INDUCTIVE INFERENCE AND REVERSE MATHEMATICS

RUPERT HÖLZL, SANJAY JAIN AND FRANK STEPHAN

Abstract. The present work investigates inductive inference from the perspective of reverse mathematics. Reverse mathematics is a framework that allows gauging the proof strength of theorems and axioms in many areas of mathematics. The present work applies its methods to basic notions of algorithmic learning theory such as Angluin's tell-tale criterion and its variants for learning in the limit and for conservative learning, as well as to the more general scenario of partial learning. These notions are studied in the reverse mathematics context for uniformly and weakly represented families of languages. The results are stated in terms of axioms referring to induction strength and to domination of weakly represented families of functions.

Keywords. Reverse Mathematics, Recursion Theory, Inductive Inference, Learning from Positive Data.

Mathematics Subject Classification. 03D70 Inductive Definability, 03D80 Applications of Recursion Theory, 03H15 Nonstandard Models of Arithmetic.

1. Introduction

It is standard practice in mathematics to use known theorems to prove others. In these cases it can often be observed that some theorem T seems to be "stronger" than another theorem U in the sense that T allows proving U, but not vice versa. In the 1970s, Friedman [12] proposed a framework that formalises this intuition and allows gauging the different strength levels of theorems found in classical mathematics. The general idea is to assume only a subset of the axioms of second order arithmetic which by itself is too weak to prove the theorems in question, and then to analyse whether one theorem implies the other over this weak base system. Of course, if we want to *exactly* determine the strength of a mathematical theorem T in this sense, then we need to look at both directions: which theorems are implied by T and which imply T? As all of mathematics is ultimately founded on axioms, it is a natural next step to extend this study to the relation between axioms and theorems, and to wonder what *axioms* are exactly equivalent to a given theorem T, that is, imply T and are implied by T.

This "inverted" approach — where one uses theorems to prove axioms instead of the other way around — explains the name of this field of study: reverse mathematics. The

---

Rupert Hölzl, Department of Mathematics, National University of Singapore, S17, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore, r@hoelzl.fr, `http://hoelzl.fr`. R. Hölzl was supported by NUS/MOE grant R146-000-184-112 (MOE2013-T2-1-062).

Sanjay Jain, Department of Computer Science, School of Computing, National University of Singapore, 13 Computing Drive, Singapore 117417, Republic of Singapore, sanjay@comp.nus.edu.sg, `http://www.comp.nus.edu.sg/~sanjay/`. S. Jain was partially supported by NUS/MOE grant R146-000-184-112 (MOE2013-T2-1-062) and by NUS grant C252-000-087-001.

Frank Stephan, Department of Mathematics and Department of Computer Science, National University of Singapore, S17, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore; fstephan@comp.nus.edu.sg, `http://www.comp.nus.edu.sg/~fstephan/`. F. Stephan was partially supported by NUS/MOE grant R146-000-184-112 (MOE2013-T2-1-062).

A preliminary version of this article appeared in the Proceedings of the 32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015) [16].

subject has developed well since its inception, in particular thanks to many substantial contributions made by Simpson and his students [24].

The methodology of reverse mathematics has been applied to many fields of classical mathematics, for example to group theory, to vector algebra, to analysis and — especially in recent years — to combinatorics, including Ramsey theory and related fields. We refer to the books of Hirschfeldt [15] and Simpson [24] which are convenient resources for the topic and give many references.

In this article we propose to apply the methodology of reverse mathematics to the field of inductive inference, or algorithmic learning theory, as introduced and studied in numerous publications during the last decades [1, 2, 4, 7, 8, 13, 18, 23, 27]. A main focus of our work will be Angluin's tell-tale criterion for learnability of families of sets [2], one of the central results in classical algorithmic learning theory. We will also study several of its variants and a number of related results. We would like to point out that reverse mathematics analyses of inductive inference have also been proposed by de Brecht and Yamamoto [6] and by Hayashi [14] but that their approaches to the topic are unrelated to ours and to each other.

As algorithmic learning theory in its classical form is usually about the learnability of a language drawn from a family of languages that was fixed in advance, using a learner constructed specifically for that family, it is important for our study to know which such families are even guaranteed to exist over a weak base system as used in reverse mathematics. To handle this complication we define two ways of representing such a family by a single subset of the natural numbers: uniformly and weakly represented families of sets. Then when the set representing a family exists in the weak base system, so does the family. Similarly we will define weakly represented families of *functions* which will be used to formulate one of the reverse mathematics axioms used here.

As we will show in this article, many results in inductive inference relate to the following three axioms from reverse mathematics, which we only state informally for the time being.

- The axiom DOM which states that for every weakly represented family of functions there exists a function growing faster than all members of the family;
- The axiom ACA_0 which states that the Turing jump of every set exists;
- The axiom $\mathrm{I}\Sigma_2$ that allows inductive arguments on sets definable via Σ_2-formulas.

