How strong is Ramsey's theorem if infinity can be weak?

Leszek Aleksander Kołodziejczyk[*],
Katarzyna W. Kowalik[*],
Keita Yokoyama[†]

January 16, 2021

Abstract

We study the first-order consequences of Ramsey's Theorem for k-colourings of n-tuples, for fixed $n, k \geq 2$, over the relatively weak second-order arithmetic theory RCA_0^*. Using the Chong-Mourad coding lemma, we show that in a model of RCA_0^* that does not satisfsy Σ_1^0 induction, RT_k^n is equivalent to its relativization to any proper Σ_1^0-definable cut, so its truth value remains unchanged in all extensions of the model with the same first-order universe.

We give a complete axiomatization of the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}_k^n$ for $n \geq 3$. We show that they form a non-finitely axiomatizable subtheory of PA whose Π_3 fragment coincides with $\mathrm{B}\Sigma_1 + \exp$ and whose $\Pi_{\ell+3}$ fragment for $\ell \geq 1$ lies between $\mathrm{I}\Sigma_\ell \Rightarrow \mathrm{B}\Sigma_{\ell+1}$ and $\mathrm{B}\Sigma_{\ell+1}$. We also give a complete axiomatization of the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}_k^2 + \neg\mathrm{I}\Sigma_1$. In general, we show that the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}_k^2$ form a subtheory of $\mathrm{I}\Sigma_2$ whose Π_3 fragment coincides with $\mathrm{B}\Sigma_1 + \exp$ and whose Π_4 fragment is strictly weaker than $\mathrm{B}\Sigma_2$ but not contained in $\mathrm{I}\Sigma_1$.

Additionally, we consider a principle Δ_2^0-RT_2^2 which is defined like RT_2^2 but with both the 2-colourings and the solutions allowed to be Δ_2^0-sets rather than just sets. We show that the behaviour of Δ_2^0-RT_2^2 over $\mathrm{RCA}_0 + \mathrm{B}\Sigma_2^0$ is in many ways analogous to that of RT_2^2 over RCA_0^*, and that $\mathrm{RCA}_0 + \mathrm{B}\Sigma_2^0 + \Delta_2^0$-$\mathrm{RT}_2^2$ is Π_4- but not Π_5-conservative over $\mathrm{B}\Sigma_2$. However, the statement we use to witness failure of Π_5-conservativity is not provable in $\mathrm{RCA}_0 + \mathrm{RT}_2^2$.

Over the last two decades, much of the research in reverse mathematics has concerned the logical strength of various principles from Ramsey theory. One of the challenging problems in this area has been to characterize the first-order consequences of Ramsey's Theorem for pairs. Despite significant progress (e.g. [4, 7, 26]), this remains open. In particular, it is not known whether Ramsey's Theorem for pairs and a fixed number of colours is Π_1^1 conservative over the Σ_2^0 collection scheme.

---

[*] Institute of Mathematics, University of Warsaw, `{lak, katarzyna.kowalik}@mimuw.edu.pl`. Partially supported by grant no. 2017/27/B/ST1/01951 of the National Science Centre, Poland.

[†] School of Information Science, Japan Advanced Institute of Science and Technology, `y-keita@jaist.ac.jp`. Partially supported by JSPS KAKENHI grant no. 19K03601.

In this paper, we study the first-order strength of Ramsey's Theorem – both for pairs and for longer tuples of fixed length – over a weaker base theory than the one normally used in reverse mathematics. Our base theory, RCA_0^*, differs from the usual system RCA_0 in that the Σ^0_1 induction axiom of the latter is replaced by induction for bounded formulas only.

The study of RCA_0^* was initiated in [30] and continued in a number of later papers, e.g. [13, 31, 20, 10]. In the context of Ramsey theory, it is important that Σ^0_1 induction is needed to show that each infinite set has arbitrarily large finite subsets. Hence, over RCA_0^* the infinite homogeneous sets witnessing various principles might be so sparse that they have "strictly smaller cardinality" than $\mathbb{N}$, so the principles can become weaker. Indeed, Yokoyama [33] showed that for each fixed n, k, RCA_0^* extended by Ramsey's Theorem for n-tuples and k colours, RT^n_k, is Π_2-conservative over $\mathrm{I}\Delta_0 + \exp$. We are able to go quite a bit beyond that result.

Recent work of Belanger [3] has demonstrated that the study of reverse mathematics over RCA_0^* is relevant to the traditional RCA_0 framework as well. In fact, a large part of our original motivation for studying Ramsey's Theorem over RCA_0^* was the desire to understand whether it can help in understanding RT^2_2 over RCA_0. The jury is still out on that. However, it has turned out that Ramsey theory in RCA_0^* is a highly interesting topic in its own right. It gives rise to new examples of principles that are partially conservative but not Π^1_1-conservative over the base theory, and it has intriguing connections to the model theory of first-order arithmetic.

After discussing the necessary background in a preliminary Section 1, we begin the paper proper in Section 2 by proving that in models of RCA_0^* that are *not* models of RCA_0, RT^n_k is equivalent to its relativizations to Σ^0_1-definable cuts. One consequence of that result is that in some models of RCA_0^*, Ramsey's Theorem is computably true. This is not the case in the standard model of arithmetic or in any other model of RCA_0.

In Section 3, we use the equivalence from Section 2 to give an axiomatization of the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}^n_k$ where $n \geq 3$. In each case, this turns out to be an unusual fragment of Peano Arithmetic that is Π_3- but not Π_4-conservative over $\mathrm{B}\Sigma_1 + \exp$. Moreover, it is not contained in $\mathrm{I}\Sigma_\ell$ for any ℓ.

We then consider Ramsey's Theorem for pairs. We are not able to give a complete axiomatization of its first-order consequences over RCA_0^*, but in Section 4 we obtain some partial results. In particular, we do axiomatize these consequences over $\neg\mathrm{I}\Sigma_1$. We also show that $\mathrm{RCA}_0^* + \mathrm{RT}^2_2$ is not conservative over (the lightface theory) $\mathrm{I}\Sigma_1$.

Then, in Section 5, we take a look at the question whether our results say anything about Ramsey's Theorem for pairs over RCA_0. We consider a principle that can be viewed as a "jumped version" of RT^2_2, and we show that it is not Π_5-conservative over $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2$. We also show that the most obvious sentence witnessing the lack of conservativity is unprovable in $\mathrm{RCA}_0 + \mathrm{RT}^2_2$. However, the proof of unprovability, which is based on a possibly unexpected technique (proof speedup), no longer works for slightly weaker sentences.

1 Preliminaries

We assume that the reader has some familiarity with fragments of second-order arithmetic, as described in [29] or [14]. We also assume familiarity with some basic facts about first-order arithmetic and its models – most or all of the necessary information can be found in [14], and [17] covers more than enough.

The symbol ω stands for the set of standard natural numbers. In contrast, $\mathbb{N}$ stands for the set of natural numbers as formalized in the given theory we are studying – in a nonstandard model, this is the first-order universe of the model.

Notation like Σ^0_ℓ, Π^0_ℓ represents the usual formula classes defined in terms of first-order quantifier alternations, but allowing second-order free variables. On the other hand, notation without the superscript 0, like Σ_ℓ, Π_ℓ, represents analogously defined classes of first-order, or "lightface", formulas – that is, without any second-order variables at all. If we want to specify the second-order parameters appearing in a Σ^0_ℓ formula, we use notation like $\Sigma_\ell(\bar{X})$. We extend these conventions to naming theories: thus, for example, $\mathrm{B}\Sigma^0_2$ is the fragment of second-order arithmetic axiomatized by Δ^0_0 induction and Σ^0_2 collection, whereas $\mathrm{B}\Sigma_2$ is the fragment of first-order arithmetic axiomatized by Δ_0 induction and Σ_2 collection.

Remark. In formulating the results presented in the paper, we had to make the decision whether to state them in purely arithmetical, lightface, form, or in Π^1_1 form, allowing the appearance of (typically universally quantified) second-order parameters. We opted to use the lightface version most of the time, with the tacit understanding that our results of the form "first-order scheme T implies first-order sentence ψ" (as for instance Lemma 6) typically have a natural relativization of the form "for all X, $T(X)$ implies $\psi(X)$" that can be proved by essentially the same argument. On the other hand, we did allow second-order parameters whenever we found it advisable, for instance because it was necessary to state the result properly (as in Theorem 14) or needed for later applications (as in the case of Theorem 3).

Recall that for $\ell \geq 1$ the theory $\mathrm{I}\Sigma_\ell$ proves (in fact, is equivalent to over $\mathrm{I}\Delta_0$) the scheme of *strong* Σ_ℓ *collection*, that is,

$$\forall v\, \exists w\, \forall x \leq v\, (\exists y\, \sigma(x,y) \Rightarrow \exists y \leq w\, \sigma(x,y)),$$

where $\sigma(x,y)$ is a Σ_ℓ formula, possibly with parameters.

The theory RCA^*_0 is obtained from RCA_0 by weakening the $\mathrm{I}\Sigma^0_1$ axiom to $\mathrm{B}\Sigma^0_1$ and adding the axiom exp that explicitly guarantees the totality of exponentiation. The first-order consequences of RCA^*_0 are axiomatized by $\mathrm{B}\Sigma_1 + \mathrm{exp}$.

When we consider a model $(M, \mathcal{X})$ of some fragment of second-order arithmetic (or simply work inside this fragment without reference to a specific model), a *set* is an element of the second-order universe, i.e. an element of $\mathcal{X}$. In contrast, a *definable set* is any subset of M that is definable in $(M, \mathcal{X})$, but does not have to belong to $\mathcal{X}$. A definable set is a Δ^0_ℓ*-definable set*, or simply a Δ^0_ℓ*-set* (resp., a Σ^0_ℓ*-definable set* or Σ^0_ℓ*-set*) if it happens to be definable by a Δ^0_ℓ (resp. Σ^0_n) formula. The notions of a Δ_ℓ*-set* and Σ_ℓ*-set* are defined analogously.

Since most of the models we study only satisfy Δ^0_1-comprehension, Δ^0_ℓ-sets for $\ell \geq 2$ and Σ^0_ℓ-sets for $\ell \geq 1$ will not always be sets. However, using appropriate universal formulas, we can quantify over Δ^0_ℓ- or over Σ^0_ℓ-sets using second-order quantifiers (e.g. "for every X, and every equivalent pair of a $\Sigma^0_\ell(X)$ and a $\Pi^0_\ell(X)$ formula, ..."). On the other hand, quantification over Δ_ℓ- or over Σ_ℓ-sets is first-order. We write Δ_ℓ-Def(M) (resp. Δ^0_ℓ-Def$(M, \mathcal{X})$) for the collection of Δ_ℓ-definable subsets of M (resp. the subsets of M that are Δ^0_ℓ-definable in $(M, \mathcal{X})$).

