Reverse mathematics of a color basis theorem

Caleb Davis, Jeffry Hirst, Silva Keohulian,
Brody Miller, Jessica Ross

November 12, 2025

Abstract

The infinite pigeonhole theorem asserts that if $f : \mathbb{N} \to m$ is a function with a finite range, then there is a $j < m$ such that the set $\{n \in \mathbb{N} \mid f(n) = j\}$ is infinite. This article uses the techniques of reverse mathematics and Weihrauch analysis to examine the strength of a theorem that finds all the values that occur infinitely often in the range of a function.

For a function $f : \mathbb{N} \to m$ with a finite range, the *color basis* for f is the set $B \subseteq [0, m)$ such that $c \in B$ if and only if c appears infinitely often in the range. More formally, $B = \{c < m \mid \forall b \exists n (n > b \land f(n) = c)\}$. The next section examines the strength of the existence of color bases in reverse mathematics. The following three sections extend the examination via Weihrauch analysis and higher order reverse mathematics. Preliminary versions of these results were presented at RaTLoCC 2024 [13] under the title of pigeonhole basis theorems. The terminology has been changed to avoid confusion with computational basis results by Monin and Patey [11].

1 Reverse mathematics: Induction and comprehension

The study of reverse mathematics is founded on a hierarchy of subsystems of second order arithmetic, described in detail in the texts of Dzhafarov and Mummert [5] and Simpson [14]. The base system RCA_0 includes induction restricted to Σ^0_1 formulas and a set existence axiom for computable sets (formalized by Δ^0_1 definability). As a consequence of the restriction on induction,

RCA_0 cannot prove the Π^0_1 bounding scheme, defined by

$$\mathsf{B}\Pi^0_1 : (\forall x < a)(\exists y)(\forall z)\theta(x, y, z) \to (\exists b)(\forall x < a)(\exists y < b)(\forall z)\theta(x, y, z)$$

where θ is a Σ^0_0 formula. Indeed, over RCA_0 there is a strict hierarchy of bounding and induction schemes, with $\mathsf{I}\Sigma^0_n$ weaker than $\mathsf{B}\Pi^0_n$ weaker than $\mathsf{I}\Sigma^0_{n+1}$ for all n. See Chapter 6 of Dzhafarov and Mummert [5] for details. The following theorem relates $\mathsf{B}\Pi^0_1$ to the infinite pigeonhole principle, often called Ramsey's theorem for singletons.

Theorem 1. (RCA_0) *The following are equivalent:*

(1) $\mathsf{B}\Pi^0_1$.

(2) RT^1: *If $f : \mathbb{N} \to m$ then for some $j < m$, the set $\{n \mid f(n) = j\}$ is infinite.*

The proof of Theorem 1 appeared initially in Hirst's thesis [7], but is more readily accessible in the texts of Dzhafarov and Mummert [5] (Theorem 6.5.1) and Weber [15] (Theorem 9.5.1). While RT^1 ensures that the color basis for a function is not empty, over RCA_0 the existence of the color basis is strictly stronger, as shown by the following theorem.

Theorem 2. (RCA_0) *The following are equivalent:*

(1) CB: *Every $f : \mathbb{N} \to m$ has a color basis.*

(2) $\mathsf{I}\Sigma^0_2$: *Induction restricted to Σ^0_2 formulas.*

Proof. Working in RCA_0, by Exercise II.3.13 of Simpson [14], the induction scheme $\mathsf{I}\Sigma^0_2$ is equivalent to bounded Π^0_2 comprehension. Recall that the color basis of f is defined by $B = \{c < m \mid \forall b \exists n(n > b \wedge f(n) = c)\}$, which is a bounded Π^0_2 set. Thus item (1) follows from item (2).

To show the converse, suppose $m \in \mathbb{N}$ and $\theta(c, b, n)$ is a Σ^0_0 formula. Our goal is to use CB to prove that the set $\{c < m \mid \forall b \exists n \theta(c, b, n)\}$ exists. Using a bijection identifying triples (c, b, n) in $m \times \mathbb{N} \times \mathbb{N}$ with integer codes, define $f : \mathbb{N} \to m + 1$ by

$$f(c, b, n) = \begin{cases} c \text{ if } n \text{ is the least } t \leq n \text{ such that } (\forall j \leq b)(\exists k \leq t)\theta(c, j, k) \\ m \text{ otherwise.} \end{cases}$$