The remainder of this paper is organised as follows: In Section 2 we will introduce important definitions from reverse mathematics and inductive inference. Section 3 will discuss what bearing the tell-tale criterion has on the learnability of weakly represented families. The axiom DOM will be identified as necessary and sufficient for this. In Section 4 we will then follow Angluin's approach more closely and investigate the learnability of uniformly represented families, the closest equivalent in reverse mathematics of the indexed families that Angluin originally studied. We will show that for the learnability of uniformly represented families the degree of effectiveness of the bound in the tell-tale criterion is crucial. In this context we will also study conservative learning, where the axiom ACA_0 will be of relevance. Section 5 then focusses on sufficient criteria for learning from the classical theory and shows that they are sufficient for the learnability of uniformly represented families in reverse mathematics as well. However, for weakly represented families we will again require the additional axiom DOM. In Section 6 we will study the situation for partial learning where the axiom $\mathrm{I}\Sigma_2$ will play an important role.

2. Preliminaries

2.1. **Reverse mathematics.** In the practice of reverse mathematics we study proper subsets of the axioms of second order arithmetic and analyse the properties of possible models of these axiom sets. Such a model is of the form $(M, +, \cdot, <, 0, 1, \mathcal{S})$, where M is a (not necessarily standard) model of the natural numbers and $\mathcal{S}$ is a class of subsets of M. The minimal axiom system over which we will work is called RCA_0. Informally speaking, the axioms of this system guarantee that $\mathcal{S}$ contains at least all recursive sets and is closed under Turing join and Turing reduction. Furthermore, the axioms ensure that the system satisfies Σ_1-induction with parameters from $\mathcal{S}$.

More precisely, RCA_0 postulates that $(M, +, <, \cdot, 0, 1)$ behaves sufficiently similar to the natural numbers, in the following sense, and that $\mathcal{S}$ satisfies the following closure properties:

- The ordering $<$ is linear, transitive and antireflexive, and 0 is its least element;
- The successor mapping $x \mapsto x+1$ satisfies that $x < x+1$ and $x < y \Leftrightarrow x+1 < y+1$ as well as that 0 is the only $x \in M$ which is not equal to $y + 1$ for some $y \in M$;
- Addition $+$ is inductively defined from the successor mapping via $x + 0 = x$ and $x + (y + 1) = (x + y) + 1$;
- The ordering $<$ satisfies $x < y \Leftrightarrow \exists z\ [x + z + 1 = y]$;
- Multiplication $\cdot$ is inductively defined by $x \cdot 0 = 0$ and $x \cdot (y + 1) = (x \cdot y) + x$;
- The second order model satisfies Σ_1-induction, that is, if $I \subseteq M$ is defined by a Σ_1-formula using parameters from $\mathcal{S}$ and if I satisfies for all e the implication $(\forall d < e\ [d \in I]) \Rightarrow e \in I$, then $I = M$;
- The set $\mathcal{S}$ contains the empty set $\emptyset$;
- The set $\mathcal{S}$ is a Turing ideal, that is, if $I, J \in \mathcal{S}$ then

$$I \oplus J = \{i + i \colon i \in I\} \cup \{j + j + 1 \colon j \in J\} \in \mathcal{S},$$

  and if $J \in \mathcal{S}$ can be defined via both a Σ_1-definition with parameters in $\mathcal{S}$ and a Π_1-definition with parameters in $\mathcal{S}$, then $J \in \mathcal{S}$.

Note that it is the last two statements that ensure that $\mathcal{S}$ contains all sets recursive in the model $(M, +, <, \cdot, 0, 1)$ and that $\mathcal{S}$ is closed under Turing join and Turing reducibility.

When M is the standard model of the natural numbers, we call every $(M, +, \cdot, <, 0, 1, \mathcal{S})$ an ω-model, and otherwise a nonstandard model. Due to their well-behavedness compared to nonstandard models, ω-models are better understood. However, various complicated results in reverse mathematics were only obtained through the use of nonstandard models [9, 10]. The ω-model where $\mathcal{S}$ contains *exactly* the recursive sets is called the minimal model of RCA_0. Of course there are many models of RCA_0 that are much richer than the minimal model; for example, there exists an ω-model where $\mathcal{S} = \mathcal{P}(M)$, the power set of M.[1] There also exist many intermediate models between those two extremes.

By identifying a function with its graph we can also informally talk about the functions from M to M that exist in a model (that is, whose graphs are in $\mathcal{S}$). Note that Σ_1-induction implies in particular that even in nonstandard models of RCA_0 all numbers of the form $\max_{i<n} f(i)$ exist for all $n \in M$ and all functions $f \in \mathcal{S}$.

In an informal way, we will often think of the sets in $\mathcal{S}$ as being the recursive sets, even for sets that are not recursive in the usual sense of recursion theory. This seemingly strange fact can be understood as follows: Often in the reverse mathematics context we wonder whether a certain object Y exists in a model, or what additional axiom — say, for example, the axiom $\mathrm{I}\Sigma_2$ mentioned in the introduction — is needed to ensure its existence. We are then allowed to derive the existence of Y from any object X already existing in the model, no matter whether X is recursive in the usual sense; that is, in such a situation X is as good as recursive for our purposes.