For $\ell \geq 1$, let $\mathrm{Sat}_\ell(x, y)$ be the usual universal Σ_ℓ formula and let $\mathrm{Sat}_\ell(x, y, X)$ be the usual universal Σ^0_ℓ formula with the unique second-order variable X. Then $0^{(\ell)}$ is the Σ_ℓ definable set $\{e : \mathrm{Sat}_\ell(e, e)\}$; we write $0'$ for $0^{(1)}$. Similarly, if A is a set, then $A^{(\ell)}$ is $\{e : \mathrm{Sat}_\ell(e, e, A)\}$; this notion is generalized in a natural way to the case where A is merely a definable set. Note that $\mathrm{B}\Sigma_\ell$ is enough to prove that $0^{(\ell+1)}$ and $(0^{(\ell)})'$ are mutually Δ_1-definable.

For $n, k \in \omega$, RT^n_k stands for the usual formulation of Ramsey's Theorem for pairs in second-order arithmetic: "for every function $f \colon [\mathbb{N}]^n \to k$, there is an infinite homogeneous set H for f". Importantly, "H is infinite" is understood here as "H is unbounded", i.e. for every $x \in \mathbb{N}$ there is $H \ni y \geq x$. If Σ^0_1 induction fails, this does not imply that H contains an x-element finite subset for every x. Ramsey's Theorem formulated in terms of the latter notion is easily seen to imply $\mathrm{I}\Sigma^0_1$ [33].

A *cut* in a model of arithmetic M is any subset $I \subseteq M$ which contains 0 and is closed downwards and under successor; note that if $I \neq M$, it will never be a "set" in the sense of belonging to whatever second-order arithmetic structure there might be on M. A *definable cut* is a cut that happens to be a definable set. If $(M, \mathcal{X}) \models \mathrm{RCA}^*_0$, and I is a Σ^0_1-definable cut in M, then there is an infinite set $A \in \mathcal{X}$ of cardinality I, i.e. $A = \{a_i : i \in I\}$ enumerated in increasing order.

For an element s of a model M, $(s)_{\mathrm{Ack}}$ stands for $\{a \in M : M \models a \in_{\mathrm{Ack}} s\}$, where $\in_{\mathrm{Ack}}$ is the usual Ackermann interpretation of set theory in arithmetic ("the a-th bit in the binary notation for s is 1"). Given a proper cut $I \subseteq M$, the collection $\mathrm{Cod}(M/I)$ of *subsets of I coded in M* is $\{(s)_{\mathrm{Ack}} \cap I : s \in M\}$. If M satisfies induction for any of the classes of formulas Γ that we consider in this paper, this will coincide with $\{A \cap I : A \text{ a } \Gamma\text{-definable subset of } M\}$.

The collection $\mathrm{Cod}(M/\omega)$ is commonly referred to as the *standard system* of M and denoted by $\mathrm{SSy}(M)$. The following combination of standard model-theoretic facts discussed in [17] and well-known results on RT^n_k presented e.g. in [14] will often be used without notice.

Fact. Let $\mathcal{S} \subseteq \mathcal{P}(\omega)$ be such that $(\omega, \mathcal{S}) \models \mathrm{WKL}_0$ (such a family $\mathcal{S}$ is known as a *Scott set*). If $\mathcal{S}$ is countable, then for every $\ell \geq 1$ there exists a model $M \models \mathrm{B}\Sigma_\ell$ such that ω is Σ_ℓ-definable in M and $\mathrm{SSy}(M) = \mathcal{S}$.

For each fixed $n \geq 2$, there exist countable Scott sets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that $(\omega, \mathcal{S}_1) \models \mathrm{RT}^n_2$ and $(\omega, \mathcal{S}_1) \not\models \mathrm{RT}^n_2$.

We will sometimes want to abuse notation and use $\mathrm{Cod}(M/I)$ for the collection of binary (as opposed to unary) relations on I coded in M, that is for $\{(s)_{\mathrm{Ack}} \cap \{\langle i, j\rangle : i, j \in I\} : s \in M\}$ where $\langle \cdot, \cdot \rangle$ is the usual Cantor pairing function. If I is not closed under multiplication, then such binary relations might not be elements of $\mathrm{Cod}(M/I)$ in the strict sense, but that should not lead to any confusion.

We define the iterated exponential function $\exp_n(x)$ by: $\exp_0(x) = x$, and $\exp_{n+1}(x) = 2^{\exp_n(x)}$.

2 Characterization in terms of cuts

In this section, we prove a basic result which underlies our subsequent analysis of Ramsey's Theorem over RCA_0^*: if Σ_1^0 induction fails but Σ_1^0 collection holds, then Ramsey's Theorem is equivalent to its own relativization to a proper Σ_1^0-definable cut. To prove this, we make use of an important fact about coding sets in models of collection.

Lemma 1 ([5]). *Let $(M, \mathcal{X}) \models \mathrm{RCA}_0^* + \mathrm{B}\Sigma_n^0$. Then for every pair of bounded disjoint Σ_n^0-definable sets $X, Y \subseteq M$ there exists $A \in \mathcal{X}$ such that $A \cap (X \cup Y) = X$.*

Corollary 2. *Let $(M, \mathcal{X}) \models \mathrm{RCA}_0^* + \mathrm{B}\Sigma_n^0$ and let $I \subseteq M$ be a proper cut in M. If $X \subseteq I$ is such that both X and $I \setminus X$ are Σ_n^0-definable sets, then $X \in \mathrm{Cod}(M/I)$.*

Theorem 3. *Let $(M, \mathcal{X}) \models \mathrm{RCA}_0^*$ and let $I \subseteq M$ be a Σ_1^0-definable proper cut in M. Then for every $n, k \in \omega$:*

$$(M, \mathcal{X}) \models \mathrm{RT}_k^n \textit{ iff } (I, \mathrm{Cod}(M/I)) \models \mathrm{RT}_k^n. \tag{1}$$

Proof. Let $(M, \mathcal{X})$ be a model of RCA_0^* and let $I \subseteq M$ be a Σ_1^0-definable proper cut. Let $A \in \mathcal{X}$ be an infinite subset of M which can be enumerated in increasing order as $\{a_i : i \in I\}$. We may assume that $0 \in A$. Fix standard n, k.

Suppose $(M, \mathcal{X}) \models \mathrm{RT}_k^n$. Let $f\colon [I]^n \to k$ be coded by $c \in M$. We can use f to define a colouring $\check{f}\colon [A]^n \to k$ in the following way:

$$\check{f}(a_{i_1}, \ldots, a_{i_n}) = f(i_1, \ldots, i_n).$$

In fact, it is easy to generalize the definition of $\check{f}$ to obtain a colouring of $[M]^n$, which we will continue to call $\check{f}$:

$$\check{f}(x_1, \ldots, x_n) = \begin{cases} f(i_1, \ldots, i_n) & \text{if } i_1 < \ldots < i_n \in I \text{ are such that} \\ & x_1 \in [a_{i_1}, a_{i_1+1}), \ldots, x_n \in [a_{i_n}, a_{i_n+1}), \\ 0 & \text{if there are no such } i_1, \ldots, i_n. \end{cases}$$

Note that $\check{f}$ is $\Delta_1(A, c)$-definable, so $\check{f} \in \mathcal{X}$. By RT_k^n, there exists an infinite $H \in \mathcal{X}$ homogeneous for $\check{f}$. By Corollary 2, the $\Sigma_1(H)$-definable set

$$\hat{H} = \{i \in I : H \cap [a_i, a_{i+1}) \neq \emptyset\}$$

is in $\mathrm{Cod}(M/I)$. Clearly, $\hat{H}$ is cofinal in I and homogeneous for f.

In the other direction, suppose $(I, \mathrm{Cod}(M/I)) \models \mathrm{RT}_k^n$. Consider a colouring $f\colon [M]^n \to k$. By Corollary 2, the colouring $\hat{f}\colon [I]^n \to k$ given by

$$\hat{f}(i_1, \ldots, i_n) = f(a_{i_1}, \ldots, a_{i_n})$$

is in $\mathrm{Cod}(M/I)$. Since $(I, \mathrm{Cod}(M/I)) \models \mathrm{RT}_k^n$, there is $\mathrm{Cod}(M/I) \ni H \subseteq I$ cofinal in I and homogeneous for $\hat{f}$. Then the set $\check{H} = \{a_i : i \in H\}$ is in $\mathcal{X}$ and it is an infinite subset of M homogeneous for f. □

Remark. Note that the left-hand side of the equivalence (1) in Theorem 3 does not depend on the choice of the cut I, while the right-hand side does not depend on $\mathcal{X}$, as long as I is Σ^0_1-definable in $(M, \mathcal{X})$. Thus, Theorem 3 means that over RCA^*_0, once $\mathrm{I}\Sigma^0_1$ fails, Ramsey's Theorem becomes in some sense a first-order property. In particular, it can be satisfied in some structures of the form $(M, \Delta_1\text{-}\mathrm{Def}(M))$ ("computably true in M"). We investigate this phenomenon further in the next two sections of the paper.

3 Ramsey for triples and beyond

We now use the characterization provided by Theorem 3 to study the first-order consequences of $\mathrm{RCA}^*_0 + \mathrm{RT}^n_k$ for $n \geq 3$. We begin with the easy but useful observation that, just like over RCA_0, the strength of Ramsey's Theorem for n-tuples does not increase if we consider a larger but fixed number of colours.

Lemma 4. *For each* $n, k \geq 2$, $\mathrm{RCA}^*_0 \vdash (\mathrm{RT}^n_k \Leftrightarrow \mathrm{RT}^n_{k+1})$.

Proof. Assume $\mathrm{RCA}^*_0 + \mathrm{RT}^n_k$ and let $f\colon [\mathbb{N}]^n \to k+1$. Consider the colouring $g\colon [\mathbb{N}]^n \to k$ given by $g(\bar{x}) = \min(f(\bar{x}), k-1)$. Let A be an infinite homogeneous set for g and let $\{a_i : i \in I\}$ be an increasing enumeration of A. (Here I may be either a proper Σ^0_1-definable cut or $\mathbb{N}$, depending on A.)

