PHP*, written* $\mathrm{PHP}^a_b(f)$*, says that* f *is not a surjection* $a \twoheadrightarrow b$*. If* Γ *is a class of functions,* $\mathrm{PHP}(\Gamma)$ *is the set* $\{\mathrm{PHP}(f) : f \in \Gamma\}$.

By the weak pigeonhole principle, WPHP, we mean one of the above in the case $b \geqslant a^2$. The particular meaning should be clear from the context. Possible Γs considered will be: polynomial time functions in models of PV or S^1_2, Σ^b_i functions in models of S^i_2 and Δ_0 functions in models of $\mathrm{I}\Delta_0$ or $\mathrm{PA}^{\mathrm{top}}$.

2 Amplification

Lemma 2.1 *Suppose $M \vDash S_2^i$, $a, b, c \in M$, $a < b$ and f is a Σ_i^b definable injection $a^2 \hookrightarrow a$ with a parameter c. Then there is a Σ_i^b definable injection $F : b \hookrightarrow a$ with parameters a, b, c. Similarly a Σ_i^b surjection $g : a \twoheadrightarrow a^2$ can be amplified to $G : a \twoheadrightarrow b$.*

Proof. Consider f as an injection $a \times a \hookrightarrow a$. The function F is given by the following algorithm. Let d be least such that $d > b$ and $d = a^\delta$ for some δ (where δ is an integer in the structure). Thus d is Δ_1^b definable from a and b. Consider the binary tree consisting of $2\delta + 1$ nodes labelled $w_1, \ldots, w_{\delta+1}, x_1, \ldots, x_\delta$. w_1 is the root (at the bottom), and for each $i \leqslant \delta$, w_i has a left-hand child x_i and a right-hand child w_{i+1}. No other nodes have any children. For $x < d$, consider x as a sequence of δ numerals $x_i < a$, so that $x = \sum_{i=1}^{\delta} x_i a^{i-1}$. Give each node x_i its corresponding value and set $w_{\delta+1} = 0$. If $f : a^2 \hookrightarrow a$, for each $i \leqslant \delta$ (working down from the top) put $w_i = f(x_i, w_{i+1})$. F is the function taking x to w_1.

Formally, $F(x) = y$ if and only if

$$\exists w < d\,,\, w_1 = y \wedge w_{\delta+1} = 0 \wedge \forall 1 \leqslant i \leqslant \delta\, (f(x_i, w_{i+1}) = w_i).$$

Induction up the tree shows that F is an injection $d \hookrightarrow a$.

The surjective case is similar: if $g : a \twoheadrightarrow a \times a$ has projections g_1,g_2 then $x = G(y)$ is computed by setting $w_1 = y$ and for each $i \leqslant \delta$ putting $x_i = g_1(w_i)$ and $w_{i+1} = g_2(w_i)$, this time working up from the bottom. Induction shows that G is a surjection $a \twoheadrightarrow d$. $\square$

Notice that the only part of the model used in this proof is the initial segment up to $\max(d, c)$; in particular, all quantification can be done over this initial segment.

Corollary 2.2 *Suppose $M \vDash$ PV, $a, b, c \in M$, $a < b$ and f is a PV function symbol such that $f(c, x)$ is an injection $a^2 \hookrightarrow a$ (x here is a placeholder). Then there is a PV function symbol F such that $F(c, b, a, x)$ is an injection $b \hookrightarrow a$. If in addition $M \vDash S_2^1$, then a PV surjection $g : a \twoheadrightarrow a^2$ can similarly be amplified to a surjection $G : a \twoheadrightarrow b$.*

Proof. In lemma 2.1, if the function f is polynomial time then so is the function F. The same holds for g and G. The sentence expressing that F is an injection whenever f is, namely

$$\forall c \forall b \forall a < b \forall x_1 < x_2 < b (F(x_1) \neq F(x_2) \vee \exists y_1 < y_2 < a^2 (f(y_1) = f(y_2))),$$

is $\forall\Sigma^b_1$, and is provable in S^1_2 by the lemma, so will be provable in PV by the conservativity of PV over S^1_2 [9]. The sentence expressing that G is a surjection whenever g is has higher quantifier complexity, so we cannot use this conservativity result here and need the extra assumption that $M \vDash S^1_2$. $\Box$

Let $u(e, x, c)$ be the universal PV function symbol, which calculates the output of the Turing machine with code e run on input x for time $|c|$.

Lemma 2.3 *The theories surjective* WPHP(PV) *with parameters,*

$$\forall a\, \forall e\, \forall c\, \mathrm{PHP}^a_{a^2}(u(e, x, c))$$

and surjective WPHP(PV) *without parameters,*

$$\{\forall a\, \mathrm{PHP}^a_{a^2}(f(x)) : f \text{ a PV function symbol}\}$$

are equivalent over S^1_2.

Proof. One direction is trivial; for the other, suppose $M \vDash S^1_2$, $a, e, c \in M$ and $u(e, x, c)$ is a surjection $a \twoheadrightarrow a^2$. Find $d > e, c, a$. By corollary 2.2, there is a PV function symbol G such that $G(d-1, e, c, a, x)$ is a surjection $a \twoheadrightarrow d^{10}$ (G has to calculate from $d - 1$ the parameter d^{10} needed for the lemma). Since $d - 1, e, c, a < d$, $G(x_1, \ldots, x_5)$ is a parameter-free surjection from d^5 onto d^{10}. $\Box$

3 Building end-extensions

Theorem 3.1 (Paris, Wilkie, Woods [17]) *If $K \vDash \mathrm{PA}^{\mathrm{top}}$ is of the form $[0, a^\varepsilon)$ for $a, \varepsilon \in K$, and K defines a function f which is a surjection $a \twoheadrightarrow a^\varepsilon$, then K has an end-extension to $J \vDash \mathrm{PA}^{\mathrm{top}}$ of the form $[0, a^{2^\varepsilon})$.* $\Box$

In this section we define a hierarchy of theories S^i_0 that bear a similar relationship to the S^i_2 hierarchy as $\mathrm{PA}^{\mathrm{top}}$ does to $\mathrm{I}\Delta_0$, and show that a similar result to theorem 3.1 holds in S^i_0, although with a smaller increase in size. We observe that the construction used in the proof of this theorem shows that any such end-extension of K is definable inside K, and so is determined by K.

The proof of theorem 3.1 is essentially an amplification of f to a surjection $a \twoheadrightarrow a^{2^\varepsilon}$, and is similar to the amplification of pseudorandom number generators to pseudorandom function generators using a complete binary tree (see, for example [4]). The amplification result above, and the construction below, are based on a similar idea. However they use less induction than the construction in [17] and are only able to use one "branch" of a binary tree. This is why the increase in size below is from a^ε to a^{ε^2}, rather than to a^{2^ε}. Notice that our amplification construction above is similar to the one used to polynomially increase the stretching factor of a pseudorandom number generator.

Definition 3.2 *S_0^i is the theory $BASIC'+$ (there is a greatest element b) $+(|b|^k$ exists for all $k \in \mathbb{N})$ together with the length induction axiom*

$$\forall \bar{c}\,,\ \phi(\bar{c},0) \wedge \forall x < |b|^k\, (\phi(\bar{c},x) \to \phi(\bar{c},x+1)) \to \phi(\bar{c},|b|^k)$$

for every Σ_i^b formula ϕ and every $k \in \mathbb{N}$.

Here we abuse notation slightly and alter the definition of a Σ_i^b formula in the natural way.

Definition 3.3 *A formula is Σ_i^b if it contains $i-1$ quantifier alternations beginning with an existential quantifier (we do not need to insist on bounded quantifiers, since we have a greatest element) and ignoring sharply bounded quantifiers, which will now also include those of the form $\forall x < |y|^k$ or $\exists x < |y|^k$, for any $k \in \mathbb{N}$.*

If S and $X \subseteq S$ are sets in a structure, X is said to be Δ_i^b in S if both X and $S \setminus X$ are definable by Σ_i^b formulas.

