

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

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

April 4, 2022

## 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  $\text{RCA}_0^*$ . Using the Chong-Mourad coding lemma, we show that in a model of  $\text{RCA}_0^*$  that does not satisfy  $\Sigma_1^0$  induction,  $\text{RT}_k^n$  is equivalent to its relativization to any proper  $\Sigma_1^0$ -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  $\text{RCA}_0^* + \text{RT}_k^n$  for  $n \geq 3$ . We show that they form a non-finitely axiomatizable subtheory of PA whose  $\Pi_3$  fragment coincides with  $\text{B}\Sigma_1 + \text{exp}$  and whose  $\Pi_{\ell+3}$  fragment for  $\ell \geq 1$  lies between  $\text{I}\Sigma_\ell \Rightarrow \text{B}\Sigma_{\ell+1}$  and  $\text{B}\Sigma_{\ell+1}$ . We also give a complete axiomatization of the first-order consequences of  $\text{RCA}_0^* + \text{RT}_k^2 + \neg\text{I}\Sigma_1$ . In general, we show that the first-order consequences of  $\text{RCA}_0^* + \text{RT}_k^2$  form a subtheory of  $\text{I}\Sigma_2$  whose  $\Pi_3$  fragment coincides with  $\text{B}\Sigma_1 + \text{exp}$  and whose  $\Pi_4$  fragment is strictly weaker than  $\text{B}\Sigma_2$  but not contained in  $\text{I}\Sigma_1$ .

Additionally, we consider a principle  $\Delta_2^0\text{-RT}_2^2$  which is defined like  $\text{RT}_2^2$  but with both the 2-colourings and the solutions allowed to be  $\Delta_2^0$ -sets rather than just sets. We show that the behaviour of  $\Delta_2^0\text{-RT}_2^2$  over  $\text{RCA}_0 + \text{B}\Sigma_2^0$  is in many ways analogous to that of  $\text{RT}_2^2$  over  $\text{RCA}_0^*$ , and that  $\text{RCA}_0 + \text{B}\Sigma_2^0 + \Delta_2^0\text{-RT}_2^2$  is  $\Pi_4$ - but not  $\Pi_5$ -conservative over  $\text{B}\Sigma_2$ . However, the statement we use to witness failure of  $\Pi_5$ -conservativity is not provable in  $\text{RCA}_0 + \text{RT}_2^2$ .

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, 27]), this remains open. In particular, it is not known whether Ramsey's Theorem for pairs and a fixed number of colours is  $\Pi_1^1$  conservative over the  $\Sigma_2^0$  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.

†Mathematical Institute, Tohoku University, keita.yokoyama.c2@tohoku.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,  $\text{RCA}_0^*$ , differs from the usual system  $\text{RCA}_0$  in that the  $\Sigma_1^0$  induction axiom of the latter is replaced by induction for bounded formulas only.

The study of  $\text{RCA}_0^*$  was initiated in [31] and continued in a number of later papers, e.g. [14, 32, 21, 10]. In the context of Ramsey theory, it is important that  $\Sigma_1^0$  induction is needed to show that each infinite set has arbitrarily large finite subsets. Hence, over  $\text{RCA}_0^*$  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, the third author [34] showed that for each fixed  $n, k$ ,  $\text{RCA}_0^*$  extended by Ramsey’s Theorem for  $n$ -tuples and  $k$  colours,  $\text{RT}_k^n$ , is  $\Pi_2$ -conservative over  $\text{I}\Delta_0 + \text{exp}$ . 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  $\text{RCA}_0^*$  is relevant to the traditional  $\text{RCA}_0$  framework as well. In fact, a large part of our original motivation for studying Ramsey’s Theorem over  $\text{RCA}_0^*$  was the desire to understand whether it can help in understanding  $\text{RT}_2^2$  over  $\text{RCA}_0$ . The jury is still out on that. However, it has turned out that Ramsey theory in  $\text{RCA}_0^*$  is a highly interesting topic in its own right. It gives rise to new examples of principles that are partially conservative but not  $\Pi_1^1$ -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  $\text{RCA}_0^*$  that are *not* models of  $\text{RCA}_0$ ,  $\text{RT}_k^n$  is equivalent to its relativizations to  $\Sigma_1^0$ -definable cuts. One consequence of that result is that in some models of  $\text{RCA}_0^*$ , Ramsey’s Theorem is computably true. This is not the case in the standard model of arithmetic or in any other model of  $\text{RCA}_0$ .

In Section 3, we use the equivalence from Section 2 to give an axiomatization of the first-order consequences of  $\text{RCA}_0^* + \text{RT}_k^n$  where  $n \geq 3$ . In each case, this turns out to be an unusual fragment of Peano Arithmetic that is  $\Pi_3$ - but not  $\Pi_4$ -conservative over  $\text{B}\Sigma_1 + \text{exp}$ . Moreover, it is not contained in  $\text{I}\Sigma_\ell$  for any  $\ell$ .

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

Then, in Section 5, we take a look at the question whether our results say anything about Ramsey’s Theorem for pairs over  $\text{RCA}_0$ . We consider a principle that can be viewed as a “jumped version” of  $\text{RT}_2^2$ , and we show that it is not  $\Pi_5$ -conservative over  $\text{RCA}_0 + \text{B}\Sigma_2^0$ . We also show that the most obvious sentence witnessing the lack of conservativity is unprovable in  $\text{RCA}_0 + \text{RT}_2^2$ . 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 [30] or [15]. 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 [15], and [18] covers more than enough.

The symbol  $\omega$  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  $\Sigma_\ell^0$ ,  $\Pi_\ell^0$  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  $\Sigma_\ell$ ,  $\Pi_\ell$ , 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  $\Sigma_\ell^0$  formula, we use notation like  $\Sigma_\ell^0(\bar{X})$ . We extend these conventions to naming theories: thus, for example,  $\text{B}\Sigma_2^0$  is the fragment of second-order arithmetic axiomatized by  $\Delta_0^0$  induction and  $\Sigma_2^0$  collection, whereas  $\text{B}\Sigma_2$  is the fragment of first-order arithmetic axiomatized by  $\Delta_0$  induction and  $\Sigma_2$  collection. (We follow the custom of some other authors in assuming also that all theories  $\text{I}\Sigma_\ell$ ,  $\text{B}\Sigma_\ell$  etc. contain  $\text{PA}^-$ , the theory of nonnegative parts of discretely ordered rings.)

*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  $\Pi_1^1$  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  $\psi$ ” (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  $\text{I}\Sigma_\ell$  proves (in fact, is equivalent to over  $\text{I}\Delta_0$ ) the scheme of *strong  $\Sigma_\ell$  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  $\Sigma_\ell$  formula, possibly with parameters.

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

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  $\Delta_\ell^0$ -*definable set*, or simply a  $\Delta_\ell^0$ -*set* (resp., a  $\Sigma_\ell^0$ -*definable set* or  $\Sigma_\ell^0$ -*set*) if it happens to be definable by a  $\Delta_\ell^0$  (resp.  $\Sigma_\ell^0$ ) formula. The notions of a  $\Delta_\ell$ -*set* and  $\Sigma_\ell$ -*set* are defined analogously.