If A is j-homogeneous for g with $j < k-1$, then A is also j-homogeneous for f, so we are done. Otherwise, A is $(k-1)$-homogeneous for g, which means that $f\restriction_{[A]^n}$ takes at most the two values $k-1$ and k. Define a 2-colouring of $[\mathbb{N}]^n$ by:

$$\check{f}(x_1, \ldots, x_n) = \begin{cases} f(a_{i_1}, \ldots, a_{i_n}) - k + 1 & \text{if } i_1 < \ldots < i_n \in I \text{ are such that} \\ & x_1 \in [a_{i_1}, a_{i_1+1}), \ldots, x_n \in [a_{i_n}, a_{i_n+1}), \\ 0 & \text{if there are no such } i_1, \ldots, i_n. \end{cases}$$

Let H be an infinite homogeneous set for $\check{f}$. Then the set

$$H' := \{a_i : i \in I \text{ and } H \cap [a_i, a_{i+1}) \neq \emptyset\}$$

exists by Δ^0_1-comprehension: it is clearly Σ^0_1-definable, and its complement is the union of $\mathbb{N} \setminus A$ and the Σ^0_1-definable set $\{a_i : \exists a \in A\, (a > a_i \text{ and } H \cap [a_i, a) = \emptyset)\}$. Moreover, H' is infinite and homogeneous for f. □

Definition 5. For $\ell \geq 1, n, k \geq 2$, let $\Delta_\ell\text{-}\mathrm{RT}^n_k$ be the first-order statement: "for every Δ_ℓ-definable k-colouring of $[\mathbb{N}]^n$, there is a Δ_ℓ-definable infinite homogeneous set".

Thus, a model M satisfies $\Delta_\ell\text{-}\mathrm{RT}^n_k$ exactly if $(M, \Delta_\ell\text{-}\mathrm{Def}(M)) \models \mathrm{RT}^n_k$.

It is well known that each $\Delta_\ell\text{-}\mathrm{RT}^n_k$ is false in the standard model. However, the usual argument makes use of a nontrivial amount of induction.

Lemma 6. *For each* $n \geq 2$*:*

(a) $\mathrm{I}\Sigma_1$ *proves that there is a* Δ_1*-definable* 2*-colouring of* $[\mathbb{N}]^n$ *with no* Σ_1*-definable infinite homogeneous set,*

(b) for each $l \geq 1$, $\mathrm{I}\Sigma_{\ell+1}$ proves that there is a Δ_ℓ-definable 2-colouring of $[\mathbb{N}]^n$ with no $\Sigma_{\ell+1}$-definable infinite homogeneous set.

Proof. Clearly, it is enough to prove the statement for $n = 2$.

The proof of (b) is just a formalization of the usual proof due to [15] in $\mathrm{I}\Sigma_{\ell+1}$. The place where $\Sigma_{\ell+1}$-induction is used is when we are given a hypothetical $\Delta_{\ell+1}$-definable infinite homogeneous set with code e, and we want to reach a contradiction by looking at the first $2e + 2$ elements of this set. To do this, we need to know that the set actually has at least $2e + 2$ elements, and this is justified by proving "for every x, the $\Delta_{\ell+1}$-set with code e has a finite subset with at least x elements" by induction on x.

To prove (a), one could formalize Specker's construction [32] of a computable 2-colouring of pairs with no r.e. homogeneous set within $\mathrm{I}\Sigma_1$. Instead of that, we choose to formalize a weaker variant of the argument of [15] proving (b) for $\ell = 1$. We define a computable function $f\colon [\mathbb{N}]^2 \to 2$ in the following way. At stage s, we determine the values $f(n, s)$ for $n < s$. To do this, we consider all Σ_1 formulas with codes $0, \ldots, \lfloor (s-1)/2 \rfloor$. Given $e \leq \lfloor (s-1)/2 \rfloor$, if e is the code of a Σ_1 formula $\exists v\, \delta(x, v)$ and there are at least $2e+2$ elements $x < s$ such that $\exists v \leq s\, \mathrm{Sat}_0(\ulcorner \delta \urcorner, (x, v))$ holds, then choose the smallest two such elements x_0, x_1 for which $f(x_0, s), f(x_1, s)$ have not yet been defined, and let $f(x_i, s) = i$. Otherwise, do nothing. Once all the formulas with codes $0, \ldots, \lfloor (s-1)/2 \rfloor$ have been dealt with, complete stage s by letting $f(x, s) = 0$ for all those $x < s$ for which $f(x, s)$ was not defined earlier.

Now if the formula $\exists v\, \delta(x, v)$ with code e defines an infinite homogeneous set for f, we can use Σ_1 induction to conclude that there are at least $2e + 2$ elements x such that $\exists v\, \delta(x, v)$ holds. Consider the $2e+2$ smallest such elements, say $x_0 < \ldots < x_{2e+1}$. By another application of Σ_1 induction, there is some $s > \max(2e, x_{2e+1})$ such that for $x \leq x_{2e+1}$, if $\exists v\, \delta(x, v)$, then $\exists v \leq s\, \delta(x, v)$. Since there are infinitely many elements x such that $\exists v\, \delta(x, v)$, we can also assume that $\exists v\, \delta(s, v)$. But the lower bounds on s imply that at stage s there will be some $i < j \leq 2e + 1$ such that $\exists v\, \delta(x_i, v), \exists v\, \delta(x_j, v)$, and $f(x_i, s) \neq f(x_j, s)$. This is a contradiction, because all three elements x, x', s satisfy a formula that defines a homogeneous set for f. □

Lemma 7. *Let $(M, \mathcal{X}) \models \mathrm{RCA}_0^* + \mathrm{RT}_2^n$ where $n \geq 3$ and assume that $M \models \mathrm{I}\Sigma_\ell$. Then $0^{(\ell)} \in \mathcal{X}$. As a consequence, $\Delta_{\ell+1}$-$\mathrm{Def}(M) \subseteq \mathcal{X}$ and $M \models \mathrm{B}\Sigma_{\ell+1}$.*

Proof. Let $M \models \mathrm{RCA}_0^* + \mathrm{RT}_2^n + \mathrm{I}\Sigma_\ell$. We will prove by induction on $j \leq \ell$ that $0^{(j)} \in \mathcal{X}$. For $j = \ell$, this will immediately imply $\Delta_{\ell+1}$-$\mathrm{Def}(M) \subseteq \mathcal{X}$ and $M \models \mathrm{B}\Sigma_{\ell+1}$ because $(M, \mathcal{X})$ satisfies Δ_1^0 comprehension and $\mathrm{B}\Sigma_1^0$.

The base step of the induction holds by Δ_1^0-comprehension in $(M, \mathcal{X})$. So, let $j < \ell$ and assume that $0^{(j)} \in \mathcal{X}$. We have to prove that $0^{(j+1)} \in \mathcal{X}$.

Consider the usual computable instance of RT_2^3 whose solutions compute $0'$ and relativize it to $0^{(j)}$:

$$f(x, y, z) = \begin{cases} 0 & \text{if there is a } \Sigma_{j+1} \text{ sentence } \exists v\, \pi(v) \text{ with code at most } x \\ & \text{such that } \forall v \leq y\, \mathrm{Sat}_j(\ulcorner \neg\pi \urcorner, v) \wedge \exists v \leq z\, \neg \mathrm{Sat}_j(\ulcorner \neg\pi \urcorner, v), \\ 1 & \text{otherwise.} \end{cases}$$

The colouring f is $\Delta_1(0^{(j)})$-definable, so $f \in \mathcal{X}$. By RT_2^n, there exists an infinite $H \in \mathcal{X}$ homogeneous for f. We claim that H cannot be 0-homogeneous for f.

To see this, note that by $\mathrm{I}\Sigma_\ell$ we have strong Σ_{j+1} collection, so for any given x there is a bound w such that for any Σ_{i+1} sentence with code below x, if the sentence is true, then there is a witness for it below w. Thus, for any $z > y \geq w$, we must have $f(x, y, z) = 1$, which implies that no infinite set can be 0-homogeneous for f.

So, H is 1-homogeneous for f. We can now compute $0^{(j+1)}$ with oracle access to $0^{(i)} \oplus H$ as follows: given a Σ_{j+1} sentence $\exists v\, \pi(v)$, find some $x \in H$ above the code for the sentence, find $y \in H$ above x, and use $0^{(j)}$ to determine whether $\exists v \leq y\, \pi(v)$ holds; if is does not, then neither does $\exists v\, \pi(v)$. Both $0^{(j)}$ and H are in $\mathcal{X}$, so $0^{(j+1)} \in \mathcal{X}$ as well. □

We are now ready to give an axiomatization of the first-order part of $\mathrm{RCA}_0^* + \mathrm{RT}_2^n$ for $n \geq 3$. Afterwards, we will study the relationship of this theory to the usual fragments of first-order arithmetic.

Theorem 8. *Let $n \geq 3$ and let R^n be the theory:*

$$\left\{ (\mathrm{B}\Sigma_{\ell+1} \wedge \exp) \vee \bigvee_{j=1}^{\ell} \Delta_j\text{-}\mathrm{RT}_2^n : \ell \in \omega \right\}. \tag{2}$$

Then R^n axiomatizes the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}_2^n$.

Proof. Fix $n \geq 3$ and let R^n be as in (2).

We first argue that for every $M \models \mathrm{R}^n$ there is a family of sets $\mathcal{X} \subseteq \mathcal{P}(M)$ such that $(M, \mathcal{X}) \models \mathrm{RCA}_0^* + \mathrm{RT}_2^n$, which will mean that $\mathrm{R}^n \not\models \psi$ implies $\mathrm{RCA}_0^* + \mathrm{RT}_2^n \not\models \psi$ for each arithmetical sentence ψ. So, let $M \models \mathrm{R}^n$. If $M \models \mathrm{PA}$, then $(M, \mathrm{Def}(M))$ is a model of ACA_0 and, *a fortiori*, of $\mathrm{RCA}_0^* + \mathrm{RT}_2^n$.

Otherwise, let $\ell \in \omega$ be the smallest such that $M \models \neg \mathrm{I}\Sigma_{\ell+1}$. For each $j = 1, \ldots, \ell$, it follows from Lemma 6 that there is a Δ_j-definable 2-colouring of $[M]^n$ with no Δ_j-definable homogeneous set, so R^n implies that $\mathrm{B}\Sigma_{\ell+1} + \exp$ must hold in M. Moreover, since $\mathrm{B}\Sigma_{\ell+2}$ fails, it must be the case that $M \models \Delta_{\ell+1}\text{-}\mathrm{RT}_2^n$. Thus $(M, \Delta_{\ell+1}\text{-}\mathrm{Def}(M)) \models \mathrm{RCA}_0^* + \mathrm{RT}_2^n$.

In the other direction, we assume that $(M, \mathcal{X}) \models \mathrm{RCA}_0^* + \mathrm{RT}_2^n$ and prove that $M \models \mathrm{R}^n$. This is clear if $M \models \mathrm{PA}$. Otherwise, let ℓ be such that $M \models \neg \mathrm{B}\Sigma_{\ell+1}$. Let $j \leq \ell$ be the largest such that $M \models \mathrm{I}\Sigma_j$. By Lemma 7, $M \models \mathrm{B}\Sigma_{j+1}$, so in particular $j < \ell$. Moreover, $\Delta_{j+1}\text{-}\mathrm{Def}(M) \subseteq \mathcal{X}$. We now argue that $(M, \Delta_{j+1}\text{-}\mathrm{Def}(M)) \models \mathrm{RT}_2^n$, which will complete the argument.