So a typical model of S_0^i is an initial segment $M \restriction a$ of a model M of S_2^i in which a is not a length.

Lemma 3.4 *In S_0^1 we can define a function $bit(x,i)$ for the i-th bit of x. Using this, for any $K \vDash S_0^1$, if $a \in K$ is a power of 2 then we can define a Σ_1^b coding function x_i for the ith numeral in the a-ary expansion of x. The most significant bits or numerals will come last.*

Proof. Using the $|\ |$ relation we can define a limited amount of exponentiation and define the function $bit(x,i)$ in the normal way in terms of powers of 2. Its properties can be proved as in [2], using induction and our axiom that every number is either even or odd. □

We define a principle $\mathrm{mPHP}_a^{a^2}$, "multifunction WPHP", which states that there is no (single-valued) function with domain a subset of a and range all of a^2. So if a function $f : a \to a^2$ violates surjective WPHP, then it also violates mPHP; if a function $f : a^2 \to a$ violates injective WPHP, then the inverse of f is a surjection from the range of f onto a^2 and thus violates mPHP. The reason for the name is that mPHP is equivalent to the statement: there is no injective many-valued function from a^2 into a. This last form is the version of WPHP of which a T_2^2 proof is given in [13].

Definition 3.5 *We say that* $\mathrm{mPHP}_a^{a^2}(\Gamma)$ *holds if there is no pair of formulas* $r(x)$ *and* $f(x,y)$ *in the class* Γ *with parameters in* $[0,a)$ *such that* f *is a surjection from* $r([0,a))$ *onto* $[0,a^2)$*. That is, it holds if*

$$\neg[\forall y < a^2\, \exists x < a\, (r(x) \wedge f(x,y)) \wedge \forall x < a\, (r(x) \to \exists! y < a^2\, f(x,y))]$$

holds for all such r, f.

Lemma 3.6 *Suppose* $K \vDash S_0^j$ *is of the form* $[0, a^\varepsilon)$*, where* $a = 2^\alpha$ *some* α *and* K *does not satisfy* $\mathrm{mPHP}_a^{a^2}(\Sigma_j^b)$*. Then for any* $l \in \mathbb{N}$ *there is a* Σ_j^b *subset* S *of* $[0,a)$ *such that we can define a* Σ_j^b *coding relation* $\langle x \rangle_i = y$*, which defines a function from* $S \times \varepsilon^l$ *to* $[0,a)$*, but is not defined for* x *outside* S*. This can be used to code any* Σ_j^b *definable* ε^l*-length sequence of elements of* $[0,a)$ *as an element of* S *(possibly two elements of* S *will code the same sequence).*

Proof. We will call elements of $[0,a)$ "numerals", and use the function x_i to treat any element of K as a sequence of ε numerals. Let $r(x)$ and $f(x,y)$ be the Σ_j^b formulas violating mPHP. We first amplify f to a function g with range a^ε rather than a^2. We use the same construction as lemma 2.1. Define

$$\begin{aligned} g(x,y) \quad \Leftrightarrow \quad & \exists w\, , w_1 = x \wedge \forall 1 \leqslant i < \varepsilon (r(w_i) \wedge f(w_i, (y_i, w_{i+1}))) \\ & \qquad \qquad \wedge r(w_\varepsilon) \wedge f(w_\varepsilon, (y_\varepsilon, 0)) \end{aligned}$$

and

$$s(x) \Leftrightarrow x < a \wedge \exists y < a^\varepsilon\, g(x,y).$$

Then g and s are Σ_j^b and by induction g is a surjection from $s([0,a))$ onto $[0,a^\varepsilon)$. Furthermore, if $r([0,a)) = [0,a)$ then $s([0,a)) = [0,a)$ and if f is 1-1 then g is.

Thus we can encode an ε-length sequence of numerals as a single element of $s([0,a))$. To encode ε^l-length sequences, we use a complete ε-ary tree of (standard) height l. We label the leaves of the tree with the sequence $\beta_1 \ldots \beta_{\varepsilon^l}$ which we want to encode, and then label the other nodes so that if a node is labelled w, then $w \in s$ and its children are labelled $g(w)_1 \ldots g(w)_\varepsilon$. We define $\langle x \rangle_i = y$ to hold if, in the tree with x at the root, the leaf at the end of the path naturally given by $i < \varepsilon^l$ (considered as an l-tuple in $\varepsilon \times \ldots \times \varepsilon$) is labelled y. Let $S(x)$ be the formula $x < a \wedge \forall 1 \leqslant i \leqslant \varepsilon^l \, \exists y < a \, \langle x \rangle_i = y$.

To show formally that this is a coding relation with the required property, let $\phi(i,y)$ be a Σ^b_j formula, possibly with parameters, such that $\forall 1 \leqslant i \leqslant \varepsilon^l \, \exists y < a \, \phi(i,y)$. We claim that we can find $x \in S$ encoding a sequence satisfying ϕ, ie. $\forall 1 \leqslant i \leqslant \varepsilon^l \, \phi(i, \langle x \rangle_i)$. This is done by considering, in turn, each level of the tree described in the previous paragraph and showing that each node on that level has a suitable label.

First look at the level immediately below the leaves. Let $\phi'(i,y)$ be the formula stating that y is a suitable label for the ith node on this level:

$$s(y) \wedge \forall 1 \leqslant k \leqslant \varepsilon \, \phi((i-1) \cdot \varepsilon + k, g(y)_k)$$

(where g is our surjection $s([0,a)) \twoheadrightarrow [0,a^\varepsilon))$. Since we can encode, using first comprehension and then g, any Σ^b_j definable ε-length sequence of numerals as a single element of s, we can show that all these nodes can be labelled, ie. $\forall 1 \leqslant i \leqslant \varepsilon^{l-1} \, \exists y < a \, \phi'(i,y)$. The formula $\phi'(i,y)$ is still Σ^b_j, so we can repeat this step $l-1$ times for all the lower levels in the tree to find $y \in S$ (there may be more than one such y) encoding a suitable ε^l-length sequence via $\langle y \rangle_i$. $\square$

Note that in the theorem below the restriction that a should be a power of 2 can easily be dispensed with, since given $K \vDash S^i_0$ of the form $[0,b)$ we can always construct a model of the form $[0,b^2)$ from the cartesian product $K \times K$, and this model will certainly be determined up to isomorphism over K. So we can find a suitable power of 2 when we need one.

Theorem 3.7 *Suppose $j,k,l \in \mathbb{N}$, $j \geqslant 1$, $k \geqslant 0$, $l \geqslant 2$. Let K be a model of S^{j+k}_0 of the form $[0,a^\varepsilon)$, where $a = 2^\alpha$ for some α. Suppose $K \vDash \neg\mathrm{mPHP}^{a^2}_a(\Sigma^b_j)$. Then*

1. *K has an end extension to a model J of S^{k+1}_0 of the form $[0,a^{\varepsilon^l})$. Furthermore this end extension is definable inside K, in the sense that*

there is a Σ^b_j subset S of $[0,a)$ on which we can define relations $=_J$, $<_J$, $+_J$, $\cdot_J$ and $|\ |_J$ which are Δ^b_j in S such that $=_J$ is an equivalence relation on S and J is the structure $S/_{=_J}$ with the relations induced by $<_J$, $+_J$, $\cdot_J$ and $|\ |_J$;

2. *If I is any end-extension of K to a model of S^j_0 of the form $[0,a^{\varepsilon^l})$, then I is relatively categorical over K with respect to the theory $S^j_0+(a^{\varepsilon^l}-1$ is the greatest element).*

(The most useful case is $j=1$, $k=0$.)