Since most of the models we study only satisfy  $\Delta_1^0$ -comprehension,  $\Delta_\ell^0$ -sets for  $\ell \geq 2$  and  $\Sigma_\ell^0$ -sets for  $\ell \geq 1$  will not always be sets. However, using appropri-

ate universal formulas, we can quantify over  $\Delta_\ell^0$ - or over  $\Sigma_\ell^0$ -sets using second-order quantifiers (e.g. “for every  $X$ , and every equivalent pair of a  $\Sigma_\ell^0(X)$  and a  $\Pi_\ell^0(X)$  formula, ...”). On the other hand, quantification over  $\Delta_\ell$ - or over  $\Sigma_\ell$ -sets is first-order. We write  $\Delta_\ell\text{-Def}(M)$  (resp.  $\Delta_\ell^0\text{-Def}(M, \mathcal{X})$ ) for the collection of  $\Delta_\ell$ -definable subsets of  $M$  (resp. the subsets of  $M$  that are  $\Delta_\ell^0$ -definable in  $(M, \mathcal{X})$ ).

For  $\ell \geq 1$ , let  $\text{Sat}_\ell(x, y)$  be the usual universal  $\Sigma_\ell$  formula and let  $\text{Sat}_\ell(x, y, X)$  be the usual universal  $\Sigma_\ell^0$  formula with the unique second-order variable  $X$ . Then  $0^{(\ell)}$  is the  $\Sigma_\ell$  definable set  $\{e : \text{Sat}_\ell(e, e)\}$ ; we write  $0'$  for  $0^{(1)}$ . Similarly, if  $A$  is a set, then  $A^{(\ell)}$  is  $\{e : \text{Sat}_\ell(e, e, A)\}$ ; this notion is generalized in a natural way to the case where  $A$  is merely a definable set. Note that  $\text{B}\Sigma_\ell$  is enough to prove that  $0^{(\ell+1)}$  and  $(0^{(\ell)})'$  are mutually  $\Delta_1$ -definable.

For  $n, k \in \omega$ ,  $\text{RT}_k^n$  stands for the usual formulation of Ramsey’s Theorem for  $n$ -tuples in second-order arithmetic: “for every function  $f : [\mathbb{N}]^n \rightarrow k$ , 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  $\Sigma_1^0$  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  $\text{I}\Sigma_1^0$  [34].

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  $(M, \mathcal{X}) \models \text{RCA}_0^*$ , and  $I$  is a  $\Sigma_1^0$ -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  $M$ ,  $(s)_{\text{Ack}}$  stands for  $\{a \in M : M \models a \in_{\text{Ack}} s\}$ , where  $\in_{\text{Ack}}$  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  $\text{Cod}(M/I)$  of *subsets of  $I$  coded in  $M$*  is  $\{(s)_{\text{Ack}} \cap I : s \in M\}$ . If  $M$  satisfies induction for any of the classes of formulas  $\Gamma$  that we consider in this paper, this will coincide with  $\{A \cap I : A \text{ a } \Gamma\text{-definable subset of } M\}$ .

The collection  $\text{Cod}(M/\omega)$  is commonly referred to as the *standard system* of  $M$  and denoted by  $\text{SSy}(M)$ . The following combination of standard model-theoretic facts discussed in [18] and well-known results on  $\text{RT}_k^n$  presented e.g. in [15] will often be used without notice, for instance in the proof of Theorem 10.

**Fact.** Let  $\mathcal{S} \subseteq \mathcal{P}(\omega)$  be such that  $(\omega, \mathcal{S}) \models \text{WKL}_0$  (such a family  $\mathcal{S}$  is known as a *Scott set*). If  $\mathcal{S}$  is countable, then for every consistent computably axiomatized theory  $T \supseteq \text{I}\Delta_0 + \text{exp}$  there exists a model  $M \models T$  such that  $\text{SSy}(M) = \mathcal{S}$ , and for every  $\ell \geq 1$  there exists  $M \models \text{B}\Sigma_\ell$  such that  $\omega$  is  $\Sigma_\ell$ -definable in  $M$  and  $\text{SSy}(M) = \mathcal{S}$ .

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

We will sometimes want to abuse notation and use  $\text{Cod}(M/I)$  for the collection of  $n$ -ary relations on  $I$  coded in  $M$ , that is for

$$\{(s)_{\text{Ack}} \cap \{\langle i_1, \dots, i_n \rangle : i_1, \dots, i_n \in I\} : s \in M\},$$

for some fixed  $n > 1$ . Here  $\langle i_1, \dots, i_n \rangle$  is defined in terms of the usual Cantor pairing function. If  $I$  is not closed under multiplication, then such coded  $n$ -ary

relations might not be elements of  $\text{Cod}(M/I)$  in the strict sense, but that should not lead to any confusion.

We define the iterated exponential function  $\text{exp}_n(x)$  by:  $\text{exp}_0(x) = x$ , and  $\text{exp}_{n+1}(x) = 2^{\text{exp}_n(x)}$ .

## 2 Characterization in terms of cuts

In this section, we prove a basic result which underlies our subsequent analysis of Ramsey's Theorem over  $\text{RCA}_0^*$ : if  $\Sigma_1^0$  induction fails but  $\Sigma_1^0$  collection holds, then Ramsey's Theorem is equivalent to its own relativization to a proper  $\Sigma_1^0$ -definable cut. To prove this, we make use of an important fact about coding sets in models of collection.

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

**Corollary 2.** *Let  $(M, \mathcal{X}) \models \text{RCA}_0^* + \text{B}\Sigma_n^0$  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  $\Sigma_n^0$ -definable sets, then  $X \in \text{Cod}(M/I)$ .*

**Theorem 3.** *Let  $(M, \mathcal{X}) \models \text{RCA}_0^*$  and let  $I \subseteq M$  be a  $\Sigma_1^0$ -definable proper cut in  $M$ . Then for every  $n, k \in \omega$ :*

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

*Remark.* Theorem 3 has now been generalized from  $\text{RT}_k^n$  to a larger class of statements in [11].

*Proof.* Let  $(M, \mathcal{X})$  be a model of  $\text{RCA}_0^*$  and let  $I \subseteq M$  be a  $\Sigma_1^0$ -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  $(M, \mathcal{X}) \models \text{RT}_k^n$ . Let  $f: [I]^n \rightarrow k$  be coded by  $c \in M$ . We can use  $f$  to define a colouring  $\check{f}: [A]^n \rightarrow k$  in the following way:

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

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  $\check{f}$ :

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

Note that  $\check{f}$  is  $\Delta_1(A, c)$ -definable, so  $\check{f} \in \mathcal{X}$ . By  $\text{RT}_k^n$ , 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  $\text{Cod}(M/I)$ . Clearly,  $\hat{H}$  is cofinal in  $I$  and homogeneous for  $f$ .