Recursive comprehension proves the existence of f. Note that for a fixed c_0, if $\forall b \exists n \theta(c_0, b, n)$, then RCA_0 proves that for each b there is a unique least t such that $(\forall j \leq b)(\exists k \leq t)\theta(c_0, j, k)$. In this situation, c_0 appears in the range of f once for each value of b, and so c_0 is in the color basis for f. On the other hand, for any fixed c_1 satisfying $\neg\forall b \exists n \theta(c_1, b, n)$, if b_1 witnesses $\forall n \neg\theta(c_1, b_1, n)$, then c_1 appears in the range of f no more than b_1 times. In this situation, c_1 is not in the color basis for f. Summarizing, the values less than m that are in the color basis for f are exactly the set $\{c < m \mid \forall b \exists n \theta(c, b, n)\}$ as desired. $\square$

At RaTLoCC 2024 [13], Professor Schnoebelen (LSV, CNRS, ENS Paris-Saclay) asked if requiring the color bases of item (1) of Theorem 2 to be nonempty would affect the reverse mathematical strength. Interestingly, the strength of item (1) is unchanged by this revision. The scheme $\mathsf{I}\Sigma^0_2$ implies $\mathsf{B}\Pi^0_1$, so item (2) implies the revised item (1). The converse follows immediately from the given proof.

In light of known results on reverse mathematics of matroids, the connection of the color basis theorem and $\mathsf{I}\Sigma^0_2$ is not so surprising. Matroids capture the fundamental notions of basis and dimension in a combinatorial setting. Theorem 5 of Hirst and Mummert's article [8] shows the equivalence of a matroid basis theorem and $\mathsf{I}\Sigma^0_2$. Informally, a matroid resembles the vectors in a vector space, and an e-matroid as defined below is an enumeration of dependent sets.

Definition. An e-matroid is a pair (M, e) consisting of a non-empty set M and a function $e : \mathbb{N} \to [M]^{<\mathbb{N}}$ enumerating the finite dependent subsets of M. The enumeration e satisfies the following conditions:

(1) The empty set is independent. Formally, $\forall n (e(n) \neq \emptyset)$.

(2) Finite supersets of dependent sets are dependent. Formally,

$$(\forall n)(\forall Y \in M^{<\mathbb{N}})(e(n) \subseteq Y \to \exists m(e(m) = Y)).$$

(3) (Exchange principle) If X and Y are independent with $|X| < |Y|$, then Y contains an element that is independent of X. That is, if X and Y are independent and $|X| < |Y|$, then $(\exists y \in Y)(\forall n)(e(n) \neq X \cup \{y\})$.

The set M is often used as a shorthand for the matroid (M, e). A finite set B spans M if every proper extension is dependent. Formally, B *spans* M means

$$(\forall x \in M)(x \notin B \to (\exists n)(e(n) = B \cup \{x\})).$$

A finite subset B is a *basis* for M if B spans M and B is independent.

The e-matroid terminology can be used to add another equivalence to Theorem 2.

Theorem 3. (RCA_0) *The following are equivalent:*

(1) EMB: *If there is a bound b for the dimension of an e-matroid (M, e), that is, if every set of size greater than b is dependent, then M has a finite basis.*

(2) CB: *Every $f : \mathbb{N} \to m$ has a color basis.*

(3) $\mathsf{I}\Sigma^0_2$: *Induction restricted to Σ^0_2 formulas.*

Proof. The shortest proof is to note that Theorem 2 shows the equivalence of CB and $\mathsf{I}\Sigma^0_2$, and Theorem 5 of Hirst and Mummert [8] shows the equivalence of EMB and $\mathsf{I}\Sigma^0_2$. $\square$

Of course, direct proofs of the equivalence of the first two items of Theorem 3 are possible. In particular, see the comment following the proof of Lemma 10 below.

The subsystem $\mathsf{ACA_0}$ includes a set comprehension axiom that asserts the existence of arithmetically definable sets. Many results in reverse mathematics prove equivalences between familiar mathematical theorems and $\mathsf{ACA_0}$. Finding color bases for sequences of functions yields such a result.

Theorem 4. (RCA_0) *The following are equivalent:*

(1) $\mathsf{ACA_0}$.