Proof. We will first construct J, and then show it is an end extension. Each element of J will be constructed in the natural way as a sequence of numerals of length ε^l. We use the coding function $\langle x\rangle_i$ and the set S given by lemma 3.6, and the fact that we can define addition and multiplication numeral-wise.

Define, for b,c in S,

$$\begin{aligned}
b =_J c \quad &\Leftrightarrow \quad \forall 1\leqslant i\leqslant\varepsilon^l\ \langle b\rangle_i = \langle c\rangle_i;\\
b <_J c \quad &\Leftrightarrow \quad \exists 1\leqslant i\leqslant\varepsilon^l\ (\langle b\rangle_i < \langle c\rangle_i \wedge \forall i<t\leqslant\varepsilon^l \langle b\rangle_t = \langle c\rangle_t);\\
(|b|=c)_J \quad &\Leftrightarrow \quad (b=0\wedge c=0)\vee \exists 1\leqslant i\leqslant\varepsilon^l\ ,\\
&\qquad \langle b\rangle_i \neq 0 \wedge c = \alpha\cdot(i-1)+|\langle b\rangle_i| \wedge \forall i<t\leqslant\varepsilon^l \langle b\rangle_t = 0.
\end{aligned}$$

These are Δ^b_j in S because we only apply $\langle x\rangle_i$ to members of S, on which it behaves like a function, so that we can negate the subformulas containing it without having to increase the quantifier complexity. For example, to write the definition of equality in J more fully, we have

$$\begin{aligned}
b =_J c \quad &\Leftrightarrow \quad S(b)\wedge S(c)\wedge \forall 1\leqslant i\leqslant\varepsilon^l\ \exists y\ \langle b\rangle_i = y \wedge \langle c\rangle_i = y\\
b \neq_J c \quad &\Leftrightarrow \quad S(b)\wedge S(c)\wedge \exists 1\leqslant i\leqslant\varepsilon^l\ \exists y\,\exists z\ \langle b\rangle_i = y \wedge \langle c\rangle_i = z \wedge y\neq z.
\end{aligned}$$

For addition, first note that in S^1_0 we can define Σ^b_1 *modulo addition* and *carry* functions $A(x,y,z)$ and $C(x,y,z)$ in K such that if $x,y,z<a$, then $A(x,y,z), C(x,y,z) < a$ and $x+y+z = a\cdot C(x,y,z)+A(x,y,z)$. We add our ε^l-length sequences numeral by numeral, and use a variable w to encode the sequence of numerals carried.

Define, for $c,d,e\in S$,

$$(c+d=e)_J \Leftrightarrow \exists w\ ,\ S(w)\wedge[\langle w\rangle_1 = C(\langle c\rangle_1,\langle d\rangle_1,0)$$

$$\wedge \forall 1 < i \leqslant \varepsilon^l \langle w \rangle_i = C(\langle c \rangle_i, \langle d \rangle_i, \langle w \rangle_{i-1})]$$
$$\wedge [\langle e \rangle_1 = A(\langle c \rangle_1, \langle d \rangle_1, 0) \wedge \langle w \rangle_\varepsilon = 0$$
$$\wedge \forall 1 < i \leqslant \varepsilon^l \langle e \rangle_i = A(\langle c \rangle_i, \langle d \rangle_i, \langle w \rangle_{i-1})].$$

We can always find a $w \in S$ witnessing the first expression in square brackets, so we can define the complement in S of this relation by putting a negation in front of the second expression in square brackets. Hence it is Δ_j^b in S.

Multiplication is defined in a similar way. We need to be able to encode $(2\varepsilon^l + 1) \times (\varepsilon^l)^2$ matrices of numerals, and by lemma 3.6 there is a Σ_j^b subset T of $[0, a)$ on which we can do this via a Σ_j^b coding relation $(x)_k^i = y$, for $i = 1, \ldots, (\varepsilon^l)^2$ and $k = 1, \ldots, 2\varepsilon^l + 1$. For any $b, c \in S$, there are two such matrices π, ζ which encode the multiplication of b by c as follows (of course there may be more than one member of T coding each of these matrices). Each row of π contains the product of a numeral of b and a numeral of c, suitably offset. In particular if $\langle b \rangle_k \langle c \rangle_i = u + av$, for some $u, v < a$, then row $(i-1)\varepsilon^l + k$ of π has u in the $(i+k)$th place, v in the $(i+k+1)$st place, and zero everywhere else. Each row i of ζ encodes the sum of rows 1 to i of π. Define, for $d \in S$, $(b \cdot c = d)_J$ if for some π and ζ as above $\forall 1 \leqslant k \leqslant \varepsilon^l \ \langle c \rangle_k = (\zeta)_k^{(\varepsilon^l)^2}$ and all the other entries in row $(\varepsilon^l)^2$ of ζ are 0.

This completes the construction of the model $J = S/_{=_J}$. For $x \in S$ we will write $[x]$ for the equivalence class of x under $=_J$. To establish the $BASIC'$ axioms, we first observe that by $\Sigma_j^b - LIND$ in K, $<_J$ is a total ordering on J. Then consider the formula

$$\phi(x, y) \Leftrightarrow S(y) \wedge \forall 1 \leqslant i \leqslant \varepsilon \ \langle y \rangle_i = x_i \wedge \forall \varepsilon < i \leqslant \varepsilon^l \ \langle y \rangle_i = 0$$

and the map

$$\sigma : x \mapsto \{y \in S : K \vDash \phi(x, y)\}.$$

This is a map $K \to J$ and the relations on S have been defined so that it is an isomorphism onto an initial segment of J. Hence J is an end-extension of K. Standard methods now show that $J \vDash BASIC'$.

Let $\gamma = \alpha \cdot \varepsilon$ and $[\gamma'] = \sigma(\gamma)$, so that the sharply bounded quantifiers are precisely those that are bounded by some standard power of γ (in K) or some standard power of $[\gamma']$ (in J).

We claim that, for $n \geqslant 0$, for every Σ_{n+1}^b (or Π_{n+1}^b) formula $\theta(\bar{x})$, there is a Σ_{n+j}^b (respectively Π_{n+j}^b) formula $\theta_J(\bar{x})$ such that for all $\bar{b} \in S$,

$$J \vDash \theta([\bar{b}]) \Leftrightarrow K \vDash \theta_J(\bar{b}). \qquad \text{(i)}$$

and if θ is sharply bounded then we can find both a Σ_j^b formula θ_J^Σ and a Π_j^b formula θ_J^Π satisfying (i).

We prove the last case first, by induction on the number of quantifiers in θ. We know how to translate quantifier free formulas into formulas that are Δ_j^b in S, and this is precisely the property that we require. Now suppose θ is of the form $\forall y < [\gamma']^m \, \chi(\bar{x}, y)$ for some $m \in \mathbb{N}$, where χ is sharply bounded. Then for any $\bar{b} \in S$, by the definition of the isomorphism σ,

$$\begin{aligned} J \vDash \theta([\bar{b}]) &\Leftrightarrow \forall i \in K,\ i < \gamma^m,\ J \vDash \chi([\bar{b}], \sigma(i)) \\ &\Leftrightarrow \forall i \in K,\ i < \gamma^m,\ \forall c \in \sigma(i), K \vDash \chi_J^\Pi(\bar{b}, c) \\ &\Leftrightarrow K \vDash \forall i < \gamma^m \, \forall x \, (S(x) \wedge \phi(i, x) \rightarrow \chi_J^\Pi(\bar{b}, x)) \end{aligned}$$

where χ_J^Π is given by the induction hypothesis and is Π_j^b. Since the equivalence class $\sigma(i)$ is never empty, this is in turn equivalent to

$$K \vDash \forall i < \gamma^m \, \exists x \, (S(x) \wedge \phi(i, x) \wedge \chi_J^\Sigma(\bar{b}, x))$$

where χ_J^Σ is given by the induction hypothesis and is Σ_j^b. We deal similarly with sharply bounded existential quantifiers.