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

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

is in  $\text{Cod}(M/I)$ . Since  $(I, \text{Cod}(M/I)) \models \text{RT}_k^n$ , there is  $\text{Cod}(M/I) \ni H \subseteq I$  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  $f$ .  $\square$

*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  $\Sigma_1^0$ -definable in  $(M, \mathcal{X})$ . Thus, Theorem 3 means that over  $\text{RCA}_0^*$ , once  $\text{I}\Sigma_1^0$  fails, Ramsey's Theorem becomes in some sense a first-order property. In particular, it can be satisfied in some structures of the form  $(M, \Delta_1\text{-Def}(M))$  ("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  $\text{RCA}_0^* + \text{RT}_k^n$  for  $n \geq 3$ . We begin with the easy but useful observation that, just like over  $\text{RCA}_0$ , 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  $n, k \geq 2$ ,  $\text{RCA}_0^* \vdash (\text{RT}_k^n \Leftrightarrow \text{RT}_{k+1}^n)$ .*

*Proof.* Assume  $\text{RCA}_0^* + \text{RT}_k^n$  and let  $f: [\mathbb{N}]^n \rightarrow k+1$ . Consider the colouring  $g: [\mathbb{N}]^n \rightarrow k$  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  $\Sigma_1^0$ -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  $f \upharpoonright_{[A]^n}$  takes at most the two values  $k-1$  and  $k$ . Define a 2-colouring of  $[\mathbb{N}]^n$  by:

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

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  $\Delta_1^0$ -comprehension: it is clearly  $\Sigma_1^0$ -definable, and its complement is the union of  $\mathbb{N} \setminus A$  and the  $\Sigma_1^0$ -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  $f$ .  $\square$

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

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

It is well known that each  $\Delta_\ell\text{-RT}_k^n$  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)  $\text{I}\Sigma_1$  proves that there is a  $\Delta_1$ -definable 2-colouring of  $[\mathbb{N}]^n$  with no  $\Sigma_1$ -definable infinite homogeneous set,
- (b) for each  $l \geq 1$ ,  $\text{I}\Sigma_{\ell+1}$  proves that there is a  $\Delta_\ell$ -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 [16] in  $\text{I}\Sigma_{\ell+1}$ . 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 [33] of a computable 2-colouring of pairs with no r.e. homogeneous set within  $\text{I}\Sigma_1$ . Instead of that, we choose to formalize a weaker variant of the argument of [16] proving (b) for  $\ell = 1$ . We define a computable function  $f: [\mathbb{N}]^2 \rightarrow 2$  in the following way. At stage  $s$ , we determine the values  $f(n, s)$  for  $n < s$ . To do this, we consider all  $\Sigma_1$  formulas with codes  $0, \dots, \lfloor (s-1)/2 \rfloor$ , in that order. Given  $e \leq \lfloor (s-1)/2 \rfloor$ , if  $e$  is the code of a  $\Sigma_1$  formula  $\exists v \delta(x, v)$  and there are at least  $2e + 2$  elements  $x < s$  such that  $\exists v \leq s \text{Sat}_0(\ulcorner \delta \urcorner, (x, v))$  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  $0, \dots, \lfloor (s-1)/2 \rfloor$  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  $\Sigma_1$  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  $x_0 < \dots < x_{2e+1}$ . By another application of  $\Sigma_1$  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  $f$ .  $\square$

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

*Proof.* Let  $M \models \text{RCA}_0^* + \text{RT}_2^n + \text{I}\Sigma_\ell$ . We will prove by induction on  $j \leq \ell$  that  $0^{(j)} \in \mathcal{X}$ . For  $j = \ell$ , this will immediately imply  $\Delta_{\ell+1}\text{-Def}(M) \subseteq \mathcal{X}$  and  $M \models \text{B}\Sigma_{\ell+1}$  because  $(M, \mathcal{X})$  satisfies  $\Delta_1^0$  comprehension and  $\text{B}\Sigma_1^0$ .

The base step of the induction holds by  $\Delta_1^0$ -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  $\text{RT}_2^3$  whose solutions compute  $0'$  and relativize it to  $0^{(j)}$ :

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

The colouring  $f$  is  $\Delta_1(0^{(j)})$ -definable, so  $f \in \mathcal{X}$ . By  $\text{RT}_2^n$ , 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  $\text{I}\Sigma_\ell$  we have strong  $\Sigma_{j+1}$  collection, so for any given  $x$  there is a bound  $w$  such that for any  $\Sigma_{j+1}$  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  $0^{(j)} \oplus H$  as follows: given a  $\Sigma_{j+1}$  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 it 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.  $\square$

We are now ready to give an axiomatization of the first-order part of  $\text{RCA}_0^* + \text{RT}_2^n$  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  $\text{R}^n$  be the theory:*

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

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

*Remark.* By Lemma 6, already  $\text{I}\Sigma_\ell$  is inconsistent with  $\Delta_\ell\text{-RT}_2^n$ , so an equivalent axiomatization of  $\text{R}^n$  would consist of  $\text{B}\Sigma_1 + \text{exp}$  and the sentences  $\text{B}\Sigma_\ell \Rightarrow (\text{B}\Sigma_{\ell+1} \vee \Delta_\ell\text{-RT}_2^n)$  for all  $\ell \geq 1$ .

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

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

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

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



Let  $I$  be a  $\Sigma_{j+1}$ -definable proper cut in  $M$ . The cut  $I$  is  $\Sigma_1^0$ -definable in  $(M, \Delta_{j+1}\text{-Def}(M))$  and thus also in  $(M, \mathcal{X})$ . Moreover, both of these structures satisfy  $\text{RCA}_0^*$ . Therefore, Theorem 3 and the fact that  $(M, \mathcal{X}) \models \text{RCA}_0^* + \text{RT}_2^n$  let us conclude that  $(M, \Delta_{j+1}\text{-Def}(M)) \models \text{RT}_2^n$  as well.  $\square$

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

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

Kaye [19] showed that  $\text{IB} + \text{exp}$  implies the theory of all  $\kappa$ -like models of arithmetic (for  $\kappa$  possibly singular). It is now known (see [13, Section 3.3], [2, Section 6]) that  $\text{IB} + \text{exp}$  is actually strictly stronger than the theory of all  $\kappa$ -like models.

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

- (a) *the first-order consequences of  $\text{RCA}_0^* + \text{RT}_2^n$  are strictly in between  $\text{IB} + \text{exp}$  and  $\text{PA}$ ; as a result, they are not finitely axiomatizable.*
- (b) *the  $\Pi_3$  consequences of  $\text{RCA}_0^* + \text{RT}_2^n$  coincide with  $\text{B}\Sigma_1 + \text{exp}$ ; for  $\ell \geq 1$ , the  $\Pi_{\ell+3}$  consequences are strictly in between*

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

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

*Proof.* We first prove (b). As in Theorem 8, we let  $\text{R}^n$  stand for the first-order consequences of  $\text{RCA}_0^* + \text{RT}_2^n$ .