Let I be a Σ_{j+1}-definable proper cut in M. The cut I is Σ_1^0-definable in $(M, \Delta_{j+1}\text{-}\mathrm{Def}(M))$ and thus also in $(M, \mathcal{X})$. Moreover, both of these structures satisfy RCA_0^*. Therefore, Theorem 3 and the fact that $(M, \mathcal{X}) \models \mathrm{RCA}_0^* + \mathrm{RT}_2^n$ let us conclude that $(M, \Delta_{j+1}\text{-}\mathrm{Def}(M)) \models \mathrm{RT}_2^n$ as well. □

Definition 9. The theory IB is axiomatized by $\mathrm{B}\Sigma_1$ and the set of sentences

$$\{\mathrm{I}\Sigma_\ell \Rightarrow \mathrm{B}\Sigma_{\ell+1} : \ell \geq 1\}.$$

Kaye [18] showed that $\mathrm{IB} + \exp$ implies the theory of all κ-like models of arithmetic (for κ possibly singular). It is now known (see [12, Section 3.3], [2, Section 6]) that $\mathrm{IB} + \exp$ is actually strictly stronger than the theory of all κ-like models.

Theorem 10. *Let $n \geq 3$. Then:*

(a) the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}^n_2$ *are strictly in between* $\mathrm{IB} + \exp$ *and* PA*; as a result, they are not finitely axiomatizable.*

(b) the Π_3 *consequences of* $\mathrm{RCA}_0^* + \mathrm{RT}^n_2$ *coincide with* $\mathrm{B}\Sigma_1 + \exp$*; for* $\ell \geq 1$*, the* $\Pi_{\ell+3}$ *consequences are strictly in between*

$$\mathrm{B}\Sigma_1 + \exp + \bigwedge_{1 \leq j \leq \ell} (\mathrm{I}\Sigma_j \Rightarrow \mathrm{B}\Sigma_{j+1})$$

and $\mathrm{B}\Sigma_{\ell+1}$.

Proof. We first prove (b). As in Theorem 8, we let R^n stand for the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}^n_2$.

It follows immediately from the definition of RCA_0^* and Lemma 7 that the $\Pi_{\ell+3}$ consequences of R^n include $\mathrm{B}\Sigma_1 + \exp$ and $\mathrm{I}\Sigma_\ell \Rightarrow \mathrm{B}\Sigma_{\ell+1}$ for each $j \leq \ell$. For $\ell \geq 1$, the inclusion is strict, because the statement

$$(\mathrm{B}\Sigma_{\ell+1} \wedge \exp) \vee \bigvee_{j=1}^{\ell} \Delta_j\text{-}\mathrm{RT}^n_2$$

is $\Pi_{\ell+3}$ but not provable in $\mathrm{B}\Sigma_1 + \exp + \bigwedge_{1 \leq j \leq \ell}(\mathrm{I}\Sigma_j \Rightarrow \mathrm{B}\Sigma_{j+1})$. To see the unprovability, consider a model $M \models \mathrm{B}\Sigma_\ell + \exp$ such that ω is Σ_ℓ-definable in M and $(\omega, \mathrm{SSy}(M)) \not\models \mathrm{RT}^n_2$. Then, clearly, $M \models \mathrm{I}\Sigma_j \Rightarrow \mathrm{B}\Sigma_{j+1}$ for each $j \leq \ell$; in fact, M is a model of IB. However, Lemma 6 implies that $(M, \Delta_j\text{-}\mathrm{Def}(M)) \not\models \mathrm{RT}^n_2$ for each $1 \leq j \leq \ell - 1$. On the other hand, $(M, \Delta_\ell\text{-}\mathrm{Def}(M))$ is a model of RCA_0^* in which ω is Σ^0_1-definable, so by Theorem 3 and the choice of $\mathrm{SSy}(M)$ it does not satisfy RT^n_2 either.

Using a model M chosen similarly but with $(\omega, \mathrm{SSy}(M)) \models \mathrm{RT}^n_2$, we get $(M, \Delta_\ell\text{-}\mathrm{Def}(M)) \models \mathrm{RT}^n_2 + \neg\mathrm{B}\Sigma_{\ell+1}$. Thus, R^n does not prove $\mathrm{B}\Sigma_{\ell+1}$ for $\ell \geq 1$.

To see that all $\Pi_{\ell+3}$ consequences of R^n follow from $\mathrm{B}\Sigma_{\ell+1}$ for $\ell \geq 1$ let the $\Sigma_{\ell+3}$ formula $\psi := \exists x\, \forall y\, \exists z\, \pi(x, y, z)$ be consistent with $\mathrm{B}\Sigma_{\ell+1}$, let $K \models \mathrm{B}\Sigma_{\ell+1} \wedge \psi$ be such that $(\omega, \mathrm{SSy}(K)) \models \mathrm{RT}^n_2$, and let $a \in K$ be a witness for the initial existential quantifier in ψ. By $\mathrm{B}\Sigma_{\ell+1}$, the function

$$\begin{aligned} f(y) = {} & \text{least } w > y \text{ such that } \forall y' \!\leq y\, \exists z \!\leq w\, \pi(a, y', z) \\ & \quad \text{and ``true } \Sigma_\ell \text{ sentences with codes} \leq y \text{ are witnessed} \leq w\text{''} \end{aligned}$$

is total and $\Delta_{\ell+1}$-definable in K. Let M be the cut $\sup_K(\{f^m(a) : m \in \omega\})$. Then $M \models \mathrm{B}\Sigma_{\ell+1} \wedge \psi$ and ω is $\Sigma_{\ell+1}$-definable in M. Since $(\omega, \mathrm{SSy}(M)) \models \mathrm{RT}^n_2$, we get $(M, \Delta_\ell\text{-}\mathrm{Def}(M)) \models \mathrm{RT}^n_2$ by Theorem 3, so $M \models \mathrm{R}^n \wedge \psi$.

The proof that the Π_3 consequences of R^n follow from $\mathrm{B}\Sigma_1 + \exp$ is very similar, except that the function f is now defined by

$$f(y) = \text{least } w > 2^y \text{ such that } \forall y' \!\leq y\, \exists z \!\leq w\, \pi(a, y', z),$$

where π is now a Δ_0 formula. The difference is due to the fact that for $\ell = 0$ we no longer have to care about elementarity between the cut M and the model K to ensure that $M \models \mathrm{B}\Sigma_{\ell+1} \wedge \psi$, but we need to guarantee that $M \models \exp$.

We have thus proved (b). Regarding (a), note that the containments

$$\mathrm{IB} + \exp \subseteq \mathrm{R}^n \subsetneq \mathrm{PA}$$

follow directly from the statement of (b), and in the proof of (b) we constructed a model of $\mathrm{IB} + \exp$ not satisfying R^n. Finally, observe that IB is not contained in any $\mathrm{I}\Sigma_\ell$, so any subtheory of PA extending IB cannot be finitely axiomatizable. □

Note that the proof of Theorem 10 immediately gives the following statement, which says essentially that Lemma 6 is optimal with respect to the amount of induction used to prove the existence of colourings without simple homogeneous sets.

Corollary 11. *For each* $\ell \geq 1, n \geq 2$*, the theory* $\mathrm{B}\Sigma_\ell + \exp + \Delta_\ell\text{-}\mathrm{RT}^n_2$ *is consistent.*

Remark. As mentioned in Section 1, results such as Theorem 10 can be converted from purely arithmetical to Π^1_1 form by relativizing to second-order parameters. In Theorem 10(a), the appropriate relativization of the scheme IB takes the form $\forall X\, (\mathrm{I}\Sigma_k(X) \Rightarrow \mathrm{B}\Sigma_{k+1}(X))$ for each k. In Section 4, we will also consider a weaker relativization of IB: see the remark after Corollary 15.

Question 1. Does $\mathrm{RCA}^*_0 + \mathrm{RT}^3_2$ imply RT^4_2? More generally, does $\mathrm{RCA}^*_0 + \mathrm{RT}^n_2$ imply RT^{n+1}_2 for some/all $n \geq 3$?

4 Ramsey for pairs

We turn to the case of Ramsey's Theorem for pairs. Here, we are not able to give a complete axiomatization analogous to that of Theorem 8. Loosely speaking, our understanding of the strength of $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$ strongly depends on the amount of induction satisfied by the underlying first-order model.

Theorem 12. *Let* R^2 *stand for the first-order consequences of* $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$*. Then:*

- *(a)* $\mathrm{R}^2 \wedge \neg\mathrm{I}\Sigma_1$ *is axiomatized by* $\mathrm{B}\Sigma_1 + \exp + \Delta_1\text{-}\mathrm{RT}^2_2$*.*
- *(b)* $\mathrm{I}\Sigma_2$ *implies* R^2*.*
- *(c) Over* $\mathrm{B}\Sigma_2$*,* R^2 *is implied by, and consistent with, both the first-order consequences of* $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ *and the statement* $\Delta_2\text{-}\mathrm{RT}^2_2$*.*
- *(d)* R^2 *implies every first-order sentence* ψ *such that both* $\mathrm{B}\Sigma_2 \vdash \psi$ *and* $\mathrm{RCA}^*_0 + \neg\mathrm{I}\Sigma^0_1 \vdash \psi$*.*

Proof. We first prove (a). Clearly, if $M \models \mathrm{B}\Sigma_1 + \exp$ and $(M, \Delta_1\text{-}\mathrm{Def}(M)) \models \mathrm{RT}^2_2$, then M satisfies R^2 (as well as $\neg\mathrm{I}\Sigma_1$, by Lemma 6). On the other hand, let $(M, \mathcal{X}) \models \mathrm{RCA}^*_0 + \mathrm{RT}^2_2 + \neg\mathrm{I}\Sigma_1$. Obviously, M satisfies $\mathrm{B}\Sigma_1 + \exp$. Let I be a proper Σ_1-definable cut in M. Applying Theorem 3 two times, we get first $(I, \mathrm{Cod}(M/I)) \models \mathrm{RT}^2_2$ and then $(M, \Delta_1\text{-}\mathrm{Def}(M)) \models \mathrm{RT}^2_2$.

Statement (b) follows immediately from the result of [4] that $\mathrm{RCA}_0 + \mathrm{I}\Sigma^0_2 + \mathrm{RT}^2_2$ is conservative over $\mathrm{I}\Sigma_2$.

We turn to (c). It is clear that R^2 is implied by the first-order consequences of $\mathrm{RCA}_0 + \mathrm{RT}^2_2$. Meanwhile, R^2 is also satisfied by any model $M \models \mathrm{B}\Sigma_2 + \Delta_2\text{-}\mathrm{RT}^2_2$ since $(M, \Delta_2\text{-}\mathrm{Def}(M)) \models \mathrm{RCA}^*_0 + \mathrm{RT}^2_2$. It remains to argue that such a model exists. To see this, take $M \models \mathrm{B}\Sigma_2$ with Σ_2-definable ω and $(\omega, \mathrm{SSy}(M)) \models \mathrm{RT}^2_2$, and apply Theorem 3 to the model $(M, \Delta_2\text{-Def}(M)) \models \mathrm{RCA}^*_0$.