For the remaining cases, sharply bounded quantifiers are dealt with as above. If θ is of the form $\exists y \, \chi(\bar{x}, y)$, where χ is a Σ_{n+1}^b formula for which we have found a suitable Σ_{n+j}^b translation χ_J, we have

$$\begin{aligned} J \vDash \exists y \, \chi([\bar{b}], y) &\Leftrightarrow J \vDash \chi([\bar{b}], [c]) \quad \text{for some } c \in S \\ &\Leftrightarrow K \vDash \exists z \, (S(z) \wedge \chi_J(\bar{b}, z)). \end{aligned}$$

We treat Π_{n+1}^b formulas in a similar way.

To show that J satisfies $\Sigma_{k+1}^b - LIND$, suppose θ is a Σ_{k+1}^b formula and

$$J \vDash \theta(0) \wedge \forall x < [\gamma']^m \, (\theta(x) \rightarrow \theta(x+1)).$$

Then for the corresponding Σ_{j+k}^b formula θ_J, if we write $\theta_J(\phi(x))$ for $\exists y \, (S(y) \wedge \phi(x, y) \wedge \theta_J(y))$,

$$K \vDash \theta_J(\phi(0)) \wedge \forall x < \gamma^m \, (\theta_J(\phi(x)) \rightarrow \theta_J(\phi(x+1))).$$

So by $\Sigma_{j+k}^b - LIND$ in K, $K \vDash \theta_J(\phi(\gamma^m))$ and hence $J \vDash \theta([\gamma']^m)$.

Finally, suppose I is an end-extension of K to a model of S_0^j of the form $[0, a^{\varepsilon^l})$. Let J be the end-extension of K to a model of S_0^1 given by the construction above if we take $k = 0$. We claim that I is isomorphic to J.

We can use our ordinary coding function to consider any $x \in I$ as a sequence $x_1 \dots x_{\varepsilon^l}$ of numerals in $[0, a)$. For $x, y \in I$, $y \in S$, put

$$\phi'(x, y) \Leftrightarrow \forall 1 \leqslant i \leqslant \varepsilon^l \, \langle y \rangle_i = x_i,$$

meaning "y codes x". Just as in lemma 3.6, $\langle x \rangle_i$ can be used in I to encode any Σ_j^b definable ε^l-length sequence of numerals as an element of S, so in particular, every element of I is coded by (at least one) element of S. Conversely, by normal comprehension in I, every element of S codes an element of I. Also all the elements in an $=_J$ equivalence class in S must code the same element of I. Hence the map

$$\sigma' : x \mapsto \{y \in S : I \vDash \phi'(x, y)\}$$

is a bijection $\sigma' : I \leftrightarrow J$, and the definitions of the relations on S are set up precisely so that it is an isomorphism. $\square$

Corollary 3.8 *For $i \geqslant 1$, if $K \vDash S_0^i$ is of the form $[0, a\#a)$ and defines a Σ_1^b function violating either injective or surjective* WPHP *between a and a^2, then K has an end-extension to a model M of S_2^i in which $\#a$ is cofinal. Furthermore this end-extension is unique, in that for any model N of S_2^i with $N \restriction a\#a$ isomorphic to K, this isomorphism extends to an isomorphism $M \cong N \restriction \#a$.* $\square$

Corollary 3.9 *For $i \geqslant 1$, let M be a model of S_2^i in which either injective or surjective* WPHP *fails between a and a^2 for a Σ_1^b function with parameters in $[0, a\#a)$. Then every Σ_i^b subset of $[0, a\#a)$ definable in M with parameters from $[0, a\#a)$ (and where we allow $\#$ to appear as a function symbol in the bounds on quantifiers) is Σ_i^b definable inside $M \restriction a\#a$, that is, with all quantifiers bounded by $a\#a$.*

Proof. Use the translation in the proof of theorem 3.7.$\square$

Hence if the weak pigeonhole principle fails we can do computations from the polynomial hierarchy in constant space. However this is at the expense of introducing more sharply bounded universal quantifiers, and so increasing the time taken (in some sense).

The version of corollary 3.9 for $\mathrm{I}\Delta_0$ leads naturally to a proof by diagonalization that the (parameter-free) Δ_0 hierarchy does not collapse in a model in which WPHP fails [16]. See also [14] for a discussion of the extent to which the theory of an initial segment $M \restriction b$ of a model of true arithmetic is determined by $M \restriction a$, for $a < b$, under various assumptions about the collapse of the linear or polynomial time hierarchies.

4 Witnessing WPHP

Definition 4.1 *$S^1_2(\mathrm{PV})$ is the theory S^1_2 with all the* PV *function symbols and their defining axioms added to it, and with induction for all formulas containing these new symbols. It is conservative over S^1_2, and in this section we will identify it with S^1_2.* $\mathrm{PHP}^a_{a^2}(\mathrm{PV})$ *is the surjective* WPHP *for* PV *function symbols without parameters.*

Lemma 4.2 (Wilkie [9]) *If $S^1_2 + \forall a\, \mathrm{PHP}^a_{a^2}(\mathrm{PV}) \vdash \forall x\, \exists y\, \theta(x, y)$, with θ a Σ^b_1 formula, then there is a probabilistic polynomial time machine which will, with probability greater than $2/3$, compute on input x a value of y such that $\theta(x, y)$.* $\square$

A predicate A is in the class **ZPP** if there is a polynomial time probabilistic algorithm which gives the right answer to the question "is x in A?" with high probability and never gives a wrong answer.

Corollary 4.3 *If $S^1_2 + \forall a\, \mathrm{PHP}^a_{a^2}(\mathrm{PV})$ Δ^b_1-defines a predicate A, then $A \in$* **ZPP**. $\square$

However it is unlikely that surjective WPHP captures all of **ZPP**. In [1] an oracle is constructed with respect to which **ZPP** does not have a complete language, and it is not hard to show that if, for any sufficiently well-behaved complexity class $\mathcal{C}$ (such as **ZPP** with an oracle) there is a recursive theory T such that every predicate in $\mathcal{C}$ is Δ^b_1 definable in T and every proof in T that a predicate L is Δ^b_1 yields a machine witnessing that $L \in \mathcal{C}$ (as Buss' theorem does in the proof of lemma 4.2), then $\mathcal{C}$ has a complete language. So if there is a proof of a converse to corollary 4.3, it does not relativize.

Definition 4.4 (Rivest, Shamir, Adleman [21]) *An instance of RSA consists of a modulus n which is the product of two large primes p and q, exponents e and d which are mutually inverse modulo $(p-1)(q-1)$, a message $m < n$ and a cyphertext $c < n$ such that $c \equiv m^e \bmod n$ and $m \equiv c^d \bmod n$. We say that we can crack RSA if, given n, e and c, we can efficiently find m.*

Lemma 4.5 (Krajíček and Pudlák [11]) *Suppose there is an efficient algorithm witnessing injective* WPHP *for* PV *function symbols with parameters, that is, given any polynomial time function f there is an algorithm which, on input c, a such that $\forall x < a^2\, (f(c, x) < a)$, outputs $x_1 < x_2 < a^2$ such that $f(c, x_1) = f(c, x_2)$. Then we can crack RSA.*

Proof. (This is a more direct version of the proof in [11].) We are given c, n and e. Without loss of generality $(c, n) = 1$ or we could factorize n, recover p and q and hence find $(p-1)(q-1)$, d and m. Hence c has an inverse modulo n.

Let s be the order of $c \bmod n$, which will be the same as the order of $m \bmod n$. Now e and s must be coprime. Otherwise let $(e, s) = t > 1$ and $u = s/t$. Then $u < s$ and $s|eu$, so $c^u \equiv m^{eu} \equiv 1$, contradicting the leastness of s.