It follows immediately from the definition of  $\text{RCA}_0^*$  and Lemma 7 that the  $\Pi_{\ell+3}$  consequences of  $\text{R}^n$  include  $\text{B}\Sigma_1 + \text{exp}$  and  $\text{I}\Sigma_j \Rightarrow \text{B}\Sigma_{j+1}$  for each  $j \leq \ell$ . For  $\ell \geq 1$ , the inclusion is strict, because the statement

$$(\text{B}\Sigma_{\ell+1} \wedge \text{exp}) \vee \bigvee_{j=1}^{\ell} \Delta_j\text{-RT}_2^n$$

is  $\Pi_{\ell+3}$  but not provable in  $\text{B}\Sigma_1 + \text{exp} + \bigwedge_{1 \leq j \leq \ell} (\text{I}\Sigma_j \Rightarrow \text{B}\Sigma_{j+1})$ . To see the unprovability, consider a model  $M \models \text{B}\Sigma_\ell + \text{exp}$  such that  $\omega$  is  $\Sigma_\ell$ -definable in  $M$  and  $(\omega, \text{SSy}(M)) \not\models \text{RT}_2^n$ . Then, clearly,  $M \models \text{I}\Sigma_j \Rightarrow \text{B}\Sigma_{j+1}$  for each  $j \leq \ell$ ; in fact,  $M$  is a model of IB. However, Lemma 6 implies that  $(M, \Delta_j\text{-Def}(M)) \not\models \text{RT}_2^n$  for each  $1 \leq j \leq \ell - 1$ . On the other hand,  $(M, \Delta_\ell\text{-Def}(M))$  is a model of  $\text{RCA}_0^*$  in which  $\omega$  is  $\Sigma_1^0$ -definable, so by Theorem 3 and the choice of  $\text{SSy}(M)$  it does not satisfy  $\text{RT}_2^n$  either.

Using a model  $M$  chosen similarly but with  $(\omega, \text{SSy}(M)) \models \text{RT}_2^n$ , we get  $(M, \Delta_{\ell+1}\text{-Def}(M)) \models \text{RT}_2^n + \neg \text{B}\Sigma_{\ell+1}$ . Thus,  $\text{R}^n$  does not prove  $\text{B}\Sigma_{\ell+1}$  for  $\ell \geq 1$ .

To see that all  $\Pi_{\ell+3}$  consequences of  $\text{R}^n$  follow from  $\text{B}\Sigma_{\ell+1}$  for  $\ell \geq 1$  let the  $\Sigma_{\ell+3}$  formula  $\psi := \exists x \forall y \exists z \pi(x, y, z)$  be consistent with  $\text{B}\Sigma_{\ell+1}$ , let  $K \models \text{B}\Sigma_{\ell+1} \wedge \psi$  be such that  $(\omega, \text{SSy}(K)) \models \text{RT}_2^n$ , and let  $a \in K$  be a witness for the initial existential quantifier in  $\psi$ . By  $\text{B}\Sigma_{\ell+1}$ , the function

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

and “true  $\Sigma_\ell$  sentences with codes  $\leq y$  are witnessed  $\leq w$ ”

is total and  $\Delta_{\ell+1}$ -definable in  $K$ . Let  $M$  be the cut  $\sup_K(\{f^m(a) : m \in \omega\})$ . Then  $M \models \text{B}\Sigma_{\ell+1} \wedge \psi$  and  $\omega$  is  $\Sigma_{\ell+1}$ -definable in  $M$ . Since  $(\omega, \text{SSy}(M)) \models \text{RT}_2^n$ , we get  $(M, \Delta_\ell\text{-Def}(M)) \models \text{RT}_2^n$  by Theorem 3, so  $M \models \text{R}^n \wedge \psi$ .

The proof that the  $\Pi_3$  consequences of  $\text{R}^n$  follow from  $\text{B}\Sigma_1 + \text{exp}$  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  $\pi$  is now a  $\Delta_0$  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  $M \models \text{B}\Sigma_{\ell+1} \wedge \psi$ , but we need to guarantee that  $M \models \text{exp}$ .

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

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

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

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  $\text{B}\Sigma_\ell + \text{exp} + \Delta_\ell\text{-RT}_2^n$  is consistent.*

*Remark.* As mentioned in Section 1, results such as Theorem 10 can be converted from purely arithmetical to  $\Pi_1^1$  form by relativizing to second-order parameters. In Theorem 10(a), the appropriate relativization of the scheme  $\text{IB}$  takes the form  $\forall X (\text{I}\Sigma_k(X) \Rightarrow \text{B}\Sigma_{k+1}(X))$  for each  $k$ . In Section 4, we will also consider a weaker relativization of  $\text{IB}$ : see the remark after Corollary 15.

*Question 1.* Does  $\text{RCA}_0^* + \text{RT}_2^3$  imply  $\text{RT}_2^4$ ? More generally, does  $\text{RCA}_0^* + \text{RT}_2^n$  imply  $\text{RT}_2^{n+1}$  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  $\text{RCA}_0^* + \text{RT}_2^2$  strongly depends on the amount of induction satisfied by the underlying first-order model.

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

- (a)  $\text{R}^2 \wedge \neg\text{I}\Sigma_1$  is axiomatized by  $\text{B}\Sigma_1 + \text{exp} + \Delta_1\text{-RT}_2^2$ .
- (b)  $\text{I}\Sigma_2$  implies  $\text{R}^2$ .
- (c) Over  $\text{B}\Sigma_2$ ,  $\text{R}^2$  is implied by, and consistent with, both the first-order consequences of  $\text{RCA}_0 + \text{RT}_2^2$  and the statement  $\Delta_2\text{-RT}_2^2$ .

(d)  $R^2$  implies every first-order sentence  $\psi$  such that both  $B\Sigma_2 \vdash \psi$  and  $RCA_0^* + \neg I\Sigma_1^0 \vdash \psi$ .

*Proof.* We first prove (a). Clearly, if  $M \models B\Sigma_1 + \text{exp}$  and  $(M, \Delta_1\text{-Def}(M)) \models RT_2^2$ , then  $M$  satisfies  $R^2$  (as well as  $\neg I\Sigma_1$ , by Lemma 6). On the other hand, let  $(M, \mathcal{X}) \models RCA_0^* + RT_2^2 + \neg I\Sigma_1$ . Obviously,  $M$  satisfies  $B\Sigma_1 + \text{exp}$ . Let  $I$  be a proper  $\Sigma_1$ -definable cut in  $M$ . Applying Theorem 3 two times, we get first  $(I, \text{Cod}(M/I)) \models RT_2^2$  and then  $(M, \Delta_1\text{-Def}(M)) \models RT_2^2$ .

Statement (b) follows immediately from the result of [4] that  $RCA_0 + I\Sigma_2^0 + RT_2^2$  is conservative over  $I\Sigma_2$ .

We turn to (c). It is clear that  $R^2$  is implied by the first-order consequences of  $RCA_0 + RT_2^2$ . Meanwhile,  $R^2$  is also satisfied by any model  $M \models B\Sigma_2 + \Delta_2\text{-}RT_2^2$  since  $(M, \Delta_2\text{-Def}(M)) \models RCA_0^* + RT_2^2$ . The consistency of  $B\Sigma_2 + \Delta_2\text{-}RT_2^2$  is Corollary 11 for  $\ell = n = 2$ .