(2) *If $\langle f_i \rangle_{i \in \mathbb{N}}$ is a sequence of functions with finite ranges, then there is a function $g : \mathbb{N} \to \mathbb{N}$ such that for every n, $g(n)$ is (the code for) the color basis for f_n.*

(3) *If $\langle f_i \rangle_{i \in \mathbb{N}}$ is a sequence of functions from $\mathbb{N}$ to $\{0, 1\}$, then there is a function $g : \mathbb{N} \to \mathbb{N}$ such that for every n, $g(n)$ is (the code for) the color basis for f_n.*

Proof. We work in RCA_0 throughout. To prove that item (1) implies item (2), assume ACA_0 and let $\langle f_i \rangle_{i \in \mathbb{N}}$ satisfy the hypotheses of item (2). Then for each i, there is a unique (code for a) finite set B_i which is a color basis for f_i. The set B_i satisfies the arithmetical formula

$$j \in B_i \leftrightarrow \forall m \exists n (m < n \wedge f_i(n) = j).$$

Thus arithmetical comprehension suffices to prove the existence of the function g which maps each i to (the code for) B_i.

Item (3) is a special case of item (2), so the proof can be completed with a proof of item (1) from item (3). By Lemma III.1.3 of Simpson [14], it suffices to use item (3) to find the range of an injection $h : \mathbb{N} \to \mathbb{N}$. For each i, define f_i by:

$$f_i(n) = \begin{cases} 0 \text{ if } (\forall t \leq n)(h(t) \neq i) \\ 1 \text{ if } (\exists t \leq n)(h(t) = i). \end{cases}$$

The existence of the sequence $\langle f_i \rangle_{i \in \mathbb{N}}$ is provable in RCA_0. The color basis for f_i is $\{0\}$ if i is not in the range of h, and $\{1\}$ if i is in the range of h. Apply item (3) to find a function g such that $g(i)$ is the color basis for f_i for all i. Then the range of h is $\{i \in \mathbb{N} \mid g(i) = \{1\}\}$, and exists by recursive comprehension. □

2 Weihrauch Analysis

This section uses Weihrauch analysis to examine the color basis theorem. Introductions to Weihrauch analysis can be found in the texts of Weihrauch [16] and Dzhafarov and Mummert [5], and the works of Brattka and Gherardi e.g. [1]. The articles of Dorais et al [4] and Brattka and Rakotoniaina [3] include Weihrauch analysis of many problems related to RT^1.

Adapting the notation of Brattka and Rakotoniaina [3], we denote the Weihrauch problem related to the color basis principle by CB_+. An instance of the problem CB_+ is a pair (f, m) where m is a natural number and $f : \mathbb{N} \to m$. The solution for the problem is (the integer code for) the color basis for f. The notation CB_k is used to denote the color basis problem restricted to colorings of the form $f : \mathbb{N} \to k$ for a fixed value of k. Similarly, an instance of the Weihrauch problem EMB_+ is a triple (M, e, b) where (M, e) is an e-matroid in which every set of size $b + 1$ is dependent, and the solution is (an integer code for) a basis of (M, e). This is identical to the Weihrauch problem $\mathsf{EMB}_{<\omega}$ studied by Hirst and Mummert [8]. The notation EMB_k will be used for the restriction of the problem to a fixed bound k.

A realizer for a Weihrauch problem is a function that inputs instances of the problem and outputs solutions. Because instances can have many solutions, realizers are not unique. If P and Q are Weihrauch problems, we say P is (weakly) Weihrauch reducible to Q and write $P \leq_W Q$ if there is a computable preprocessing procedure Φ and a computable postprocessing procedure Ψ such that for any realizer R_Q for problem Q, the composition $\Psi(R_Q(\Phi(f)), f)$ is a realizer for P. Informally, Φ converts any instance f of the problem P into an instance of Q, and Ψ converts any solution of $\Phi(f)$ into a solution for f, referring to f in the conversion, if necessary.

Our results will relate color basis and matroid basis problems to two widely studied problems. The limited principle of omniscience problem, denoted LPO, accepts inputs of the form $f : \mathbb{N} \to 2$, outputs 1 if the range of f contains no zeros, and outputs 0 if zero is in the range of f. The choice principle $\mathsf{C}_\mathbb{N}$ accepts as input a function $f : \mathbb{N} \to \mathbb{N}$ that is not onto, and outputs an integer not appearing in the range of f.

Section 2 of Brattka and Rakotoniaina [3] lists a number of operators that are commonly used to combine Weihrauch problems. For example, if P is a problem, $P \times P$ denotes the problem of solving two parallel instances of P. The k-fold product is denoted by P^k, and P^* denotes the problem corresponding to P^k for arbitrary finite values of k. If P accepts a function f as an input, then the problem P' accepts sequences of functions $\langle f_n \rangle_{n\in\mathbb{N}}$ as input and outputs the result of applying P to the limit function $\lim_n f_n$. This jump operator appears in the paper of Brattka, Gherardi, and Marcone [2].