Finally, to see that (d) holds, let ψ be provable both in $\mathrm{B}\Sigma_2$ and in $\mathrm{RCA}^*_0 + \neg\mathrm{I}\Sigma^0_1$. We check that $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2 \vdash \psi$. Let $(M, \mathcal{X}) \models \mathrm{RCA}^*_0 + \mathrm{RT}^2_2$. If $(M, \mathcal{X}) \models \mathrm{RCA}_0$, then $M \models \mathrm{B}\Sigma_2$, so $M \models \psi$. Otherwise, $M \models \mathrm{RCA}^*_0 + \neg\mathrm{I}\Sigma^0_1$, so $M \models \psi$ as well. $\square$

Parts (a) and (b) of Theorem 12 give a complete axiomatization of the first-order consequences of $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$ over, respectively, $\neg\mathrm{I}\Sigma_1$ and $\mathrm{I}\Sigma_2$. However, the situation in the region between $\mathrm{I}\Sigma_1$ and $\mathrm{I}\Sigma_2$ is much less clear.

As mentioned in the introduction, it is open whether $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ is arithmetically conservative over $\mathrm{B}\Sigma_2$. Therefore, it is consistent with what we know that already $\mathrm{B}\Sigma_2$ implies the first-order consequences of $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$.

On the other hand, we will now use Theorem 12(d) to show that there are some first-order sentences provable in $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$ but not in $\mathrm{I}\Sigma_1$. It will be clear from our argument that this is not a feature of RT^2_2 specifically, but rather of all principles that imply $\mathrm{B}\Sigma^0_2$ (or even somewhat weaker statements) over RCA_0.

Definition 13. For each $\ell \geq 1$, the Σ_ℓ *cardinality scheme*, $\mathrm{C}\Sigma_\ell$, asserts that no Σ_ℓ formula defines a total injection with bounded range.

The Σ_ℓ *generalized pigeonhole principle*, $\mathrm{GPHP}(\Sigma_\ell)$, asserts that for every Σ_ℓ formula $\varphi(x, y, z)$ and every number a, there exists a number b such that there is no c for which $\varphi(\cdot, \cdot, c)$ defines an injective multifunction from b into a:

$$\forall a\, \exists b\, \forall c\, [\forall x\!<\!b\, \exists y\!<\!a\, \varphi(x, y, c) \Rightarrow \neg\forall y\!<\!a\, \exists^{\leq 1} x\!<\!b\, \varphi(x, y, c)].$$

The principle $\mathrm{C}\Sigma_\ell$ was defined in [28]. It is known that $\mathrm{I}\Sigma_\ell$ does not imply $\mathrm{C}\Sigma_{\ell+1}$ [11, Proposition 3.1]. The principle $\mathrm{GPHP}(\Sigma_\ell)$ was defined in [18], where it was also observed that the theory of all κ-like models of $\mathrm{I}\Delta_0$ implies $\mathrm{GPHP}(\Sigma_\ell)$ for all ℓ.

Clearly, $\mathrm{GPHP}(\Sigma_\ell)$ implies $\mathrm{C}\Sigma_\ell$ for each $\ell \geq 1$. For $\ell \geq 2$, $\mathrm{GPHP}(\Sigma_\ell)$ is in turn implied by $\mathrm{B}\Sigma_\ell$, since the latter is, for each $\ell \geq 1$, equivalent to the usual pigeonhole principle for Σ_ℓ maps over $\mathrm{I}\Delta_0 + \exp$ [9]. It follows from [2] that the implication from $\mathrm{B}\Sigma_\ell$ to $\mathrm{GPHP}(\Sigma_\ell)$ is strict.

$\mathrm{C}\Sigma_2$ is known to be a consequence of some theories studied in reverse mathematics that do not imply $\mathrm{B}\Sigma_2$, such as RCA_0 plus the Rainbow Ramsey Theorem for pairs [8] and RCA_0 plus the existence of 2-random reals [12].

In the theorem below, we explicitly indicate second-order variables to emphasize the role played by set parameters in the second part of the statement. Recall that $\mathrm{I}\Sigma^0_k$ (resp. $\mathrm{B}\Sigma^0_k$) means $\forall X\, \mathrm{I}\Sigma_k(X)$ (resp. $\forall X\, \mathrm{B}\Sigma_k(X)$).

Theorem 14. *For each $k, \ell \geq 1$, the following statements are provable in* RCA^*_0*:*

(a) $\forall X\, (\mathrm{B}\Sigma_\ell(X) \Rightarrow \mathrm{GPHP}(\Sigma_\ell(X)))$*,*

(b) $(\mathrm{B}\Sigma^0_k \wedge \neg\mathrm{I}\Sigma^0_k) \Rightarrow \forall X\, \mathrm{GPHP}(\Sigma_\ell(X))$*.*

Theorem 14 part (b) can be obtained by relativizing Kaye's proof of the result that any model of $\mathrm{B}\Sigma_1 + \exp + \neg\mathrm{I}\Sigma_1$ is elementarily equivalent to an $\aleph_\omega$-like structure [18, Theorem 2.4]. A model of $\neg\mathrm{I}\Sigma_1(A) + \neg\mathrm{GPHP}(\Sigma_\ell(B)) + \mathrm{B}\Sigma_1(A \oplus B) + \exp$ would also be elementary equivalent to $\aleph_\omega$-like model, but clearly such a structure can never violate the scheme GPHP(Γ) for any class of formulas Γ.

The proof of Theorem 14 we give below is considerably simpler than that of [18, Theorem 2.4]. On the other hand, both make use of an automorphism argument. It would be interesting to come up with a direct proof of GPHP(Σ_ℓ), with no model-theoretic detours, in for instance $\mathrm{B}\Sigma_1 + \exp + \neg\mathrm{I}\Sigma_1$, .

Proof. It has already been mentioned that $\mathrm{B}\Sigma_\ell + \exp$ implies GPHP(Σ_ℓ). The argument for this relativizes with no issues, thus proving part (a).

It remains to prove that $\mathrm{RCA}_0^* + \mathrm{B}\Sigma_k^0 + \neg\mathrm{I}\Sigma_k^0$ implies GPHP(Σ_ℓ^0) for any ℓ. To simplify notation, we restrict ourselves to the case where $k = 1$ and to GPHP for lightface Σ_ℓ formulas. The general case for $k \geq 1$ and a $\Sigma_\ell(B)$ formula reduces to this one by considering the model of RCA_0^* given by the $\Delta_k(A \oplus B)$-definable sets, where A is a parameter witnessing the failure of $\mathrm{I}\Sigma_k^0$.

Let (M, A) be a countable model of $\mathrm{B}\Sigma_1(A) + \exp + \neg\mathrm{I}\Sigma_1(A)$. We may assume that A itself has an increasing enumeration $A = \{a_i : i \in I\}$ for a proper cut $I \subseteq M$. By a routine compactness argument, we may also assume that for every $a \in M$ there is some $b \in M$ such that $b > \exp_m(a)$ for each $m \in \omega$. To prove that $M \models$ GPHP(Σ_ℓ), we will use a technique based on the fact that models of $\mathrm{B}\Sigma_1^0 + \exp + \neg\mathrm{I}\Sigma_1^0$ have many automorphisms [22, 23, 16].

By a standard argument (see e.g. [10, Theorem 4.6]), the model M can be end-extended to a model $K \models \mathrm{I}\Delta_0$ such that $A \in \mathrm{Cod}(K/M)$. Since elements coding A are downwards cofinal in $K \setminus M$, there is an element $d \in K$ coding A and small enough that $\exp_2(d)$ exists in K. By [25], there is a Δ_0 formula with parameter $\exp_2(d)$ that defines satisfaction for Δ_0 formulas on arguments below d. As a consequence, the structure $[0, d]$ (with addition and multiplication as ternary relations) is recursively saturated.

Now let $a \in M \setminus I$ and let $b \in M$ be such that $b > \exp_m(a)$ for each $m \in \omega$. Let $c \in M$ be arbitrary. The recursive saturation of $[0, d]$ lets us use an argument dating back to [24] (see the proof of Lemma 3.4 in [16] for a detailed argument and [23] for a brief discussion) to derive the existence of an automorphism α of $[0, d]$ such that α fixes c, d and fixes $[0, a]$ pointwise, but there is some $x < b$ with $x \neq \alpha(x) =: y$. For each $i \in I$, since $\alpha(i) = i$, $\alpha(d) = d$, and d codes A, we know that $\alpha(a_i) = a_i$. Therefore, $\alpha[M] = M$, so $\alpha{\restriction}_M$ is actually an automorphism of M. We now argue that no injective multifunction from b to a is definable in M with c as parameter. Otherwise, if f were such a multifunction, there would be some $z < a$ such that $z \in f(x)$, and therefore (since α fixes both z and c) also $z = \alpha(z) \in f(\alpha(x)) = f(y)$. By the injectivity of f, this would imply $x = y$, a contradiction. Since $c \in M$ was arbitrary, this proves that there can be no injective multifunction from b to a definable in M, so $M \models$ GPHP(Σ_ℓ) for each ℓ. □

Corollary 15. $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ *proves both* $\mathrm{C}\Sigma_2$ *and* GPHP(Σ_2).

Proof. This is a direct consequence of Theorem 12(d), Theorem 14(b), and the fact that GPHP(Σ_ℓ) implies $\mathrm{C}\Sigma_\ell$. □

Remark. Let the usual relativization of IB, namely $\forall X\, (\mathrm{I}\Sigma_k(X) \Rightarrow \mathrm{B}\Sigma_{k+1}(X))$ for each k, be called "strong", and let the "weak" relativization of IB consist of the statements $\mathrm{I}\Sigma_k^0 \Rightarrow \mathrm{B}\Sigma_{k+1}^0$ for each k. In Theorem 10, we showed that $\mathrm{RCA}_0^* + \mathrm{RT}_2^3$ implies strong relativized IB. On the other hand, Theorem 14 implies that already weak relativized IB, and even its restriction to $k < \ell$, suffices to prove $\mathrm{GPHP}(\Sigma_\ell)$.

This lets us prove Corollary 15 by exploiting the fact that $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ implies the restriction of weak relativized IB to $k = 0, 1$.

The known relationships between the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ and fragments of first-order arithmetic are summarized in the following corollary.

Corollary 16. *The first-order consequences of* $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ *follow from* $\mathrm{I}\Sigma_2$. *The* Π_3 *consequences coincide with* $\mathrm{B}\Sigma_1 + \exp$. *The* Π_4 *consequences are strictly weaker than* $\mathrm{B}\Sigma_2$ *but do not follow from* $\mathrm{I}\Sigma_1$.