Finally, to see that (d) holds, let  $\psi$  be provable both in  $B\Sigma_2$  and in  $RCA_0^* + \neg I\Sigma_1^0$ . We check that  $RCA_0^* + RT_2^2 \vdash \psi$ . Let  $(M, \mathcal{X}) \models RCA_0^* + RT_2^2$ . If  $(M, \mathcal{X}) \models RCA_0$ , then  $M \models B\Sigma_2$ , so  $M \models \psi$ . Otherwise,  $M \models RCA_0^* + \neg I\Sigma_1^0$ , so  $M \models \psi$  as well.  $\square$

Parts (a) and (b) of Theorem 12 give a complete axiomatization of the first-order consequences of  $RCA_0^* + RT_2^2$  over, respectively,  $\neg I\Sigma_1$  and  $I\Sigma_2$ . However, the situation in the region between  $I\Sigma_1$  and  $I\Sigma_2$  is much less clear.

As mentioned in the introduction, it is open whether  $RCA_0 + RT_2^2$  is arithmetically conservative over  $B\Sigma_2$ . Therefore, it is consistent with what we know that already  $B\Sigma_2$  implies the first-order consequences of  $RCA_0^* + RT_2^2$ .

On the other hand, we will now use Theorem 12(d) to show that there are some first-order sentences provable in  $RCA_0^* + RT_2^2$  but not in  $I\Sigma_1$ . 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  $B\Sigma_2^0$  (or even somewhat weaker statements) over  $RCA_0$ .

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

The  $\Sigma_\ell$  generalized pigeonhole principle,  $GPHP(\Sigma_\ell)$ , asserts that for every  $\Sigma_\ell$  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  $C\Sigma_\ell$  was defined in [29]. It is known that  $I\Sigma_\ell$  does not imply  $C\Sigma_{\ell+1}$  [12, Proposition 3.1]. The principle  $GPHP(\Sigma_\ell)$  was defined in [19], where it was also observed that the theory of all  $\kappa$ -like models of  $I\Delta_0$  implies  $GPHP(\Sigma_\ell)$  for all  $\ell$ .

Clearly,  $GPHP(\Sigma_\ell)$  implies  $C\Sigma_\ell$  for each  $\ell \geq 1$ . In turn,  $GPHP(\Sigma_\ell)$  is implied by  $B\Sigma_\ell$  for  $\ell \geq 2$ , and  $GPHP(\Sigma_1)$  is implied by  $B\Sigma_1 + \text{exp}$ , because  $B\Sigma_\ell$  is for each  $\ell \geq 1$  equivalent over  $I\Delta_0 + \text{exp}$  to the usual pigeonhole principle for  $\Sigma_\ell$  maps [9]. It follows from [2] that the implications from  $B\Sigma_\ell$  to  $GPHP(\Sigma_\ell)$  are strict.

$C\Sigma_2$  is known to be a consequence of some theories studied in reverse mathematics that do not imply  $B\Sigma_2$ , such as  $RCA_0$  plus the Rainbow Ramsey Theorem for pairs [8] and  $RCA_0$  plus the existence of 2-random reals [13].

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  $\text{IS}_k^0$  (resp.  $\text{BS}_k^0$ ) means  $\forall X \text{IS}_k(X)$  (resp.  $\forall X \text{BS}_k(X)$ ).

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

- (a)  $\forall X (\text{BS}_\ell(X) \Rightarrow \text{GPHP}(\Sigma_\ell(X)))$ ,
- (b)  $(\text{BS}_k^0 \wedge \neg \text{IS}_k^0) \Rightarrow \forall X \text{GPHP}(\Sigma_\ell(X))$ .

Theorem 14 part (b) can be obtained by relativizing Kaye's proof of the result that any model of  $\text{BS}_1 + \text{exp} + \neg \text{IS}_1$  is elementarily equivalent to an  $\aleph_\omega$ -like structure [19, Theorem 2.4]. A model of  $\neg \text{IS}_1(A) + \neg \text{GPHP}(\Sigma_\ell(B)) + \text{BS}_1(A \oplus B) + \text{exp}$  would also be elementary equivalent to  $\aleph_\omega$ -like model, but clearly such a structure can never violate the scheme  $\text{GPHP}(\Gamma)$  for any class of formulas  $\Gamma$ .

The proof of Theorem 14 we give below is considerably simpler than that of [19, 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  $\text{GPHP}(\Sigma_\ell)$ , with no model-theoretic detours, in for instance  $\text{BS}_1 + \text{exp} + \neg \text{IS}_1$ .

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

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

Let  $(M, A)$  be a countable model of  $\text{BS}_1(A) + \text{exp} + \neg \text{IS}_1(A)$ . 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  $b > \text{exp}_m(a)$  for each  $m \in \omega$ . To prove that  $M \models \text{GPHP}(\Sigma_\ell)$ , we will use a technique based on the fact that models of  $\text{BS}_1^0 + \text{exp} + \neg \text{IS}_1^0$  have many automorphisms [23, 24, 17].

By a standard argument (see e.g. [10, Theorem 4.6]), the model  $M$  can be end-extended to a model  $K \models \text{I}\Delta_0$  such that  $A \in \text{Cod}(K/M)$ . Since elements coding  $A$  are downwards cofinal in  $K \setminus M$ , there is an element  $d \in K$  coding  $A$  and small enough that  $\text{exp}_2(d)$  exists in  $K$ . By [26], there is a  $\Delta_0$  formula with parameter  $\text{exp}_2(d)$  that defines satisfaction for  $\Delta_0$  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  $b > \text{exp}_m(a)$  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 [25] (see the proof of Lemma 3.4 in [17] for a detailed argument and [24] for a brief discussion) to derive the existence of an automorphism  $\alpha$  of  $[0, d]$  such that  $\alpha$  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  $\alpha|_M$  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  $\alpha$  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  $M \models \text{GPHP}(\Sigma_\ell)$  for each  $\ell$ .  $\square$

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

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

*Remark.* Let the usual relativization of IB, namely  $\forall X (\text{I}\Sigma_k(X) \Rightarrow \text{B}\Sigma_{k+1}(X))$  for each  $k$ , be called “strong”, and let the “weak” relativization of IB consist of the statements  $\text{I}\Sigma_k^0 \Rightarrow \text{B}\Sigma_{k+1}^0$  for each  $k$ . In Theorem 10, we showed that  $\text{RCA}_0^* + \text{RT}_2^3$  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  $\text{GPHP}(\Sigma_\ell)$ .