The authors thank one of the anonymous referees for suggestions that sharpened and streamlined the results of this section. Motivated by those suggestions, the next five results yield a Weihrauch analysis of CB_+.

Lemma 5. $\mathsf{CB}_2 \leq_W \mathsf{LPO}' \times \mathsf{LPO}'$.

Proof. Suppose $f : \mathbb{N} \to 2$ is a coloring. For $b < 2$, define a sequence of functions $\langle g^b_n \rangle_{n\in\mathbb{N}}$ such that $g^b_n(t) = 0$ if either $t = 0$ and $f(j) \neq b$ for all $j < n$ or $f(t) = b$ and t is the largest such value less than n, and $g^b_n(t) = 1$

otherwise. More formally,

$$g_n^b(t) = \begin{cases} 0 \text{ if } t = 0 \wedge (\forall s < n)(f(s) \neq b) \\ 0 \text{ if } t < n \wedge f(t) = b \wedge (\forall s \in (t, n])(f(s) \neq b) \\ 1 \text{ otherwise.} \end{cases}$$

If b appears infinitely often in the range of f, then $\mathsf{LPO}(\lim_n g_n^b) = 1$. Otherwise, $\mathsf{LPO}(\lim_n g_n^b) = 0$. Thus, $\mathsf{LPO}(\lim_n g_n^0)$ determines whether or not 0 is in the color basis for f and $\mathsf{LPO}(\lim_n g_n^1)$ decides the membership of 1. □

Lemma 6. $\mathsf{LPO}' \leq_W \mathsf{CB}_2$.

Proof. Let $\langle f_n \rangle_{n\in\mathbb{N}}$ be an input for LPO'. We compute a coloring $g : \mathbb{N} \to 2$ as follows. Let $\langle (j_n, k_n) \rangle_{n\in\mathbb{N}}$ be an enumeration of $\mathbb{N} \times \mathbb{N}$. Compute a binary sequence as follows. Place a marker at (j_0, k_0) and add a one to the sequence. If $f_{j_0}(k_0) = 0$, add a zero to the sequence and increment to consider $f_{j_0+1}(k_0)$. Continue adding zeros and incrementing until a one is encountered. If a one is encountered, move the marker to (j_1, k_1) and add a one to the sequence. Continue adding zeros and incrementing the marker location in this fashion. Let $g : \mathbb{N} \to 2$ be an enumeration of this binary sequence. Note that if there is a j such that $\lim_k f_k(j) = 0$, the range of g has only finitely many ones. On the other hand, if $\lim_k f_k(j) = 1$ for all j, the marker is moved infinitely often and the range of g will have infinitely many ones. Thus the color basis of g computes $\mathsf{LPO}(\lim_n f_n)$. □

Lemma 7. $\mathsf{CB}_m \leq_W \mathsf{CB}_2^m \leq_W \mathsf{CB}_{2m}$.

Proof. To show that $\mathsf{CB}_m \leq_W \mathsf{CB}_2^m$, suppose $f : \mathbb{N} \to m$ is an input for CB_m. For $j < m$, define $g_j : \mathbb{N} \to 2$ by $g_j(t) = 1$ if $f(t) = j$ and $g_j(t) = 0$ if $f(t) \neq j$. Then j is in the color basis of f if and only if 1 is in the color basis of g_j. The sequence of color bases for $\langle g_j \rangle_{j<m}$ computes the color basis for f.

To see that $\mathsf{CB}_2^m \leq_W \mathsf{CB}_{2m}$, suppose that $\langle g_j \rangle_{j<m}$ is a sequence of m instances of CB_2. Every natural number has a unique representation of the form $mn+j$ for some n and $j < m$. Define $f : \mathbb{N} \to 2m$ by setting $f(mn+j) = 2j + g_j(n)$. For each $j < m$ and $i < 2$, i is in the color basis of g_j if and only if $2j + i$ is in the color basis of f. Thus the color bases for $\langle g_j \rangle_{j<m}$ can be computed from the color basis for f. □

Theorem 8. $\mathsf{CB}_+ \equiv_W (\mathsf{LPO}')^*$.

Proof. By Lemma 7, $\mathsf{CB}_+ \leq_W \mathsf{CB}_2^*$. By Lemma 5, $\mathsf{CB}_2 \leq_W \mathsf{LPO}' \times \mathsf{LPO}'$, so $\mathsf{CB}_2^* \leq_W (\mathsf{LPO}' \times \mathsf{LPO}')^* \equiv_W (\mathsf{LPO}')^*$. By transitivity, $\mathsf{CB}_+ \leq_W (\mathsf{LPO}')^*$.