Proof. The provability from $\mathrm{I}\Sigma_2$ is part (b) of Theorem 12. The fact that the Π_3 consequences of $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ coincide with $\mathrm{B}\Sigma_1 + \exp$ and that the Π_4 consequences are strictly weaker than $\mathrm{B}\Sigma_2$ is proved like in Theorem 10. Finally, Corollary 15 implies that $\mathrm{C}\Sigma_2$ is an example of a Π_4 sentence that follows from $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ but not $\mathrm{I}\Sigma_1$. □

Of course, quite a few questions remain. Over $\mathrm{B}\Sigma_2$, one basic issue is whether the first-order consequences of $\mathrm{RCA}_0^* + \mathrm{RT}_2^2 + \mathrm{B}\Sigma_2$ are non-trivial, and another is how closely related they are to those of $\mathrm{RCA}_0 + \mathrm{RT}_2^2$.

Question 2. Is $\mathrm{RCA}_0^* + \mathrm{RT}_2^2 + \mathrm{B}\Sigma_2$ conservative over $\mathrm{B}\Sigma_2$?

Question 3. Does $\mathrm{RCA}_0^* + \mathrm{RT}_2^2 + \mathrm{B}\Sigma_2$ imply $\psi \vee (\mathrm{B}\Sigma_2 \wedge \Delta_2\text{-}\mathrm{RT}_2^2)$ for each first-order ψ provable in $\mathrm{RCA}_0 + \mathrm{RT}_2^2$?

Over $\mathrm{I}\Sigma_1$, the basic question is:

Question 4. Does $\mathrm{RCA}_0^* + \mathrm{RT}_2^2 + \mathrm{I}\Sigma_1$ imply $\mathrm{B}\Sigma_2$?

We have no strong reasons to believe that the answer is "yes". However, it should be pointed out that, since $\mathrm{RCA}_0 + \mathrm{RT}_2^2$ proves $\mathrm{B}\Sigma_2$, answering "no" would involve constructing a model of $\mathrm{I}\Sigma_1 + \neg\mathrm{B}\Sigma_2$ that expands to a model of $\mathrm{B}\Sigma_1^0 + \neg\mathrm{I}\Sigma_1^0$ – in the terminology of [22], a model of $\mathrm{I}\Sigma_1 + \neg\mathrm{B}\Sigma_2$ that is not always semiregular. The existence of such a model itself seems to be open.

Question 5. Does there exist a model $M \models \mathrm{I}\Sigma_1 + \neg\mathrm{B}\Sigma_2$ that can be expanded to a model $(M, A) \models \mathrm{B}\Sigma_1(A) + \neg\mathrm{I}\Sigma_1(A)$?

Note that if there is M witnessing a positive answer to this question such that $\mathrm{I}\Sigma_1(A)$ fails in the expansion due to ω being $\Sigma_1(A)$-definable, then by Theorems 3 and 10 it has to be the case that $(\omega, \mathrm{SSy}(M)) \not\models \mathrm{ACA}_0$.

5 Relativizing Ramsey

In this final section, we take up the question whether our results on $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ shed any light on the problem of characterizing the first-order consequences of $\mathrm{RCA}_0 + \mathrm{RT}_2^2$. To this end, we introduce a principle in which both the instances and solutions to Ramsey's Theorem are allowed to be Δ_2^0-sets rather than sets.

Definition 17. $\Delta_2^0\text{-}\mathrm{RT}_2^2$ is the Π_2^1 statement: "for every Δ_2^0-set f which is a 2-colouring of $[\mathbb{N}]^2$, there exists an infinite homogeneous Δ_2^0-set".

Note that $\Delta_2^0\text{-}\mathrm{RT}_2^2$ is a genuine Π_2^1 statement, which should not be confused with the Π_1^1 statement relativizing $\Delta_2\text{-}\mathrm{RT}_2^2$, namely "for every set X, $\Delta_2(X)$-RT_2^2 holds". Of course, in a model of the form $(M, \Delta_1\text{-}\mathrm{Def}(M))$, the statement $\Delta_2^0\text{-}\mathrm{RT}_2^2$ will be equivalent to $\Delta_2\text{-}\mathrm{RT}_2^2$.

We are interested in studying Δ^0_2-RT^2_2 over $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2$, especially in the case where $\mathrm{I}\Sigma^0_2$ fails. The following proposition shows that in such a context, Δ^0_2-RT^2_2 behaves somewhat analogously to RT^2_2 over $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$, so we can investigate it using the methods developed in Sections 2-4.

Lemma 18. *For any model* $(M, \mathcal{X}) \models \mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2$*:* $(M, \mathcal{X}) \models \Delta^0_2$-$\mathrm{RT}^2_2$ *iff* $(M, \Delta^0_2\text{-}\mathrm{Def}(M, \mathcal{X})) \models \mathrm{RCA}^*_0 + \mathrm{RT}^2_2$*. As a consequence:*

(a) if I is a Σ^0_2-definable proper cut in $(M, \mathcal{X})$, then $(M, \mathcal{X}) \models \Delta^0_2$-$\mathrm{RT}^2_2$ iff $(I, \mathrm{Cod}(M/I) \models \mathrm{RT}^2_2$,

(b) the first-order consequences of $\mathrm{RCA}_0 + \mathrm{B}\Sigma_2 + \neg\mathrm{I}\Sigma_2 + \Delta^0_2$-$\mathrm{RT}^2_2$ *are axiomatized by* $\mathrm{B}\Sigma_2 + \Delta_2$-$\mathrm{RT}^2_2$*,*

(c) $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2 + \Delta^0_2$-$\mathrm{RT}^2_2$ *is* Π_4*- but not* Π_5*-conservative over* $\mathrm{B}\Sigma_2$*.*

Proof. The fact that a model $(M, \mathcal{X})$ satisfies $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2 + \Delta^0_2$-$\mathrm{RT}^2_2$ exactly if $(M, \Delta^0_2\text{-}\mathrm{Def}(M, \mathcal{X})) \models \mathrm{RCA}^*_0 + \mathrm{RT}^2_2$ is immediate from the definitions. Thus (a) follows from Theorem 3, because a cut I is Σ^0_2 definable in $(M, \mathcal{X}) \models \mathrm{B}\Sigma^0_2$ exactly if it is Σ^0_1-definable in $(M, \Delta^0_2\text{-}\mathrm{Def}(M, \mathcal{X}))$.

To prove (b), repeat the argument from the proof of Theorem 12(a), relativizing it to $0'$. If $M \models \mathrm{B}\Sigma_2 + \Delta_2$-$\mathrm{RT}^2_2$, then $(M, \Delta_1\text{-}\mathrm{Def}(M)) \models \mathrm{RCA}_0 + \mathrm{B}\Sigma_2 + \neg\mathrm{I}\Sigma_2 + \Delta^0_2$-$\mathrm{RT}^2_2$. In the other direction, if $(M, \mathcal{X}) \models \mathrm{RCA}_0 + \mathrm{B}\Sigma_2 + \neg\mathrm{I}\Sigma_2 + \Delta^0_2$-$\mathrm{RT}^2_2$ and I is a proper Σ_2-definable cut in M, then two applications of (a) give first $(I, \mathrm{Cod}(M/I) \models \mathrm{RT}^2_2$ and then $(M, \Delta_1\text{-}\mathrm{Def}(M)) \models \Delta^0_2$-$\mathrm{RT}^2_2$, but the latter is equivalent to $M \models \Delta_2$-RT^2_2.

To show that $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2 + \Delta^0_2$-$\mathrm{RT}^2_2$ is Π_4-conservative over $\mathrm{B}\Sigma_2$, relativize to $0'$ the argument used to prove Π_3-conservativity of $\mathrm{RCA}^*_0 + \mathrm{RT}^n_2$ over $\mathrm{B}\Sigma_1 + \exp$ in Theorem 10(b). To show lack of Π_5-conservativity, consider the sentence $\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2$-$\mathrm{RT}^2_2$. This is a Π_5 statement, and it is provable in $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2 + \Delta^0_2$-$\mathrm{RT}^2_2$ by (b). On the other hand, it is not provable in $\mathrm{B}\Sigma_2$, as can be seen by applying (a) to any model $M \models \mathrm{B}\Sigma_2$ with Σ_2-definable ω and $(\omega, \mathrm{SSy}(M)) \not\models \mathrm{RT}^2_2$. This proves (c). □

Since Lemma 18 shows that Δ^0_2-RT^2_2 is not Π_5-conservative over $\mathrm{B}\Sigma_2$, while the conservativity of $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ over $\mathrm{B}\Sigma_2$ is a well-known open problem, it is natural to ask whether RT^2_2 might imply Δ^0_2-RT^2_2, at least in the particularly relevant setting of models of $\mathrm{B}\Sigma^0_2 + \neg\mathrm{I}\Sigma^0_2$.

In Theorem 20 below, we show a negative result: there is no implication in either direction, and the sentence we used to prove lack of Π_5-conservativity of Δ^0_2-RT^2_2 is unprovable in RT^2_2. To prove this, we will have to make use of a connection between properties of infinite Δ_2-sets and the consistency of $\mathrm{I}\Sigma_1$ that may probably be considered folklore, but for which we did not find a suitable reference. So, we state the connection as a separate lemma and sketch its proof in Appendix A.

Lemma 19. *There exists a polynomial p such that $\mathrm{I}\Sigma_1$ proves:*

$$\forall x \ \big[\text{“every infinite } \Delta_2\text{-set contains at least } \exp_{p(x)}(2) \text{ elements”} \Rightarrow \mathrm{Con}_x(\mathrm{I}\Sigma_1)\big],$$

where $\mathrm{Con}_x(T)$ means that there is no inconsistency proof in T containing fewer than x symbols.

It may be worth pointing out that Lemma 19 is a quantitative version of a weakening of the well-known fact that $\mathrm{I}\Sigma_2$ is equivalent to uniform Π_4-reflection for $\mathrm{I}\Delta_0 + \exp$ (see e.g. [1, Theorem 7]). To see this, note that (over $\mathrm{I}\Delta_0 + \exp$ as a base theory) $\mathrm{I}\Sigma_2$ is equivalent to the statement that each infinite Δ_2-set contains arbitrarily large finite sets, while Π_4-reflection for $\mathrm{I}\Delta_0 + \exp$ implies $\mathrm{Con}(\mathrm{I}\Sigma_1)$.

Theorem 20. RT^2_2 *and* Δ^0_2-RT^2_2 *are incomparable over* $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2 + \neg\mathrm{I}\Sigma^0_2$. *Moreover,* $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ *does not prove* $\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2$-$\mathrm{RT}^2_2$.