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

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

**Corollary 16.** *The first-order consequences of  $\text{RCA}_0^* + \text{RT}_2^2$  follow from  $\text{I}\Sigma_2$ . The  $\Pi_3$  consequences coincide with  $\text{B}\Sigma_1 + \text{exp}$ . The  $\Pi_4$  consequences are strictly weaker than  $\text{B}\Sigma_2$  but do not follow from  $\text{I}\Sigma_1$ .*

*Proof.* The provability from  $\text{I}\Sigma_2$  is part (b) of Theorem 12. The fact that the  $\Pi_3$  consequences of  $\text{RCA}_0^* + \text{RT}_2^2$  coincide with  $\text{B}\Sigma_1 + \text{exp}$  and that the  $\Pi_4$  consequences are strictly weaker than  $\text{B}\Sigma_2$  is proved like in Theorem 10. Finally, Corollary 15 implies that  $\text{C}\Sigma_2$  is an example of a  $\Pi_4$  sentence that follows from  $\text{RCA}_0^* + \text{RT}_2^2$  but not  $\text{I}\Sigma_1$ .  $\square$

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

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

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

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

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

We have no strong reasons to believe that the answer is “yes”. However, it should be pointed out that, since  $\text{RCA}_0 + \text{RT}_2^2$  proves  $\text{B}\Sigma_2$ , answering “no” would involve constructing a model of  $\text{I}\Sigma_1 + \neg\text{B}\Sigma_2$  that expands to a model of  $\text{B}\Sigma_1^0 + \neg\text{I}\Sigma_1^0$  – in the terminology of [23], a model of  $\text{I}\Sigma_1 + \neg\text{B}\Sigma_2$  that is not always semiregular. The existence of such a model itself seems to be open.

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

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

## 5 Relativizing Ramsey

In this final section, we take up the question whether our results on  $\text{RCA}_0^* + \text{RT}_2^2$  shed any light on the problem of characterizing the first-order consequences of  $\text{RCA}_0 + \text{RT}_2^2$ . To this end, we introduce a principle in which both the instances and solutions to Ramsey's Theorem are allowed to be  $\Delta_2^0$ -sets rather than sets.

**Definition 17.**  $\Delta_2^0\text{-RT}_2^2$  is the  $\Pi_2^1$  statement: “for every  $\Delta_2^0$ -set  $f$  which is a 2-colouring of  $[\mathbb{N}]^2$ , there exists an infinite homogeneous  $\Delta_2^0$ -set”.

Note that  $\Delta_2^0\text{-RT}_2^2$  is a genuine  $\Pi_2^1$  statement, which should not be confused with the  $\Pi_1^1$  statement relativizing  $\Delta_2\text{-RT}_2^2$ , namely “for every set  $X$ ,  $\Delta_2(X)\text{-RT}_2^2$  holds”. Of course, in a model of the form  $(M, \Delta_1\text{-Def}(M))$ , the statement  $\Delta_2^0\text{-RT}_2^2$  will be equivalent to  $\Delta_2\text{-RT}_2^2$ .

We are interested in studying  $\Delta_2^0\text{-RT}_2^2$  over  $\text{RCA}_0 + \text{B}\Sigma_2^0$ , especially in the case where  $\text{I}\Sigma_2^0$  fails. The following proposition shows that in such a context,  $\Delta_2^0\text{-RT}_2^2$  behaves somewhat analogously to  $\text{RT}_2^2$  over  $\text{RCA}_0^*$ , so we can investigate it using the methods developed in Sections 2-4.

**Lemma 18.** *For any model  $(M, \mathcal{X}) \models \text{RCA}_0 + \text{B}\Sigma_2^0$ :  $(M, \mathcal{X}) \models \Delta_2^0\text{-RT}_2^2$  iff  $(M, \Delta_2^0\text{-Def}(M, \mathcal{X})) \models \text{RCA}_0^* + \text{RT}_2^2$ . As a consequence:*

- (a) *if  $I$  is a  $\Sigma_2^0$ -definable proper cut in  $(M, \mathcal{X})$ , then  $(M, \mathcal{X}) \models \Delta_2^0\text{-RT}_2^2$  iff  $(I, \text{Cod}(M/I)) \models \text{RT}_2^2$ ,*
- (b) *the first-order consequences of  $\text{RCA}_0 + \text{B}\Sigma_2 + \neg\text{I}\Sigma_2 + \Delta_2^0\text{-RT}_2^2$  are axiomatized by  $\text{B}\Sigma_2 + \Delta_2\text{-RT}_2^2$ ,*
- (c)  *$\text{RCA}_0 + \text{B}\Sigma_2^0 + \Delta_2^0\text{-RT}_2^2$  is  $\Pi_4$ - but not  $\Pi_5$ -conservative over  $\text{B}\Sigma_2$ .*

*Proof.* The fact that a model  $(M, \mathcal{X})$  satisfies  $\text{RCA}_0 + \text{B}\Sigma_2^0 + \Delta_2^0\text{-RT}_2^2$  exactly if  $(M, \Delta_2^0\text{-Def}(M, \mathcal{X})) \models \text{RCA}_0^* + \text{RT}_2^2$  is immediate from the definitions. Thus (a) follows from Theorem 3, because a cut  $I$  is  $\Sigma_2^0$  definable in  $(M, \mathcal{X}) \models \text{B}\Sigma_2^0$  exactly if it is  $\Sigma_1^0$ -definable in  $(M, \Delta_2^0\text{-Def}(M, \mathcal{X}))$ .

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

To show that  $\text{RCA}_0 + \text{B}\Sigma_2^0 + \Delta_2^0\text{-RT}_2^2$  is  $\Pi_4$ -conservative over  $\text{B}\Sigma_2$ , relativize to  $0'$  the argument used to prove  $\Pi_3$ -conservativity of  $\text{RCA}_0^* + \text{RT}_2^2$  over  $\text{B}\Sigma_1 + \text{exp}$  in Theorem 10(b). To show lack of  $\Pi_5$ -conservativity, consider the sentence  $\neg\text{I}\Sigma_2 \Rightarrow \Delta_2\text{-RT}_2^2$ . This is a  $\Pi_5$  statement, and it is provable in  $\text{RCA}_0 + \text{B}\Sigma_2^0 + \Delta_2^0\text{-RT}_2^2$  by (b). On the other hand, it is not provable in  $\text{B}\Sigma_2$ , as can be seen by applying (a) to any model  $M \models \text{B}\Sigma_2$  with  $\Sigma_2$ -definable  $\omega$  and  $(\omega, \text{SSy}(M)) \not\models \text{RT}_2^2$ . This proves (c).  $\square$

Since Lemma 18 shows that  $\Delta_2^0\text{-RT}_2^2$  is not  $\Pi_5$ -conservative over  $\text{B}\Sigma_2$ , while the conservativity of  $\text{RCA}_0 + \text{RT}_2^2$  over  $\text{B}\Sigma_2$  is a well-known open problem, it is natural to ask whether  $\text{RT}_2^2$  might imply  $\Delta_2^0\text{-RT}_2^2$ , at least in the particularly relevant setting of models of  $\text{B}\Sigma_2^0 + \neg\text{I}\Sigma_2^0$ .