By Lemma 6, $\mathsf{LPO}' \leq_W \mathsf{CB}_2$, so $(\mathsf{LPO}')^* \leq_W \mathsf{CB}_2^*$. By Lemma 7, $\mathsf{CB}_2^m \leq_W \mathsf{CB}_{2m}$, so $\mathsf{CB}_2^* \leq_W \mathsf{CB}_+$. By transitivity, $(\mathsf{LPO}')^* \leq_W \mathsf{CB}_+$. □

While not a very sharp inequality, the following corollary provides an interesting contrast to Theorem 13 in the next section.

Corollary 9. $\mathsf{LPO} <_W \mathsf{CB}_2$.

Proof. By Lemma 6, $\mathsf{LPO}' \leq_W \mathsf{CB}_2$. Constant sequences of inputs are inputs for jumps, so $\mathsf{LPO} \leq_W \mathsf{LPO}'$. Using the problem of finding limits of binary sequences as an intermediary, it is not hard to show that $\mathsf{LPO}' \not\leq_W \mathsf{LPO}$, so $\mathsf{LPO} <_W \mathsf{LPO}'$. Concatenating the reductions yields $\mathsf{LPO} <_W \mathsf{CB}_2$. □

The next three results relate the color basis and matroid basis problems.

Lemma 10. $\mathsf{CB}_+ \leq_W \mathsf{EMB}_+$.

Proof. The preprocessing procedure for an instance (f, m) of CB_+ consists of two steps. First, define $f' : \mathbb{N} \to m$ by $f'(j) = j$ for $j < m$ and $f'(j) = f(j)$ for $j \geq m$. Note that the range of f' includes all of $[0, m)$ and the color basis of f' matches that of f. Second, compute an instance of EMB_+ for f'. Let $h : \mathbb{N} \to \mathbb{N}^{<\mathbb{N}}$ computably enumerate the finite subsets of $\mathbb{N}$, repeating each subset infinitely often. Define the matroid $(\mathbb{N}, e)$ as follows. For each n, suppose $h(n) = \{x_0, \ldots, x_k\}$. If f' assigns the same value to two elements of $h(n)$, or if for some $x_j \in h(n)$ there is a $t \leq n$ such that $t > x_j$ and $f(t) = f(x_j)$, then set $e(n) = h(n)$, otherwise, set $e(n) = \{m\}$. The desired instance of EMB_+ is $(\mathbb{N}, e, m)$.

Now we will describe the postprocessing procedure. If S is any independent set for the matroid $(\mathbb{N}, e)$ and $s \in S$, then s is the largest number for which f' takes the value $f'(s)$. Let B be a basis for $(\mathbb{N}, e)$. The set $\{f'(x) \mid x \in B\}$ contains exactly those values in the range of f' which appear finitely often in the range of f'. Because f' is onto $[0, m)$, $B' = \{j < m \mid (\forall x \in B) f'(x) \neq j\}$ is the color basis for f' and thus for f. □

The proof of Lemma 10 can easily be formalized in RCA_0, providing a direct proof of one direction of Theorem 3. Our original reverse mathematics proof (not presented here) applied the preprocessing procedure to the function f, using bounded comprehension in the postprocessing stage to delete the values not in the range of f from the complement of the image of the matroid basis. The application of bounded comprehension barred a conversion of that proof to a Weihrauch reduction, so the use of f' was added to the preceding proof to address this issue.

The following result is a slight extension of the second half of Proposition 7.8 of Brattka and Rakotoniaina [3]. We adapt their proof.

Lemma 11. $\mathsf{C}_{\mathbb{N}} \equiv_W \mathsf{EMB}_1 \not\leq_W \mathsf{CB}_+$.

Proof. The first equivalence follows from the fact that the bases of a one dimensional matroid are the single element sets not listed by the matroid's enumeration. To prove the negation of the inequality, suppose by way of contradiction that $\mathsf{C}_{\mathbb{N}} \leq_W \mathsf{CB}_+$. By Fact 3.2(3) and Proposition 2.6 of Brattka, Gherardi, and Marcone [2], there is a k such that $\mathsf{C}_{\mathbb{N}} \leq_W \mathsf{CB}_k$. The range of CB_k consists of the nonempty subsets of k and so has cardinality $2^k - 1$. By Proposition 7.3 of Brattka and Rakotoniaina [3], we have $|\mathbb{N}| \leq 2^k - 1$, a contradiction. □

Theorem 12. $\mathsf{CB}_+ <_W \mathsf{EMB}_+$.