Proof. The fact that $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2 + \neg\mathrm{I}\Sigma^0_2 + \Delta^0_2$-$\mathrm{RT}^2_2$ does not prove RT^2_2 is witnessed by any structure of the form $(M, \Delta_1\text{-}\mathrm{Def}(M))$, where $M \models \mathrm{B}\Sigma_2$ has Σ_2-definable ω and $(\omega, \mathrm{SSy}(M)) \models \mathrm{RT}^2_2$. By Lemma 18(a), such a structure satisfies Δ^0_2-RT^2_2, but by Lemma 6(a) it cannot satisfy RT^2_2.

In the other direction, such a "quick and dirty" argument does not seem to be currently available: of the known constructions producing models of $\mathrm{RCA}_0 + \mathrm{RT}^2_2 + \neg\mathrm{I}\Sigma^0_2$, that of [6, 7] involves strong constraints on $\mathrm{SSy}(M)$, and that of [26, 21] does not give a Σ^0_2-definable ω. To show that $\mathrm{RCA}_0 + \mathrm{RT}^2_2 + \neg\mathrm{I}\Sigma^0_2$ does not imply Δ^0_2-RT^2_2, it is enough to prove the "Moreover" part of the statement, namely:

$$\mathrm{RCA}_0 + \mathrm{RT}^2_2 \nvdash \neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-}\mathrm{RT}^2_2.$$

This we do by means of a proof speedup argument. By [19, Lemma 3.2], $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$ proves the statement "for every k, if every infinite set contains at least k elements, then every infinite set contains at least 2^k elements". It follows immediately that $\mathrm{B}\Sigma_2 + \Delta_2$-$\mathrm{RT}^2_2$ proves "for every k, if every infinite Δ_2-set contains at least k elements, then every infinite Δ_2-set contains at least 2^k elements". This implies that the definable set

$$\{x : \text{every infinite } \Delta_2\text{-set contains at least } \exp_x(2) \text{ elements }\}$$

is a cut in $\mathrm{B}\Sigma_2 + \Delta_2$-$\mathrm{RT}^2_2$. This in turn implies (cf. [27, Theorem 3.4.1]) that, for each $n \in \omega$, there is a $\mathrm{poly}(n)$-size proof of

$$\text{"every infinite } \Delta_2\text{-set contains at least } \exp_{\exp_n(2)} 2 \text{ elements"}$$

in $\mathrm{B}\Sigma_2 + \Delta_2$-$\mathrm{RT}^2_2$. But by Lemma 19 and the fact that the exponential function dominates every polynomial, $\mathrm{I}\Sigma_1$ proves:

$$\forall x \left[\text{"every infinite } \Delta_2\text{-set contains at least } \exp_{\exp_{x+1}(2)} 2 \text{ elements"} \right.$$
$$\left. \Rightarrow \mathrm{Con}_{\exp_x(2)}(\mathrm{I}\Sigma_1)\right].$$

Thus, for each standard n there is a $\mathrm{poly}(n)$-size proof of $\mathrm{Con}_{\exp_n(2)}(\mathrm{I}\Sigma_1)$ in $\mathrm{B}\Sigma_2 + \Delta_2$-$\mathrm{RT}^2_2$.

Reasoning by cases, we can show that also $\mathrm{B}\Sigma_2 + (\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-}\mathrm{RT}^2_2)$ proves $\mathrm{Con}_{\exp_n(2)}(\mathrm{I}\Sigma_1)$ in $\mathrm{poly}(n)$-size. Indeed, either $\mathrm{I}\Sigma_2$ holds, in which case we simply have $\mathrm{Con}(\mathrm{I}\Sigma_1)$, or $\mathrm{I}\Sigma_2$ fails, in which case we have Δ_2-RT^2_2 and we can use the proof of $\mathrm{Con}_{\exp_n(2)}(\mathrm{I}\Sigma_1)$ mentioned in the previous paragraph.

However, the size of the smallest proof of $\mathrm{Con}_{\exp_n(2)}(\mathrm{I}\Sigma_1)$ in $\mathrm{I}\Sigma_1$ grows nonelementarily in n [27, Theorem 7.2.2], and by [19], $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ has no superpolynomial proof speedup over $\mathrm{I}\Sigma_1$ w.r.t. proofs of Π_3 sentences. Thus, the size of the smallest proof of $\mathrm{Con}_{\exp_n(2)}(\mathrm{I}\Sigma_1)$ in $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ also grows nonelementarily in n. Since $\mathrm{B}\Sigma_2 + (\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-}\mathrm{RT}^2_2)$ is axiomatized by a single sentence, and $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ proves $\mathrm{B}\Sigma_2$, it follows that it cannot prove $\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-}\mathrm{RT}^2_2$. ☐

Thus, the statement $\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-}\mathrm{RT}^2_2$ cannot be used to witness the potential nonconservativity of RT^2_2 over $\mathrm{B}\Sigma_2$. However, our argument for this, in addition to being somewhat roundabout, made use of the fact that $\mathrm{RCA}^*_0 + \mathrm{RT}^2_2$ proves "for every k, if every infinite set contains at least k elements, then every infinite set contains at least 2^k elements", which is shown using exponential lower bounds on finite Ramsey numbers. Thus the argument is no longer applicable to various apparently slight weakenings of $\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-}\mathrm{RT}^2_2$, for instance to statements in which RT^2_2 is replaced by a restriction to colourings for which finite Ramsey numbers are polynomial.

As an illustration, we mention two weakenings of $\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-}\mathrm{RT}^2_2$ whose status is open and seems intriguing.

Question 6. Does $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ prove one of the following the Π_5 statements:

(a) $\neg\mathrm{I}\Sigma_2 \Rightarrow \Delta_2\text{-CAC}$: if $\neg\mathrm{I}\Sigma_2$, then every Δ_2-definable partial order on $[\mathbb{N}]$ contains an infinite Δ_2-definable chain or an infinite Δ_2-definable antichain",

(b) "if $\neg\mathrm{I}\Sigma_2$, then for every Δ_1-definable 2-colouring of $[\mathbb{N}]^n$ there is a Δ_2-definable infinite homogeneous set"?

Does $\mathrm{RCA}_0 + \mathrm{B}\Sigma^0_2$ prove the statement in (b)?

Acknowledgment. We are very grateful to Tin Lok Wong for many discussions and for making some observations that inspired the work presented in this paper. In fact, we felt that Wong's contribution to the paper are such that he should be listed as a coauthor, but he does not.

We are also grateful to Marta Fiori Carones for a thorough reading of the draft and number of useful comments.

References

[1] Lev D. Beklemishev. Reflection schemes and provability algebras in formal arithmetic. *Russian Math. Surveys*, 60(2):197–268, 2005.

[2] David Belanger, Chitat Chong, Wei Wang, Tin Lok Wong, and Yue Yang. Where pigeonhole principles meet König lemmas, 2019. Preprint. Available at `arXiv:1912.03487`.

[3] David R. Belanger. Conservation theorems for the cohesiveness principle. Preprint, 2015.

[4] Peter A. Cholak, Carl G. Jockusch, and Theodore A. Slaman. On the strength of Ramsey's theorem for pairs. *J. Symb. Log.*, 66(1):1–55, 2001.

[5] C. T. Chong and K. J. Mourad. The degree of a Σ_n cut. *Ann. Pure Appl. Logic*, 48(3):227–235, 1990.

[6] C. T. Chong, Theodore A. Slaman, and Yue Yang. The metamathematics of stable Ramsey's theorem for pairs. *J. Amer. Math. Soc.*, 27(3):863–892, 2014.

[7] C. T. Chong, Theodore A. Slaman, and Yue Yang. The inductive strength of Ramsey's Theorem for Pairs. *Adv. Math.*, 308:121–141, 2017.

[8] Chris J. Conidis and Theodore A. Slaman. Random reals, the rainbow Ramsey theorem, and arithmetic conservation. *J. Symb. Log.*, 78(1):195–206, 2013.

[9] C. Dimitracopoulos and J. Paris. The pigeonhole principle and fragments of arithmetic. *Z. Math. Logik Grundlag. Math.*, 32(1):73–80, 1986.

[10] Ali Enayat and Tin Lok Wong. Unifying the model theory of first-order and second-order arithmetic via WKL_0^*. *Ann. Pure Appl. Logic*, 168(6):1247–1283, 2017.

[11] Marcia J. Groszek and Theodore A. Slaman. On Turing reducibility. Preprint, 1994.

[12] Ian Robert Haken. *Randomizing Reals and the First-Order Consequences of Randoms.* PhD thesis, UC Berkeley, 2014.

[13] Kostas Hatzikiriakou. Algebraic disguises of Σ^0_1 induction. *Arch. Math. Logic*, 29(1):47–51, 1989.

[14] Denis R. Hirschfeldt. *Slicing the truth. On the computable and reverse mathematics of combinatorial principles.* World Scientific, 2015.

[15] Carl G. Jockusch. Ramsey's theorem and recursion theory. *J. Symb. Log.*, 37(2):268–280, 1972.

[16] Richard Kaye. Model-theoretic properties characterizing Peano arithmetic. *J. Symb. Log.*, 56(3):949–963, 1991.

[17] Richard Kaye. *Models of Peano Arithmetic.* Oxford University Press, 1991.

[18] Richard Kaye. Constructing κ-like models of arithmetic. *J. London Math. Soc. (2)*, 55(1):1–10, 1997.

[19] Leszek Aleksander Kołodziejczyk, Tin Lok Wong, and Keita Yokoyama. Ramsey's theorem for pairs, collection, and proof size, 2020. Submitted. Available at `arXiv:2005.06854`.

[20] Leszek Aleksander Kołodziejczyk and Keita Yokoyama. Categorical characterizations of the natural numbers require primitive recursion. *Ann. Pure Appl. Logic*, 166(2):219–231, 2015.

[21] Leszek Aleksander Kołodziejczyk and Keita Yokoyama. Some upper bounds on ordinal-valued Ramsey numbers for colourings of pairs. *Selecta Math. (N.S.)*, 26(4):paper No. 56, 18 pages, 2020.

[22] Roman Kossak. On extensions of models of strong fragments of arithmetic. *Proc. Amer. Math. Soc.*, 108(1):223–232, 1990.

[23] Roman Kossak. A correction to: "On extensions of models of strong fragments of arithmetic" [Proc. Amer. Math. Soc. **108** (1990), no. 1, 223–232; MR0984802 (90d:03123)]. *Proc. Amer. Math. Soc.*, 112(3):913–914, 1991.

[24] Henryk Kotlarski. On elementary cuts in recursively saturated models of Peano arithmetic. *Fund. Math.*, 120(3):205–222, 1984.

[25] Hamid Lessan. *Models of arithmetic.* PhD thesis, University of Manchester, 1978.

[26] Ludovic Patey and Keita Yokoyama. The proof-theoretic strength of Ramsey's theorem for pairs and two colors. *Adv. Math.*, 330:1034–1070, 2018.

[27] Pavel Pudlák. The lengths of proofs. In S. R. Buss, editor, *Handbook of Proof Theory*, pages 547–642. Elsevier, 1998.