Use the algorithm on the function $x \mapsto c^x \bmod n$ to find $x_1 < x_2 < n^2$ with $c^{x_1} \equiv c^{x_2} \bmod n$. Let $r_0 = x_2 - x_1 \neq 0$. Then $c^{r_0} \equiv 1$ so $s|r_0$.

Remove all factors of e from r_0 by calculating

$$r_1 = \frac{r_0}{(e, r_0)}, \quad r_2 = \frac{r_1}{(e, r_1)}, \quad \ldots$$

to get r with $(e, r) = 1$. This takes at most $\log r_0$ divisions.

Since $(e, s) = 1$ and $s|r_0$ we must have $s|r_1$. Similarly $s|r_2$ etc. So $s|r$.

Calculate d' such that $d'e = 1 + tr$ some t; finally calculate $c^{d'} \bmod n$.

Then, since $s|r$ so $m^r \equiv 1$,

$$c^{d'} \equiv m^{ed'} \equiv m^{1+tr} \equiv (m^r)^t m \equiv m. \square$$

Corollary 4.6

1. *If S_2^1 proves injective* WPHP *for* PV *functions with parameters, then RSA is vulnerable to deterministic polynomial time attack.*

2. *If S_2^1 together with the surjective* WPHP *for* PV *functions proves the injective* WPHP *for* PV *functions with parameters, then RSA is vulnerable to probabilistic polynomial time attack.* $\square$

5 A conditional separation result

We give our first converse to the characterization of structures in which WPHP fails. Namely, if a model of PV satisfies WPHP, then its initial segments have more than one possible end-extension. The idea is the same as corollary 6.2 below, and works because in a model of PV, as with $\mathrm{PA}^{\mathrm{top}}$, we can take the definable closure of a set and still get a model of PV.

For the next lemma, recall that $u(e, x, c)$ is the universal PV function symbol, which calculates the output of the Turing machine with code e run on input x for time $|c|$.

Lemma 5.1 *Let $M \vDash \mathrm{PV}$. Suppose $\varepsilon, a \in M \setminus \mathbb{N}$, $\varepsilon < |a|$, $\#a$ is not cofinal in M and $M \vDash \forall c\, \mathrm{PHP}^{a^\varepsilon}_{a^{2\varepsilon}}(u(x, 0, c))$. Then there is $N \vDash \mathrm{PV}$, $(M \restriction a) \subseteq_e N \subseteq (M \restriction \#a)$ such that for some $v < a^{2\varepsilon}$ in M, $v \notin N$. Furthermore $N \not\cong M \restriction \#a$.*

Proof. Let N be the closure of $[0, a)$ in M under all PV function symbols, so $N \vDash \mathrm{PV}$, since PV is universally axiomatized. We that claim N is as required. Let $2^t > \#a$. By WPHP, for some $v < a^{2\varepsilon}$ in M,

$$M \vDash \forall y {<} a^\varepsilon\, u(y, 0, 2^t) \neq v.$$

Consider the type

$$\Gamma(z) = \{\forall \bar{b} {<} a\, f(\bar{b}) \neq z : f \text{ a PV function symbol}\}.$$

No element of N realizes this type, but v does realize it. Otherwise, we would have a function f running in time $|a|^k$, for some $k \in \mathbb{N}$, such that $v = f(\bar{b})$ for some $\bar{b} < a$. Then there is clearly some y of the form $(e, f, k, b_1, \ldots, b_n) < a^\varepsilon$, with $e, f \in \mathbb{N}$, such that the universal machine run on y simulates f with input b and halts in time $|a|^k$. Hence $v = u(y, 0, 2^t)$, a contradiction. $\Box$

A PV formula is an equality of the form $f(\bar{x}) = g(\bar{x})$ where f and g are PV function symbols.

Theorem 5.2 *Assume that* PV *proves* BB*, the* sharply bounded collection scheme

$$\forall x \forall y\, ,\, \forall 1 {\leqslant} i {\leqslant} |x|\, \exists z {<} y\, \phi(i, z) \to \exists w\, \forall 1 {\leqslant} i {\leqslant} |x|\, \phi(i, w_i)$$

for PV *formulas with parameters. Then* PV *also proves that surjective* WPHP *with parameters implies injective* WPHP *with parameters.*

Proof. Let $M \vDash \mathrm{PV}$ satisfy surjective WPHP(PV), but be such that for some $a, c \in M$ and some PV function symbol f, $M \vDash f(c, x) : a^2 \hookrightarrow a$. Choose $b > a, c$ so that $b = 2^\beta$ for some β. Amplify f to an injection $g : b\#b \hookrightarrow b$ with parameters $< b\#b$. Taking an elementary extension if necessary, assume that $\#b$ is not cofinal in M.

Take ε small and nonstandard. By lemma 5.1 we can find $N \vDash \mathrm{PV}$, a submodel of M with $[0, b\#b) \subseteq_e N$ but $v \notin N$ for some $v < (b\#b)^\varepsilon$.

Define a PV function symbol H (with parameters $< b\#b$) mapping $x < (b\#b)^\varepsilon$ to the unique $y < b^\varepsilon$ such that

$$\forall 1 \leqslant i \leqslant \varepsilon \; y_i = g([x]_i)$$

where y_i is the numeral $< b$ consisting of the β bits occurring in places $((i-1)\beta+1), \ldots, i\beta$ of y, and $[x]_i$ the numeral $< b\#b$ consisting of the β^2 bits occurring in places $((i-1)\beta^2+1), \ldots, i\beta^2$ of x. This is an alternative way of amplifying g, and as before by the $\forall\Sigma_1^b$ conservativity of S_2^1 over PV, H is an injection $(b\#b)^\varepsilon \hookrightarrow b^\varepsilon$ in both M and N.

Let $u = H(v)$. Then in M, $\forall 1 \leqslant i \leqslant \varepsilon \, \exists! w < b\#b \, g(w) = u_i$, since we may take $w = [v]_i$ and g is injective. Thus this formula is also true in N, since $M \restriction b\#b$ and $N \restriction b\#b$ are isomorphic. However $v \notin N$ and H is injective, so in N we have

$$\forall 1 \leqslant i \leqslant \varepsilon \, \exists! w < b\#b \, g(w) = u_i \, \wedge \, \forall 1 \leqslant x \leqslant (b\#b)^\varepsilon \, H(x) \neq u \tag{ii}$$

and this contradicts sharply bounded collection. $\square$

Notice that only a very weak form of collection is needed to contradict (ii).

Theorem 5.3

1. **(Zambella [22])** *If* $\mathrm{PV} + \mathrm{BB} \vdash S_2^1$, *then* $\mathrm{PV} \vdash S_2^1$.

2. **(Zambella [22], Buss [3])** *If* $\mathrm{PV} \vdash S_2^1$, *then* $\mathrm{PV} \vdash S_2$.$\square$

Corollary 5.4 *If RSA is secure against deterministic polynomial time attack, then* $\mathrm{PV} + \mathrm{BB} \nvdash S_2^1$. *If RSA is secure against randomized polynomial time attack, then* $\mathrm{PV} \nvdash \mathrm{BB}$.

Proof. The second part is by theorem 5.2. The first part is because if $\mathrm{PV} \vdash S_2$, then the injective WPHP is provable in PV, since it is provable in S_2. Hence it can be witnessed in polynomial time. $\square$

6 WPHP in recursively saturated structures

We made use in lemma 5.1 of a connection between the surjective WPHP and definability, in models of PV. In fact this connection holds in a very general setting. From now on we will begin to consider arbitrary definable functions, not just bounded ones (although in $\mathrm{PA}^{\mathrm{top}}$ these are the same thing). Notice that in $\mathrm{PA}^{\mathrm{top}}$ and $\mathrm{I}\Delta_0$ the surjective and injective WPHP for Δ_0 formulas are equivalent.