Proof. By Lemma 10, $\mathsf{CB}_+ \leq_W \mathsf{EMB}_+$. Because $\mathsf{EMB}_1 \leq_W \mathsf{EMB}_+$, the reduction $\mathsf{EMB}_+ \leq_W \mathsf{CB}_+$ would imply $\mathsf{EMB}_1 \leq_W \mathsf{CB}_+$, contradicting Lemma 11. Thus $\mathsf{CB}_+ <_W \mathsf{EMB}_+$. □

We have shown that the reverse mathematical equivalence of Theorem 3 is not replicated in the Weihrauch setting.

3 Higher order reverse mathematics

Reverse mathematics can be extended from numbers and sets of numbers to higher types, such as functions from sets to numbers or from sets to sets. A base theory RCA_0^ω and early results are presented in Kohlenbach's article [10]. This framework has been used in many articles by Normann and Sanders and by Hirst and Mummert (e.g. [12] and [9]). With the more expressive language, principles can be formulated asserting the existence of realizers for Weihrauch problems. For example, in the next theorem, the principle (LPO) asserts the existence of a realizer for the Weihrauch problem LPO. Over RCA_0^ω, (LPO) is identical to Kohlenbach's principle ($\exists^2$), which is related to Kleene's functional $\mathsf{E2}$.

Theorem 13. (RCA_0^ω) *The following are equivalent:*

(1) (LPO) *There is a functional* LPO *such that for all* $f : \mathbb{N} \to 2$*,* $\mathsf{LPO}(f) = 0$ *if and only if* $\exists t(f(t) = 0)$*. This principle is sometimes denoted* ACA_0^ω*.*

(2) (CB_2) *There is a function* CB_2 *such that for all* $f : \mathbb{N} \to 2$*,* $\mathsf{CB}_2(f)$ *is the color basis of* f*.*

Proof. To prove that item (2) implies item (1), note that RCA_0^ω proves that there is a function PRE such that for all $f : \mathbb{N} \to 2$, $\mathsf{PRE}(f)$ is a function that is constantly 1 until a zero appears in the range of f and constantly 0 afterwards. The function $\mathsf{LPO}(f)$ is the element appearing in $\mathsf{CB}_2(\mathsf{PRE}(f))$.

The underlying idea of the proof that item (1) implies (2) is that given the LPO function, RCA_0^ω can iterate it. Suppose (LPO) holds. Let $f : \mathbb{N} \to 2$ be an input for CB_2. Define the function $Z(f, n)(k)$ by setting $Z(f, n)(k) = 1$ unless k is the n^{th} number where f equals 0, in which case $Z(f, n)(k) = 0$. Note that f has at least n zeros if and only if $\mathsf{LPO}(Z(f, n)) = 0$. If f has finitely many zeros, then for all values n larger than some bound m, $\mathsf{LPO}(Z(f, n)) = 1$. The function $g(f, n) = 1 - \mathsf{LPO}(Z(f, n))$ has zeros in its range if and only if f has only finitely many zeros. Thus the function $Z'(f) = \mathsf{LPO}(g(f, n))$ takes the value 0 if f has finitely many zeros in its range and 1 if f has infinitely many zeros. Define a similar function $U'(f)$ that counts ones, so that $U'(f) = 0$ if f has finitely many ones in its range and 1 if f has infinitely many ones. The function $B(f)$ defined by

$$B(f) = \begin{cases} \{0\} & \text{if } U'(f) = 0 \wedge Z'(f) = 1 \\ \{1\} & \text{if } U'(f) = 1 \wedge Z'(f) = 0 \\ \{0, 1\} & \text{if } U'(f) = 1 \wedge Z'(f) = 1 \end{cases}$$

finds the color basis for f. $\square$

While Corollary 9 shows that the Weihrauch problems CB_2 and LPO are not Weihrauch equivalent, Theorem 13 shows that the related higher order principles (CB_2) and (LPO) are provably equivalent over RCA_0^ω. In this case, the fact that the higher order functionals can be applied sequentially makes them behave like the parallelized versions of the Weihrauch problems, which can be shown to be Weihrauch equivalent.

4 Additional equivalences

In this section, we examine two more problems that are Weihrauch equivalent to EMB_+. Both correspond to statements that are equivalent to $\mathsf{I}\Sigma^0_2$ in the reverse mathematics setting. Thus they are equivalent to the color basis problem in the reverse mathematics setting and strictly stronger in the Weihrauch setting.