[28] David Seetapun and Theodore A. Slaman. On the strength of Ramsey's theorem. *Notre Dame J. Form. Log.*, 36(4):570–582, 1995.

[29] Stephen G. Simpson. *Subsystems of Second Order Arithmetic.* Association for Symbolic Logic, 2009.

[30] Stephen G. Simpson and Rick L. Smith. Factorization of polynomials and Σ^0_1 induction. *Ann. Pure Appl. Logic*, 31(2-3):289–306, 1986.

[31] Stephen G. Simpson and Keita Yokoyama. Reverse mathematics and Peano categoricity. *Ann. Pure Appl. Logic*, 164(3):284–293, 2012.

[32] Ernst Specker. Ramsey's theorem does not hold in recursive set theory. In *Logic Colloquium '69 (Proc. Summer School and Colloq., Manchester, 1969)*, pages 439–442. North-Holland, Amsterdam, 1971.

[33] Keita Yokoyama. On the strength of Ramsey's theorem without Σ_1-induction. *MLQ Math. Log. Q.*, 59(1-2):108–111, 2013.

A Proof of Lemma 19

Lemma 19. *There exists a polynomial p such that $\mathrm{I}\Sigma_1$ proves:*

$$\forall x \left[\text{“every infinite } \Delta_2\text{-set contains at least } \exp_{p(x)}(2) \text{ elements”} \Rightarrow \mathrm{Con}_x(\mathrm{I}\Sigma_1)\right],$$

where $\mathrm{Con}_x(T)$ means that there is no inconsistency proof in T containing fewer than x symbols.

Proof. We assume that our proof system is a Tait-style calculus (see e.g. [1, Section 4.1]). Thus, $\wedge, \vee, \neg$ are our only connectives, with $\neg$ allowed to appear explicitly only in front of atoms and negation otherwise defined recursively using the De Morgan laws. The proof lines are *cedents*, or finite sets of formulas interpreted as disjunctions. The logical axioms are cedents of the form $\Gamma, \psi, \neg\psi$ for ψ atomic, as well as analogous cedents corresponding to the equality axioms (the need to allow the arbitrary set of formulas Γ to appear in axioms arises because there is no weakening rule). The most important rules from our perspective are the rules for introducing conjunctions and quantifiers:

$$\frac{\Gamma, \psi_1 \qquad \Gamma, \psi_2}{\Gamma, \psi_1 \wedge \psi_2}\,(\wedge), \qquad \frac{\Gamma, \psi(t)}{\Gamma, \exists w\, \psi}\,(\exists), \qquad \frac{\Gamma, \psi(a)}{\Gamma, \forall w\, \psi}\,(\forall),$$

where in the ($\exists$) rule t must be a term that is substitutable for w in ψ, and in the ($\forall$) rule a must be an eigenvariable, i.e. a free variable that does not appear anywhere in the conclusion of the rule. There are also natural disjunction introduction rules and the cut rule.

We may assume that $\mathrm{I}\Sigma_1$ is axiomatized by finitely many sentences $\gamma_1, \ldots, \gamma_n$, where each γ_i has the form

$$\forall \bar{v}\, \exists x\, \exists \bar{y}\, \forall \bar{z}\, \forall \bar{z}' \\ [x < v_1 \wedge [\neg\delta_i(0, \bar{z}, \bar{v}) \vee \delta_i(v_1, \bar{y}, \bar{v}) \vee (\delta_i(x, \bar{y}, \bar{v}) \wedge \neg\delta_i(x{+}1, \bar{z}', \bar{v}))]],$$

with δ_i bounded. (Using a different finite axiomatization would shorten proofs in $\mathrm{I}\Sigma_1$ by at most a constant additive factor, and using the typical axiomatization of $\mathrm{I}\Sigma_1$ as a scheme would shorten proofs at most polynomially.)

By the cut elimination theorem, which formalizes in (a fragment of) $\mathrm{I}\Sigma_1$, if there is an inconsistency proof from $\mathrm{I}\Sigma_1$ of size at most x, then for some fixed polynomial p there is a cut-free proof of the cedent

$$\neg\gamma_1, \ldots, \neg\gamma_n,$$

of size at most $\exp_{p(x)}(2)$. Working in $\mathrm{I}\Sigma_1$, let k be such that every infinite Δ_2-set contains at least $\exp_{p(k)}(2)$ elements. Let m stand for $\exp_{p(k)}(2)$. We will prove that there is no cut-free proof of $\neg\gamma_1, \ldots, \neg\gamma_n$ of size at most m, which will imply $\mathrm{Con}_k(\mathrm{I}\Sigma_1)$.

Assume to the contrary that there is such a cut-free proof, and let the lines of the proof be $C_1, \ldots, C_\ell$; note that $\ell \leq m$. For each $j = 1, \ldots, \ell$, let the *negations of* the formulas in C_j be $\xi_{j,1}, \ldots, \xi_{j,r_j}$. Note that $r_\ell = n$, each $\xi_{\ell,i}$ is γ_i, and, by the subformula property of cut-free proofs, each $\xi_{j,r}$ is a subformula of one of the ψ_i's. As usual in such a context, we regard $\varphi(t)$ as a subformula of $\mathrm{Q}x\, \varphi$ for Q a quantifier.

Define an infinite sequence of numbers by:

$$\begin{aligned} d_0 &= 0, \\ d_{j+1} &= \text{least } d > d_j \text{ s.t., if } u \text{ is the smallest number s.t. each term} \\ &\quad \text{with} \leq m \text{ symbols evaluated on arguments} \leq d_j \text{ has value} \leq u, \\ &\quad \text{then } d \geq u \text{ and, for each } i = 1, \ldots, n, \text{ each } \bar{v} \text{ and } x: \\ &\quad \max(x, \max(\bar{v})) \leq u \wedge \exists \bar{y}\, \delta_i(x, \bar{y}, \bar{v}) \Rightarrow \exists \bar{y}\, (\max(\bar{y}) \leq d \wedge \delta_i(x, \bar{y}, \bar{v})). \end{aligned}$$

Let D consist of all numbers that appear as some d_j. Note that provably in $\mathrm{I}\Sigma_1$, both D and the complement of D are Σ_2-definable, so D is a Δ_2-set, and D is infinite. By our assumption, there exists an ℓ-element finite subset of D. W.l.o.g., we may assume that the elements of this subset are $d_0, \ldots, d_{\ell-1}$.

We claim that the following statement $\eta(s)$ can be proved by Π_1 induction on $s = 0, \ldots, \ell{-}1$:

> "there exist $j \leq \ell - s$ and an assignment α of values $\leq d_s$ to the free variables in C_j such that, for every $\xi_{j,r}$ that is Σ_2, there is an assignment of values $\leq d_s$ to the variables (if any) in the unbounded existential quantifier block of $\xi_{j,r}$ that together with α makes the Π_1 part of $\xi_{j,r}$ satisfied."

Note that $\eta(s)$ is indeed a Π_1 statement (provably in $\mathrm{B}\Sigma_1 + \exp$), because all the quantifiers preceding the definition of satisfaction for Π_1 formulas are bounded. Moreover, $\eta(0)$ is true, because it is witnessed by $j = \ell$ and the empty assignment, while $\eta(\ell - 1)$ is false, because C_1 has to be a logical axiom, so an assignment witnessing the statement at $j = 1$ would have to satisfy two mutually contradictory quantifier-free formulas or falsify an equality axiom. Therefore, if the induction step goes through for $\eta(s)$, we obtain the required contradiction.

The induction step splits into cases depending on the rule used to derive C_j, where j witnesses $\eta(s)$. We consider the nontrivial cases, namely the ones corresponding to $(\wedge)$, $(\exists)$, and $(\forall)$ inferences.

If C_j was derived using the $(\wedge)$ rule, then C_j is $\Gamma, \psi_1 \wedge \psi_2$, where $\psi_1 \wedge \psi_2$ is the (necessarily Δ_0) principal formula of the inference used to derive C_j. Take the assignment α witnessing $\eta(s)$ at j, and let $j' < j$ be such that C_j is Γ, ψ_b for $b \in \{1, 2\}$ such that α satisfies $\neg\psi_b$. This j' and the unchanged assignment α witness $\eta(s + 1)$.

If C_j was derived by an $(\exists)$ inference, then C_j is $\Gamma, \exists w\, \psi$ and $C_{j'}$ is $\Gamma, \psi(t)$ for some $j' < j$. In this case, we first extend a given assignment α witnessing $\eta(s)$ at j to an assignment α' by letting all variables that are free in $C_{j'}$ but not in C_j have value 0. If $\psi(t)$ is not Π_2 (in which case $\neg\psi(t)$ is a Π_3 but not Σ_2 subformula of one of the induction axioms γ_i) or $\exists w\, \psi$ is Σ_1 (in which case $\forall w\, \neg\psi$ is Π_1 and satisfied under α, so $\neg\psi(t)$ is satisfied under α'), this is all we need to do in order to ensure that j', α' witness $\eta(s + 1)$. The remaining case is when $\psi(t)$ is Π_2 but $\exists w\, \psi$ is not. In that situation, $\neg\psi(t)$ arises from one of the γ_i's by deleting the initial universal quantifier block and substituting some terms $\bar{t}$ for the variables $\bar{v}$ appearing in that block. We know that α' satisfies $\neg\psi(t)$ (because γ_i is true), but we also have to argue that we can witness the existential quantifiers $\exists x < v_1\, \exists \bar{y}$ in $\neg\psi(t)$ by numbers below d_{s+1}. However, we know that we can find a value for x below $\alpha'(t_1)$, which is the value of a term with at most m symbols on arguments below d_s. Thus, by the definition of d_{s+1}, we can also find values for $\bar{y}$ corresponding to x in such a way that the maximum of these values is at most d_{s+1}.

Finally, if C_j was derived by a $(\forall)$ inference, then C_j is $\Gamma, \forall w\, \psi$ and $C_{j'}$ is $\Gamma, \psi(a)$ for some $j' < j$ and some variable a not appearing in C_j. Let α be an assignment witnessing $\eta(s)$ at j. There are two subcases to consider, depending on whether $\exists w\, \neg\psi$ is an unbounded Σ_2 formula or a Δ_0 formula. In the former case, we know from the inductive assumption that α satisfies $\exists w\, \neg\psi$ and that there is a number $e \leq d_s$ witnessing the quantifier $\exists w$. Then j' and the assignment $\alpha \cup \{a := e\}$ witness that $\eta(s + 1)$ holds. In the latter case, we know that α satisfies $\exists w\, \neg\psi$, and we also know that any number e witnessing the quantifier $\exists w$ must be bounded by the value of a term appearing in C_j (thus having at most m symbols) evaluated at elements of the range of α, all of which are below d_s. By definition of d_{s+1}, this means that $e \leq d_{s+1}$, so again j' and $\alpha \cup \{a := e\}$ witness that $\eta(s + 1)$ holds. □