Lemma 6.1 *Let M be a recursively saturated structure with a definable subset R containing at least two elements 0, 1 which are named in the language. For any formula $\phi(\bar{y})$, if for all $k \in \mathbb{N}$ there is no parameter free definable surjection from R^k onto $\phi(M) \cup \{0\}$, then there is $\bar{c} \in \phi(M)$ such that $\bar{c}$ is not definable with parameters from R. The converse also holds.*

Proof. Recursively enumerate all parameter free formulas as $\psi_1(\bar{x}, \bar{y})$, $\psi_2(\bar{x}, \bar{y}), \ldots$ where we assume $\bar{x}$ has arity at most i in ψ_i. Let $\Gamma(\bar{y})$ be the type

$$\{\phi(\bar{y})\} \cup \{\bigwedge_{i \leqslant m} \forall \bar{x} \subseteq R \text{ “}\bar{y} \text{ is not unique such that } \psi_i(\bar{x}, \bar{y})\text{”} : m \in \mathbb{N}\}$$

Suppose Γ is not finitely satisfied in M. Then there is a finite sequence of formulas $\psi_1, \ldots, \psi_m$ (where we now assume that each $\bar{x}$ has arity m) such that for each $\bar{c}$ satisfying ϕ, there exist $i \leqslant m$ and $\bar{d} \subseteq R$ for which $\bar{c}$ is unique such that $\psi_i(\bar{d}, \bar{c})$.

Define a surjection $f : R^{2m} \to \phi(M) \cup \{0\}$ as follows: given $(a_1, \ldots, a_m, d_1, \ldots, d_m) \in R^{2m}$, if for some $i \leqslant m$, $a_1 = \ldots = a_i = 0$, $a_{i+1} = \ldots = a_m = 1$ and $M \vDash \exists! \, \bar{y} \phi(\bar{y}) \wedge \psi_i(\bar{d}, \bar{y})$ then map $(\bar{a}, \bar{d})$ to that unique $\bar{y}$; otherwise map it to 0. This contradicts the assumption that there is no such surjection.

Hence Γ, since it is recursive, is realized in M. Clearly any element realizing it is not definable from R.

The converse direction is trivial. $\square$

Corollary 6.2 *Let M be a recursively saturated model of $\mathrm{PA}^{\mathrm{top}}$ of the form $[0, b)$ and let $a \in M$ be such that $a^k < b$ for all $k \in \mathbb{N}$. Suppose $M \vDash \mathrm{PHP}_b^{a^k}(\Delta_0)$, for every $k \in \mathbb{N}$. Then M is not relatively categorical over $[0, a)$ with respect to $Th(M)$.*

Proof. Let $K = K([0, a); M)$, the definable closure of $M \restriction a$ in M, which is elementarily equivalent to M but omits the type “y is not definable from $[0, a)$”. This type is realized in M, by lemma 6.1. So M and K are both end-extensions of $M \restriction a$, but are not isomorphic. $\square$

This is how the pigeonhole principle is typically used in the model theory of arithmetic, see for example [7], or chapter IV of [5]. We can now write down a model-theoretic characterization of the provability of WPHP in $\mathrm{I}\Delta_0$:

Corollary 6.3 $\mathrm{I}\Delta_0 \vdash \forall x\, \mathrm{PHP}^{x}_{x^2}(\Delta_0)$ *if and only if for every recursively saturated model M of* $\mathrm{PA}^{\mathrm{top}}$ *of the form* $[0, b)$ *and every $a \in M$ with $a^{\mathbb{N}} < b$, there is more than one end-extension of $M \restriction a$ to a model of* $\mathrm{PA}^{\mathrm{top}}$ *of the form* $[0, b)$ *(we assume b is definable from parameters in* $[0, a)$*).*$\square$

The last result we present in this section is Gaifman's coordinatization theorem, that (assuming WPHP) to get two different models with the same restriction to R, we do not need definable Skolem functions, but only rigidity over R. In normal arithmetical situations we will always have this, in a very strong sense.

Lemma 6.4 *Suppose in $K \vDash BASIC'$ we can define a parameter-free function $bit(x, i)$ and prove that no two numbers in K encode the same sequence of bits. Then for any $b \in K$, no two elements of K smaller than b share the same type over* $[0, |b|)$*. Hence $K \restriction b$ is rigid over $K \restriction |b|$, and by repeating the argument $K \restriction b$ is rigid over $K \restriction | \dots |b| \dots |$, for any nesting of $|\ |$s (see Kaye [7]).* $\square$

For clarity of presentation we will not prove the strongest version of the coordinatization theorem here, and will instead directly use this property that an element is uniquely given by its type.

Lemma 6.5 *Suppose M is a structure with a definable subset R such that no two elements of M have the same type over R. Then the principal types over R are realized in M by precisely the elements of M that are definable from R.*

Proof. Suppose $p(x) = tp_M(c; R)$ has a principal formula $\phi(x)$ with parameters from R. Then we must have $\exists!\, x \phi(x)$, or two elements of M would have the same type over R. Hence c is definable from R. $\square$

Theorem 6.6 (Gaifman [6]) *Suppose M is a countable recursively saturated structure in a language with no function symbols and with a definable subset R which contains all the elements named by constants. Suppose that R contains at least two elements* 0, 1 *named in the language, that there is* *no parameter-free definable surjection from any standard power of R onto M and that no two elements of M have the same type over R. Then there is $N \equiv M$ such that $N \restriction R \cong M \restriction R$ but this isomorphism cannot be extended to an isomorphism $N \cong M$.*

Proof. List the elements of $R(M)$ as $\bar{r}$. By lemma 6.1 there is $c \in M$ not definable from $\bar{r}$, and by lemma 6.5 the type $p(x) = tp_M(c; \bar{r})$ is not principal. The type

$$q(y) = \{R(y)\} \cup \{y \neq r : r \in \bar{r}\}$$

is not principal either, since it is not realized in M. Hence there is a structure $(N, \bar{r}) \equiv (M, \bar{r})$ omitting both p and q. Since $(N, \bar{r})$ omits q, $N \restriction R$ is isomorphic to $M \restriction R$. Since $(N, \bar{r})$ omits p, this isomorphism cannot be extended to an isomorphism $(N, \bar{r}) \cong (M, \bar{r})$. □

Corollary 6.7 *Let M be a countable recursively saturated model of S^1_0 of the form $[0, b)$, and let $a \in M$ be such that $|b| < a$, $a^{\mathbb{N}} < b$ and M is relatively categorical over $[0, a)$ with respect to the complete theory of M. Then $M \models \neg \mathrm{PHP}^{a^k}_b(f)$ for some $k \in \mathbb{N}$ and some definable function f.* □

This is not a very good converse to theorem 3.7, since it uses the complete theory of M and we cannot limit the quantifier complexity of f. The ideal result would be something like: relative categoricity with respect to S^1_0 implies failure of surjective WPHP for a Σ^b_1 function. However it is not clear whether this is attainable. It would mean that in S^1_2 surjective WPHP(Σ^b_1) implies injective WPHP(Σ^b_1), and we have shown that if this were true for PV function symbols (rather than for Σ^b_1 formulas) then we could crack RSA.

7 Cardinality

There are many combinatorial principles in arithmetic which are normally proved by counting arguments, but which turn out only to need approximate rather than precise counting. There have been some successes in proving these in S_2 using the weak pigeonhole principle, which could be taken to say that, as far as definable functions are concerned, n^2 is bigger than n. See for example Pudlák's proof of the Ramsey theorem [20] or the proof that there are infinitely many primes [17]. It would be nice to be able to characterize the approximate counting available in bounded arithmetic and to give a uniform way of dealing with combinatorial proofs that make use of it.

If there is no definable map from a onto b, one would sometimes like to say that the "definable cardinality" of b is bigger than that of a; Krajíček has suggested developing this idea into a theory of the definable combinatorics of a structure [10], [12].