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  $\Pi_5$ -conservativity of  $\Delta_2^0\text{-RT}_2^2$  is unprovable in  $\text{RT}_2^2$ . To prove this, we will have to make use of a connection between properties of infinite  $\Delta_2$ -sets and the consistency of  $\text{IS}_1$  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  $\text{IS}_1$  proves:*

$$\forall x \left[ \text{“every infinite } \Delta_2\text{-set contains at least } \exp_{p(x)}(2) \text{ elements”} \right. \\ \left. \Rightarrow \text{Con}_x(\text{IS}_1) \right],$$

where  $\text{Con}_x(T)$  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  $\text{IS}_2$  is equivalent to uniform  $\Pi_4$ -reflection for  $\text{I}\Delta_0 + \text{exp}$  (see e.g. [1, Theorem 7]). To see this, note that (over  $\text{I}\Delta_0 + \text{exp}$  as a base theory)  $\text{IS}_2$  is equivalent to the statement that each infinite  $\Delta_2$ -set contains arbitrarily large finite sets, while  $\Pi_4$ -reflection for  $\text{I}\Delta_0 + \text{exp}$  implies  $\text{Con}(\text{IS}_1)$ .

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

*Proof.* The fact that  $\text{RCA}_0 + \text{B}\Sigma_2^0 + \neg\text{IS}_2^0 + \Delta_2^0\text{-RT}_2^2$  does not prove  $\text{RT}_2^2$  is witnessed by any structure of the form  $(M, \Delta_1\text{-Def}(M))$ , where  $M \models \text{B}\Sigma_2$  has  $\Sigma_2$ -definable  $\omega$  and  $(\omega, \text{SSy}(M)) \models \text{RT}_2^2$ . By Lemma 18(a), such a structure satisfies  $\Delta_2^0\text{-RT}_2^2$ , but by Lemma 6(a) it cannot satisfy  $\text{RT}_2^2$ .

In the other direction, such a “quick and dirty” argument does not seem to be currently available: of the known constructions producing models of  $\text{RCA}_0 + \text{RT}_2^2 + \neg\text{IS}_2^0$ , that of [6, 7] involves strong constraints on  $\text{SSy}(M)$ , and that of [27, 22] does not give a  $\Sigma_2^0$ -definable  $\omega$ . To show that  $\text{RCA}_0 + \text{RT}_2^2 + \neg\text{IS}_2^0$  does not imply  $\Delta_2^0\text{-RT}_2^2$ , it is enough to prove the “Moreover” part of the statement, namely:

$$\text{RCA}_0 + \text{RT}_2^2 \not\vdash \neg\text{IS}_2 \Rightarrow \Delta_2\text{-RT}_2^2.$$

This we do by means of a proof speedup argument. By [20, Lemma 3.2],  $\text{RCA}_0^* + \text{RT}_2^2$  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  $\text{B}\Sigma_2 + \Delta_2\text{-RT}_2^2$  proves “for every  $k$ , if every infinite  $\Delta_2$ -set contains at least  $k$  elements, then every infinite  $\Delta_2$ -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  $\text{B}\Sigma_2 + \Delta_2\text{-RT}_2^2$ . This in turn implies (cf. [28, Theorem 3.4.1]) that, for each  $n \in \omega$ , there is a poly( $n$ )-size proof of

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

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

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

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

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

However, the size of the smallest proof of  $\text{Con}_{\exp_n(2)}(I\Sigma_1)$  in  $I\Sigma_1$  grows nonelementarily in  $n$  [28, Theorem 7.2.2], and by [20],  $\text{RCA}_0 + \text{RT}_2^2$  has no superpolynomial proof speedup over  $I\Sigma_1$  w.r.t. proofs of  $\Pi_3$  sentences. Thus, the size of the smallest proof of  $\text{Con}_{\exp_n(2)}(I\Sigma_1)$  in  $\text{RCA}_0 + \text{RT}_2^2$  also grows nonelementarily in  $n$ . Since  $B\Sigma_2 + (\neg I\Sigma_2 \Rightarrow \Delta_2\text{-RT}_2^2)$  is axiomatized by a single sentence, and  $\text{RCA}_0 + \text{RT}_2^2$  proves  $B\Sigma_2$ , it follows that it cannot prove  $\neg I\Sigma_2 \Rightarrow \Delta_2\text{-RT}_2^2$ .  $\square$

Thus, the statement  $\neg I\Sigma_2 \Rightarrow \Delta_2\text{-RT}_2^2$  cannot be used to witness the potential nonconservativity of  $\text{RT}_2^2$  over  $B\Sigma_2$ . However, our argument for this, in addition to being somewhat roundabout, made use of the fact that  $\text{RCA}_0^* + \text{RT}_2^2$  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  $\neg I\Sigma_2 \Rightarrow \Delta_2\text{-RT}_2^2$ , for instance to statements in which  $\text{RT}_2^2$  is replaced by a restriction to colourings for which finite Ramsey numbers are polynomial.

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

*Question 6.* Does  $\text{RCA}_0 + \text{RT}_2^2$  prove one of the following the  $\Pi_5$  statements:

- (a)  $\neg I\Sigma_2 \Rightarrow \Delta_2\text{-CAC}$ : if  $\neg I\Sigma_2$ , then every  $\Delta_2$ -definable partial order on  $[\mathbb{N}]$  contains an infinite  $\Delta_2$ -definable chain or an infinite  $\Delta_2$ -definable antichain”,
- (b) “if  $\neg I\Sigma_2$ , then for every  $\Delta_1$ -definable 2-colouring of  $[\mathbb{N}]^2$  there is a  $\Delta_2$ -definable infinite homogeneous set”?

Does  $\text{RCA}_0 + B\Sigma_2^0$  prove the statement in (b)?

The fact that  $\text{RCA}_0 + B\Sigma_2^0$  does not prove the statement  $\neg I\Sigma_2 \Rightarrow \Delta_2\text{-CAC}$  appearing in (a) can be shown using the methods of [11].

**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 feel that way himself.

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, C. T. Chong, Wei Wang, Tin Lok Wong, and Yue Yang. Where pigeonhole principles meet König lemmas. *Trans. Amer. Math. Soc.*, 374(11):8275–8303, 2021.
- [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  $\Sigma_n$  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  $\text{WKL}_0^*$ . *Ann. Pure Appl. Logic*, 168(6):1247–1283, 2017.
- [11] Marta Fiori-Carones, Leszek Aleksander Kołodziejczyk, and Katarzyna W. Kowalik. Weaker cousins of Ramsey’s theorem over a weak base theory. *Ann. Pure Appl. Logic*, 172(10):Paper No. 103028, 22 pages, 2021.
- [12] Marcia J. Groszek and Theodore A. Slaman. On Turing reducibility. Preprint, 1994.
- [13] Ian Robert Haken. *Randomizing Reals and the First-Order Consequences of Randoms*. PhD thesis, UC Berkeley, 2014.
- [14] Kostas Hatzikiriakou. Algebraic disguises of  $\Sigma_1^0$  induction. *Arch. Math. Logic*, 29(1):47–51, 1989.
- [15] Denis R. Hirschfeldt. *Slicing the truth. On the computable and reverse mathematics of combinatorial principles*. World Scientific, 2015.
- [16] Carl G. Jockusch. Ramsey’s theorem and recursion theory. *J. Symb. Log.*, 37(2):268–280, 1972.