The first problem is graph theoretic. Here graphs are represented by a set of vertices and a set of undirected edges, where each edge is a pair of vertices. The vertices v_0 and v_n lie in the same connected component if there is a path $v_0, v_1, \dots v_n$ such that for each i, (v_i, v_{i+1}) is an edge of G. An instance of the Weihrauch problem GAC_+ is a triple (V, E, n) consisting of a graph with vertices V and edges E with at most n distinct connected components. A solution of the problem is a set of vertices consisting of exactly one vertex from each connected component. The notation GAC_+ stands for *G*raph *A*nti*C*hain, where vertices are comparable if they lie in the same connected component. This terminology matches that of Hirst and Mummert [8].

The second problem concerns finite partitions of sets. A sequence of functions $\langle e_i \mid i \in I\rangle$ is an enumerated partition of a set S if (1) for every $s \in S$ there are values i and m such that $e_i(m) = s$, and (2) if $e_i(m) = e_j(n)$ for some i, j, m, and n, then $\forall m \exists n (e_i(m) = e_j(n))$. Informally, the functions e_i enumerate the disjoint cells in a partition of S. Cells may be enumerated by more than one function, in varying orders. Every element of S is contained in some cell.

An instance of the partition problem P_+ is a triple $(S, \langle e_i \mid i \in I\rangle, n)$ where the set S is partitioned by $\langle e_i \mid i \in I\rangle$ and the partition has at most n cells. The solution is a set of indices that include exactly one enumeration for each cell of the partition. The problem P_+ can be thought of as choosing one vertex from each edge of a hypergraph with finitely many disjoint edges, where each edge is enumerated rather than being presented as a set.

Results about graphs with infinitely many connected components and partitions with infinitely many cells can be found in the article of Gura, Hirst, and Mummert [6]. The following theorem adds the partition problem to the list of Weihrauch equivalences in Theorem 17 of Hirst and Mummert [8].

Theorem 14. $\mathsf{GAC}_+ \equiv_W \mathsf{P}_+ \equiv_W \mathsf{EMB}_+$.

Proof. To see that $\mathsf{GAC}_+ \leq_W \mathsf{P}_+$, suppose (V, E, n) is an input for GAC_+. Compute an associated partition problem by defining $e_v : \mathbb{N} \to V$ by $e_v(t) = v'$ if t codes a path from v to v' and $e_v(t) = v$ otherwise. Letting Φ denote this preprocessing computation, $\Phi(V, E, n)$ is the partition problem $(V, \langle e_v \mid v \in V\rangle, n)$. Any solution of this partition problem will consists of exactly one vertex from each connected component of the original graph, so the postprocessing computation is trivial.

To see that $\mathsf{P}_+ \leq_W \mathsf{EMB}_+$, suppose $(S, \langle e_i \mid i \in I\rangle, n)$ is a partition problem. Let s_0 denote an element not appearing in S. Let $\langle F_n\rangle_{n\in\mathbb{N}}$ be an enumeration of the finite subsets of $S \cup \{s_0\}$, where each subset appears infinitely often. Let (M, e) be the matroid on $S \cup \{s_0\}$ defined by setting $e(m) = F_m$ if either (1) $s_0 \in F_m$, or (2) there are values $t_0 < t_1$ and i all less than m such that $e_i(t_0) \in F_m$, $e_i(t_1) \in F_m$, and $e_i(t_0) \neq e_i(t_1)$. Otherwise, let $e(m) = \{s_0\}$. The independent sets of (M, e) consist of finite lists of elements of S lying in distinct partition cells. Any solution of the matroid problem (M, e, n) must span (M, e) and so will consist of exactly one element from each cell in the partition. For this reduction also, the postprocessing computation is trivial. Theorem 17 of Hirst and Mummert [8] includes the reduction $\mathsf{EMB}_+ \leq_W \mathsf{GAC}_+$. By transitivity of Weihrauch reducibility, all three problems are Weihrauch equivalent. The reductions here and in the Hirst and Mummert result do not use the initial input in the postprocessing, so the result holds for strong Weihrauch reducibility. $\square$

The proof of the preceding result is easily modified to yield a reverse mathematical equivalence.

Theorem 15. (RCA_0) *The following are equivalent:*

(1) *If the enumerations $\langle e_i \mid i \in I\rangle$ partition S into at most n cells, then there is a finite set consisting of exactly one element from each cell.*

(2) $\mathsf{I}\Sigma^0_2$.