We present a simple application of Vaught's two-cardinal theorem (see [6]) to give a result in this direction, that in a countable, recursively saturated model of a theory with Skolem functions (such as $\mathrm{PA}^{\mathrm{top}}$), we can choose any definable set R and extend the model to one in which R is unchanged, hence still countable, but every other definable set has greater cardinality than R if and only if it has greater definable cardinality than any standard power of R. This may be of some use in formalizing approximate counting arguments in S_2, and leads to an interesting characterization of the polynomial size sets in models of S_2.

In our setting we can sharpen the two-cardinal theorem slightly, using resplendence. First, however, we will use the normal version to prove a second model-theoretic statement equivalent to the provability of WPHP in $\mathrm{I}\Delta_0$.

Theorem 7.1 $\mathrm{I}\Delta_0(\alpha) \vdash \forall x\, \mathrm{PHP}^{x}_{x^2}(\Delta_0(\alpha))$ *if and only if for every countable model K of $\mathrm{PA}^{\mathrm{top}}(\alpha)$ of the form $[0, b)$ and every $a \in K$ for which $a^{\mathbb{N}} < b$ there is some uncountable $J \succeq K$ in which $J \restriction a$ is countable.*

Proof. For one direction, extend K to a recursively saturated structure K', and let I be the definable closure of $K' \restriction a$ in K'. Then as in the previous section, $I \preceq K'$, $I \restriction a = K' \restriction a$ but by WPHP, $I \subset K'$ so we can apply the two-cardinal theorem to get J. For the other direction, if there is a definable surjection $a \twoheadrightarrow a^2$ then we can amplify it as in the proof of lemma 2.1 to a surjection $a \twoheadrightarrow b$. So J and $J \restriction a$ must have the same cardinality. $\square$

Lemma 7.2 (Resplendence [8]) *Suppose M is a countable recursively saturated L-structure in a recursive language L, the language L' is a recursive extension of L and T is a recursively axiomatized L'-theory. Then, if $Th(M) + T$ is consistent, there is an expansion of M to L' satisfying T.* $\square$

Lemma 7.3 *Let L be a recursive language, $\phi(x)$, $\psi(x)$ parameter free L-formulas and M, N countable L-structures such that M is recursively saturated, $N \preceq M$, $\phi(N) = \phi(M)$ and $\psi(N) \subset \psi(M)$. Then there is $M' \succeq M$ such that $\phi(M) = \phi(M')$, $\psi(M) \subset \psi(M')$ and $M \cong M'$.*

Proof. Let $L^+ = L \cup \{H, f\}$ where f is a one-place function and H is a one-place predicate. Writing χ^H for the relativization of χ to H, let T be the following set of sentences:

1. H is the range of f;
2. $\forall \bar{x} \subseteq H(\chi^H(\bar{x}) \leftrightarrow \chi(\bar{x}))$ for each L-formula χ;
3. $\forall \bar{x}\,(\theta(\bar{x}) \leftrightarrow \theta(f(\bar{x})))$ for each atomic L-formula θ;
4. $\forall x\,(\phi(x) \rightarrow H(x))$;
5. $\exists x\,(\psi(x) \wedge \neg H(x))$.

By the proof of Vaught's two-cardinal theorem (in [6]), there are structures U, V with $M \preceq V$ such that $U \preceq V$, $\phi(U) = \phi(V)$, $\psi(U) \subset \psi(V)$ and there is an isomorphism $V \cong U$. So we may expand V to an L^+ structure satisfying T by interpreting H as membership of U and f as the isomorphism $V \cong U$.

T is a recursive theory, and we have shown that $Th(M) \cup T$ is consistent. Hence by lemma 7.2 we can expand M to an L^+ structure satisfying T.

So M is isomorphic to an elementary submodel M^- of itself, with $\phi(M^-) = \phi(M)$ and $\psi(M^-) \subset \psi(M)$. By identifying M with M^-, we can find an elementary extension M' of M with the properties required. $\square$

Lemma 7.4 *The union of a countable elementary chain $\{M_\gamma : \gamma < \delta\}$ of countable, recursively saturated structures isomorphic to M_0 is a countable, recursively saturated structure isomorphic to M_0.*

Proof. The union is recursively saturated and realizes the same types as M_0. $\square$

Theorem 7.5 *Let M be a countable, recursively saturated structure with definable Skolem functions in a recursive language. Let R be a definable subset of M containing at least two elements. Then we can find $N \succeq M$ such that $R(N) = R(M)$ but the countable definable subsets $\phi(N)$ of N (with parameters from N) are precisely those for which there is a definable surjection (with parameters from N) from some standard power of $R(N)$ onto $\phi(N)$. Every other definable subset is uncountable.*

Proof. By the existence of Skolem functions there are two definable elements of R; add names 0, 1 to the language for these. We will construct an elementary chain $\{M_\beta : \beta < \omega_1\}$, with $M_0 = M$, of pairwise isomorphic structures such that for all $\beta < \omega_1$, $R(M_\beta) = R(M_0)$ and for any formula $\phi(x)$ with parameters from M_β, if there is no definable surjection in M_β from any standard power of R onto $\phi(M_\beta)$ then $\phi(M_{\beta+1}) \supset \phi(M_\beta)$. By lemma 7.4 we can put $M_\delta = \bigcup_{\beta<\delta} M_\beta$ for δ a limit.

For the successor step, enumerate as $\phi_1(x), \phi_2(x), \dots$ the formulas with parameters from M_β for which there is no surjection (with parameters from M_β) from any standard power of $R(M_\beta)$ onto $\phi_i(M_\beta)$, and for which $\phi_i(M_\beta)$ is non-empty. Let $\bar{m}_i \subseteq M_\beta$ be the tuple of parameters appearing in ϕ_i. Writing R for $R(M_\beta) = R(M_0)$, let $K_1 := K(R \cup \bar{m}_1; M_\beta)$ be the definable closure of $R \cup \bar{m}_1$ in M_β. There is no surjection with parameter $\bar{m}_1$ from any standard power of R onto $\phi_1(M_\beta)$, so there is certainly no surjection onto $\phi_1(M_\beta) \cup \{0\}$. Thus, temporarily adding $\bar{m}_1$ to the language, by lemma 6.1 there is $c \in \phi_1(M_\beta)$ not definable from R with parameter $\bar{m}_1$; so $c \notin K_1$.

Now $K_1 \preceq M_\beta$, $R(K_1) = R(M_\beta)$ and $\phi_1(M_\beta) \supset \phi_1(K_1)$ so by lemma 7.3 there is $M_\beta^1 \succeq M_\beta$ with $M_\beta^1 \cong M_\beta$, $R(M_\beta^1) = R(M_\beta)$ and $\phi_1(M_\beta^1) \supset \phi_1(M_\beta)$. Similarly, if we let $K_2 = K(R \cup \bar{m}_2; M_\beta)$ then $\phi_2(M_\beta^1) \supset \phi_2(K_2)$ so we can find $M_\beta^2 \succeq M_\beta^1$ with $M_\beta^2 \cong M_\beta^1$, $R(M_\beta^2) = R(M_\beta)$ and $\phi_2(M_\beta^2) \supset \phi_2(M_\beta^1)$. Repeating this step for $\phi_3, \phi_4, \dots$ gives an elementary chain $M_\beta \preceq M_\beta^1 \preceq M_\beta^2 \preceq \dots$ and taking the union of the chain gives us, by lemma 7.4, $M_{\beta+1} \cong M_\beta$ with the properties required.