- [17] Richard Kaye. Model-theoretic properties characterizing Peano arithmetic. *J. Symb. Log.*, 56(3):949–963, 1991.
- [18] Richard Kaye. *Models of Peano Arithmetic*. Oxford University Press, 1991.
- [19] Richard Kaye. Constructing  $\kappa$ -like models of arithmetic. *J. London Math. Soc. (2)*, 55(1):1–10, 1997.
- [20] 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](https://arxiv.org/abs/2005.06854).
- [21] Leszek Aleksander Kołodziejczyk and Keita Yokoyama. Categorical characterizations of the natural numbers require primitive recursion. *Ann. Pure Appl. Logic*, 166(2):219–231, 2015.
- [22] 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.
- [23] Roman Kossak. On extensions of models of strong fragments of arithmetic. *Proc. Amer. Math. Soc.*, 108(1):223–232, 1990.
- [24] 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.
- [25] Henryk Kotlarski. On elementary cuts in recursively saturated models of Peano arithmetic. *Fund. Math.*, 120(3):205–222, 1984.
- [26] Hamid Lessan. *Models of arithmetic*. PhD thesis, University of Manchester, 1978.
- [27] Ludovic Patey and Keita Yokoyama. The proof-theoretic strength of Ramsey’s theorem for pairs and two colors. *Adv. Math.*, 330:1034–1070, 2018.
- [28] Pavel Pudlák. The lengths of proofs. In S. R. Buss, editor, *Handbook of Proof Theory*, pages 547–642. Elsevier, 1998.
- [29] David Seetapun and Theodore A. Slaman. On the strength of Ramsey’s theorem. *Notre Dame J. Form. Log.*, 36(4):570–582, 1995.
- [30] Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Association for Symbolic Logic, 2009.
- [31] Stephen G. Simpson and Rick L. Smith. Factorization of polynomials and  $\Sigma_1^0$  induction. *Ann. Pure Appl. Logic*, 31(2-3):289–306, 1986.
- [32] Stephen G. Simpson and Keita Yokoyama. Reverse mathematics and Peano categoricity. *Ann. Pure Appl. Logic*, 164(3):284–293, 2012.
- [33] 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.
- [34] Keita Yokoyama. On the strength of Ramsey’s theorem without  $\Sigma_1$ -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  $\mathbb{I}\Sigma_1$  proves:*

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

where  $\text{Con}_x(T)$  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  $\psi$  atomic, as well as analogous cedents corresponding to the equality axioms (the need to allow the arbitrary set of formulas  $\Gamma$  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:

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

where in the  $(\exists)$  rule  $t$  must be a term that is substitutable for  $w$  in  $\psi$ , 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  $\mathbb{I}\Sigma_1$  is axiomatized by finitely many sentences  $\gamma_1, \dots, \gamma_n$ , where each  $\gamma_i$  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  $\delta_i$  bounded. (Using a different finite axiomatization would shorten proofs in  $\mathbb{I}\Sigma_1$  by at most a constant additive factor, and using the typical axiomatization of  $\mathbb{I}\Sigma_1$  as a scheme would shorten proofs at most polynomially.)

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

$$\neg \gamma_1, \dots, \neg \gamma_n,$$

of size at most  $\exp_{p(x)}(2)$ . Working in  $\mathbb{I}\Sigma_1$ , let  $k$  be such that every infinite  $\Delta_2$ -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  $\neg \gamma_1, \dots, \neg \gamma_n$  of size at most  $m$ , which will imply  $\text{Con}_k(\mathbb{I}\Sigma_1)$ .

Assume to the contrary that there is such a cut-free proof, and let the lines of the proof be  $C_1, \dots, C_\ell$ ; note that  $\ell \leq m$ . For each  $j = 1, \dots, \ell$ , let the *negations* of the formulas in  $C_j$  be  $\xi_{j,1}, \dots, \xi_{j,r_j}$ . Note that  $r_\ell = n$ , each  $\xi_{\ell,i}$  is  $\gamma_i$ , and, by the subformula property of cut-free proofs, each  $\xi_{j,r}$  is a subformula

of one of the  $\psi_i$ 's. As usual in such a context, we regard  $\varphi(t)$  as a subformula of  $Qx \varphi$  for  $Q$  a quantifier.

Define an infinite sequence of numbers by:

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

Let  $D$  consist of all numbers that appear as some  $d_j$ . Note that provably in  $\text{I}\Sigma_1$ , both  $D$  and the complement of  $D$  are  $\Sigma_2$ -definable, so  $D$  is a  $\Delta_2$ -set, and  $D$  is infinite. By our assumption, there exists an  $\ell$ -element finite subset of  $D$ . W.l.o.g., we may assume that the elements of this subset are  $d_0, \dots, d_{\ell-1}$ .

We claim that the following statement  $\eta(s)$  can be proved by  $\Pi_1$  induction on  $s = 0, \dots, \ell - 1$ :

“there exist  $j \leq \ell - s$  and an assignment  $\alpha$  of values  $\leq d_s$  to the free variables in  $C_j$  such that, for every  $\xi_{j,r}$  that is  $\Sigma_2$ , 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  $\alpha$  makes the  $\Pi_1$  part of  $\xi_{j,r}$  satisfied.”

Note that  $\eta(s)$  is indeed a  $\Pi_1$  statement (provably in  $\text{B}\Sigma_1 + \text{exp}$ ), because all the quantifiers preceding the definition of satisfaction for  $\Pi_1$  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  $\Delta_0$ ) principal formula of the inference used to derive  $C_j$ . Take the assignment  $\alpha$  witnessing  $\eta(s)$  at  $j$ , and let  $j' < j$  be such that  $C_j$  is  $\Gamma, \psi_b$  for  $b \in \{1, 2\}$  such that  $\alpha$  satisfies  $\neg\psi_b$ . This  $j'$  and the unchanged assignment  $\alpha$  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  $\alpha$  witnessing  $\eta(s)$  at  $j$  to an assignment  $\alpha'$  by letting all variables that are free in  $C_{j'}$  but not in  $C_j$  have value 0. If  $\psi(t)$  is not  $\Pi_2$  (in which case  $\neg\psi(t)$  is a  $\Pi_3$  but not  $\Sigma_2$  subformula of one of the induction axioms  $\gamma_i$ ) or  $\exists w \psi$  is  $\Sigma_1$  (in which case  $\forall w \neg\psi$  is  $\Pi_1$  and satisfied under  $\alpha$ , so  $\neg\psi(t)$  is satisfied under  $\alpha'$ ), this is all we need to do in order to ensure that  $j', \alpha'$  witness  $\eta(s + 1)$ . The remaining case is when  $\psi(t)$  is  $\Pi_2$  but  $\exists w \psi$  is not. In that situation,  $\neg\psi(t)$  arises from one of the  $\gamma_i$ '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  $\alpha'$  satisfies  $\neg\psi(t)$  (because  $\gamma_i$  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  $\alpha$  be an assignment witnessing  $\eta(s)$  at  $j$ . There are two subcases to consider, depending on whether  $\exists w \neg \psi$  is an unbounded  $\Sigma_2$  formula or a  $\Delta_0$  formula. In the former case, we know from the inductive assumption that  $\alpha$  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  $\alpha$  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  $\alpha$ , 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.  $\square$