Proof. The construction used in the first Weihrauch reduction in Theorem 14 can be adapted to show that item (1) implies that every graph with finitely many components can be decomposed into its connected components. The second construction can be adapted to show that the fact that every finite dimensional matroid has a basis implies item (1). The equivalence of the graph and matroid statements with $\mathsf{I}\Sigma^0_2$ appears as Theorem 5 of Hirst and Mummert [8]. $\square$

Among the combinatorial statements equivalent to $\mathsf{I}\Sigma^0_2$ that are used in this paper, all the Weihrauch versions are equivalent, with the exception of the strictly weaker color basis problem. It would be interesting to know if there are other $\mathsf{I}\Sigma^0_2$ equivalent problems that are weak in the Weihrauch setting.

Bibliography

[1] Vasco Brattka and Guido Gherardi, *Effective choice and boundedness principles in computable analysis*, Bull. Symbolic Logic **17** (2011), no. 1, 73–117, DOI 10.2178/bsl/1294186663. MR2760117

[2] Vasco Brattka, Guido Gherardi, and Alberto Marcone, *The Bolzano-Weierstrass theorem is the jump of weak Kőnig's lemma*, Ann. Pure Appl. Logic **163** (2012), no. 6, 623–655, DOI 10.1016/j.apal.2011.10.006. MR2889550

[3] Vasco Brattka and Tahina Rakotoniaina, *On the uniform computational content of Ramsey's theorem*, J. Symb. Log. **82** (2017), no. 4, 1278–1316, DOI 10.1017/jsl.2017.43. MR3743611

[4] François G. Dorais, Damir D. Dzhafarov, Jeffry L. Hirst, Joseph R. Mileti, and Paul Shafer, *On uniform relationships between combinatorial problems*, Trans. Amer. Math. Soc. **368** (2016), no. 2, 1321–1359, DOI 10.1090/tran/6465. MR3430365

[5] Damir D. Dzhafarov and Carl Mummert, *Reverse mathematics—problems, reductions, and proofs*, Theory and Applications of Computability, Springer, Cham, 2022. MR4472209

[6] Kirill Gura, Jeffry L. Hirst, and Carl Mummert, *On the existence of a connected component of a graph*, Computability **4** (2015), no. 2, 103–117, DOI 10.3233/COM-150039. MR3393974

[7] Jeffry Lynn Hirst, *Combinatorics in subsystems of second order arithmetic*, 1987. Thesis (Ph.D.)–The Pennsylvania State University. MR2635978

[8] Jeffry L. Hirst and Carl Mummert, *Reverse mathematics of matroids*, Computability and complexity, Lecture Notes in Comput. Sci., vol. 10010, Springer, Cham, 2017, pp. 143–159. MR3629720

[9] ———, *Banach's theorem in higher-order reverse mathematics*, Computability **12** (2023), no. 3, 203–225, DOI 10.3233/com-230453. MR4668958

[10] Ulrich Kohlenbach, *Higher order reverse mathematics*, Reverse mathematics 2001, Lect. Notes Log., vol. 21, Assoc. Symbol. Logic, La Jolla, CA, 2005, pp. 281–295. MR2185441

[11] Benoit Monin and Ludovic Patey, *Partition genericity and pigeonhole basis theorems*, J. Symb. Log. **89** (2024), no. 2, 829–857, DOI 10.1017/jsl.2022.69. MR4747983

[12] Dag Normann and Sam Sanders, *On the uncountability of* $\mathbb{R}$, J. Symb. Log. **87** (2022), no. 4, 1474–1521, DOI 10.1017/jsl.2022.27. MR4510829

[13] RaTLoCC 2024, *Ramsey Theory in Logic, Combinatorics and Complexity, IV Edition* (June 11-14, 2024), `https://sites.google.com/uniroma1.it/lomec/events/ratlocc`. Accessed August 6, 2024.

[14] Stephen G. Simpson, *Subsystems of second order arithmetic*, 2nd ed., Perspectives in Logic, Cambridge University Press, Cambridge; Association for Symbolic Logic, Poughkeepsie, NY, 2009. MR2517689

[15] Rebecca Weber, *Computability theory*, Student Mathematical Library, vol. 62, American Mathematical Society, Providence, RI, 2012. MR2920681

[16] Klaus Weihrauch, *Computable analysis*, Texts in Theoretical Computer Science. An EATCS Series, Springer-Verlag, Berlin, 2000. An introduction. MR1795407