Let $N = \bigcup_{\beta<\omega_1} M_\beta$. Suppose $\phi(x)$ is a formula with parameters $\bar{n} \subseteq N$ such that there is no surjection definable with parameters from N from any standard power of R onto $\phi(N)$. Suppose $\bar{n} \subseteq M_\beta$ for some $\beta < \omega_1$. Then for each $\beta \leqslant \gamma < \omega_1$, there is no surjection with parameters from M_γ from any standard power of R onto $\phi(M_\gamma)$, by elementariness. So by construction $\phi(M_{\gamma+1}) \supset \phi(M_\gamma)$. Hence $\phi(N)$ is uncountable.

Conversely, if there is a surjection from R^k onto $\phi(N)$, for $k \in \mathbb{N}$, then $\phi(N)$ must be countable because R^k is. $\Box$

Corollary 7.6 *If K is a countable, recursively saturated model of* $\mathrm{PA}^{\mathrm{top}}(\alpha)$ *of the form $[0, b)$ containing an element a such that $K \vDash \mathrm{PHP}_b^{a^k}(\Delta_0(\alpha))$ for every $k \in \mathbb{N}$, then there is $J \succeq K$ with $J \restriction a = K \restriction a$ but with $J \restriction c$ uncountable for every $c > a^{\mathbb{N}}$.* $\Box$

There are similar, rather stronger results for full Peano arithmetic in Paris and Mills [15], but these make heavy use of precise counting.

One cannot in general repeat this increase in size more than once. For suppose we have a countable recursively saturated structure $K \vDash \mathrm{PA}^{\mathrm{top}} +$

$\forall x\, \mathrm{PHP}^{x}_{x^2}(\Delta_0)$ of the form $[0, b)$, with $|b|^{\mathbb{N}} < b$. We can find $J \succeq K$ of cardinality $\aleph_1$ with $J \restriction |b| = K \restriction |b|$. If we could go on to find an elementary extension I of J with cardinality $\aleph_2$ and with $I \restriction |b| = K \restriction |b|$, then this would imply a violation of the continuum hypothesis, since the function that takes an element of I to its set of non-zero bits is an injection from $[0, b)$ into the power set of $[0, |b|)$.

So if we could find a way of adding a predicate α to a model of $\mathrm{PA}^{\mathrm{top}}$ which ensured either that we could not increase the cardinality of part of the model in this way, or that, whenever we could increase the cardinality, we could do so more than once, we would have gone some way towards showing that $\mathrm{WPHP}(\alpha)$ is independent of $\mathrm{I}\Delta_0(\alpha)$.

We give one more application of the two-cardinal theorem.

Corollary 7.7 *Suppose M is a countable model of S_2, $a \in M$ and $\#a$ is cofinal in M. Then there exists an uncountable $N \succeq_{\Pi_1} M$ in which the coded sets are precisely the countable bounded Δ_0 sets.*

Proof. Let M' be a recursively saturated extension of M, so $b > \#a$ for some $b \in M'$. Let $B = M' \restriction b$, so $B \vDash \mathrm{PA}^{\mathrm{top}}$ and B is recursively saturated. Let $C \succeq B$ be given by theorem 7.5, taking R to be the definable set "$x < |a|$". Let $N = C \restriction \#a$.

Note that for each $k \in \mathbb{N}$, $M \restriction 2^{|a|^k} \preceq N \restriction 2^{|a|^k}$. Hence $M \preceq_{\Delta_0} N$ and if $N \vDash \exists x\, \theta(\bar{m}, x)$ for θ a Δ_0 formula and $\bar{m} \subseteq M$, we must have $N \vDash \exists x < 2^{|a|^k}\, \theta(\bar{m}, x)$ for some $k \in \mathbb{N}$; so $M \vDash \exists x < 2^{|a|^k}\, \theta(\bar{m}, x)$. This shows that $M \preceq_{\Pi_1} N$.

Now suppose S is a subset of N coded as a sequence $(\sigma)_1, \ldots, (\sigma)_l$ for some $\sigma \in N$. Then $l < |a|^k$ for some $k \in \mathbb{N}$, and $N \restriction |a|^k$ is countable, so S must be countable.

Conversely, suppose that $S \subseteq N \restriction 2^{|a|^k}$ is countable and is defined by a Δ_0 formula $\phi(x)$. Then S is also definable by ϕ in C, so by the construction of C there exist $l \in \mathbb{N}$ and a definable function f such that f is a surjection from $|a|^l$ onto S.

Since S is bounded by $2^{|a|^k}$, we have that $C \vDash \forall i < |a|^l\, f(i) < 2^{|a|^k}$. So by comprehension in C, there is some $\sigma < 2^{|a|^{k+l}}$ in C with $C \vDash \forall i < |a|^l\, f(i) = (\sigma)_i$. Thus $C \vDash \forall x < 2^{|a|^k},\ \phi(x) \leftrightarrow \exists i < |a|^l\, (\sigma)_i = x$. This is a Δ_0 formula, so is also true in N. Hence S is coded in N, by σ. $\square$

Acknowledgements

I would like to thank Alex Wilkie for his guidance and Jan Krajíček for the many conversations which led to this work, and for his comments on earlier versions of this paper. I would also like to thank Jan Krajíček and the Mathematical Institute in Prague for their hospitality on my visits to the Czech Republic.

References

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

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

[3] S. Buss. Relating the bounded arithmetic and polynomial time hierarchies. *Annals of Pure and Applied Logic*, 75(1–2):67–77, 1995.

[4] O. Goldreich. *Foundations of Cryptography (Fragments of a Book).* Unpublished, 1995.

[5] P. Hájek and P. Pudlák. *The Metamathematics of First Order Arithmetic.* Springer, 1993.

[6] W. Hodges. *Model Theory.* Cambridge University Press, 1993.

[7] R. Kaye. A galois correspondence for countable recursively saturated models of peano arithmetic. In R. Kaye and D. Macpherson, editors, *Automorphisms of First-Order Structures*, pages 293–312. Clarendon Press, 1994.

[8] R. Kaye and D. Macpherson. Recursive saturation. In R. Kaye and D. Macpherson, editors, *Automorphisms of First-Order Structures*, pages 243–256. Clarendon Press, 1994.

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

[10] J. Krajíček. Uniform families of polynomial equations over a finite field and structures admitting an Euler characteristic of definable sets. *Proceedings of the London Mathematical Society*, 81(2):257–284, 2000.

[11] J. Krajíček and P. Pudlák. Some consequences of cryptographical conjectures for S^1_2 and EF. *Information and Computation*, 140(1):82–89, 1998.

[12] J. Krajíček and T. Scanlon. Combinatorics with definable sets: Euler characteristics and Grothendieck rings. *Bulletin of Symbolic Logic*, 3(3):311–330, 2000.

[13] 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.

[14] J. Paris and C. Dimitracopoulos. Truth definitions for Δ_0 formulae. In *Logic and Algorithmic*, number 30 in Monographies de l'Enseignement Mathématique, pages 317–329. Université de Genève, 1982.

[15] J. Paris and G. Mills. Closure properties of countable non-standard integers. *Fundamenta Mathematica*, 103:205–215, 1979.

[16] J. Paris and A. Wilkie. Counting problems in bounded arithmetic. In *Methods in Mathematical Logic*, number 1130 in Lecture Notes in Mathematics, pages 317–340. Springer, 1985.

[17] 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.

[18] A. Pillay. $\aleph_0$-categoricity over a predicate. *Notre Dame Journal of Formal Logic*, 24(4):527–536, 1983.

[19] A. Pillay and S. Shelah. Classification theory over a predicate I. *Notre Dame Journal of Formal Logic*, 26(4):361–376, 1985.

[20] P. Pudlak. Ramsey's theorem in bounded arithmetic. In E. Börger, H. Kleine Büning, M. Richter, and W. Schönfeld, editors, *Computer Science Logic: Proceedings of the 4th Workshop, CSL '90*. 1991.

[21] A. Shamir R. Rivest and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21:120–126, 1978.

[22] D. Zambella. Notes on polynomially bounded arithmetic. *Journal of Symbolic Logic*, 61(3):942–966, 1996.