[1]Nonstandard models cannot have $\mathcal{S} = \mathcal{P}(M)$ as such models would violate Σ_1-induction.

In this article the notion of finiteness of a set is of high importance, as it will feature prominently in Angluin's tell-tale criterion. We therefore point out that in the reverse mathematics setting some care is required with regard to this, as the universe M of the model $\mathcal{S}$ may be nonstandard. We therefore fix the term "finite" for $I \subseteq M$ to mean that I has an upper bound in M, and "infinite" to mean that no such bound exists. Furthermore, it should be noted that the term "finite sets" will always mean "finite sets in $\mathcal{S}$," and that these sets are precisely those E for which there is an element $e \in M$ with $e = \sum_{d \in E} 2^d$.

2.2. **Representing families of functions.** As explained in the introduction, in the course of this article we will need to represent families of sets and of functions by single sets in $\mathcal{S}$. This will then allow us to talk about such families existing or not existing in a given model.

We use Cantor's pairing function $\langle x, y\rangle = (x+y)\cdot(x+y+1)/2 + y$ and extend it appropriately to triples, quadruples and so on. Then the most straight-forward way to encode a family of sets $\{A_e\}_{e\in M}$ into a single set $A \in \mathcal{S}$ is by letting $x \in A_e \Leftrightarrow \langle e, x\rangle \in A$. Then $\{A_e\}_{e\in M}$ is called a *uniformly represented family of sets* and A its *representation set*.

Similarly one can define a *uniformly represented family of functions* by $F_e(x) = F(\langle e, x\rangle)$ using a single representation function $F \in \mathcal{S}$. This latter notion was generalised to the notion of weakly represented families of functions as follows [17]. Assume that a set $F \in \mathcal{S}$ satisfies the following conditions:

(1) For all e, x, y, z, y', z': If $\langle e, x, y, z\rangle, \langle e, x, y', z'\rangle \in F$ then $y = y'$ and $z = z'$;
(2) If $\langle e, x, y, z\rangle \in F$ and $x' < x$ then there exist y' and z' such that $\langle e, x', y', z'\rangle \in F$ and $\langle e, x', y', z'\rangle < \langle e, x, y, z\rangle$.

We let $D = \{e \colon \forall x\, \exists y, z\, [\langle e, x, y, z\rangle \in F]\}$ and call $\{F_e\}_{e\in D}$ the *weakly represented family of functions* defined by the *representation set* F where, for each $e \in D$, $F_e(x)$ is the unique y such that $\langle e, x, y, z\rangle \in F$ for some z. We call D the *index set* of $\{F_e\}_{e\in D}$.

The motivation behind Condition (2) above is to ensure that for each function F_e the quadruples encoding the function values $F_e(x)$ appear in F in order ascending in the function argument x, even if F_e is not monotone; for this purpose we use the fourth component of the quadruples as padding parameter. Condition (1) ensures that for each e, x there is at most one code defining $F_e(x)$. If an index e is in D it is called *valid*, otherwise *invalid*. Invalid indices do not encode functions.

We will say that $F_e(x)$ *can be retrieved in time at most* s if $c = \langle e, x, F_e(x), z\rangle \in F$ is bounded by s. The intuition of time is explained by the fact that $F_e(x)$ or the padding parameter z could be very large, so that it will depend on c how far we need to search in F to determine the function value $F_e(x)$ for a given x.

Note that the notion of weakly represented families of functions is much more general than that of uniformly represented families of functions; for example, for a fixed $A \in \mathcal{S}$ the family of all A-recursive functions is weakly representable but in general not uniformly representable.

In this article, in addition to weakly represented families of functions, also weakly represented families of *sets* will be used, namely as classes containing learning targets. A *weakly represented family of sets* is simply a weakly represented family $\{F_e\}_{e\in D}$ of functions where every F_e is $\{0,1\}$-valued, so that they can be interpreted as characteristic functions of sets. In this case we will denote them by A_e rather than F_e, and the representation set F by A. Note that Dzhafarov and Mummert [11] have considered the more general concept of *enumerated families of sets*.

To avoid confusion between both types of weakly represented families we emphasize here that while some of the reverse mathematics axioms used in this article will concern families of *functions*, all learnability results will concern families of *sets* and we will mostly omit the words "of sets" when talking about them.

2.3. **Reverse mathematics axioms.** We are now ready to state formally the three axioms from reverse mathematics that will be shown in this article to interact with numerous results on algorithmic learning.

- DOM states that for every weakly represented family $\{F_e\}_{e \in D}$ of functions there is a function $f \in \mathcal{S}$ dominating this family, that is,
$$\forall e \in D \ \exists x \ \forall y > x \ [F_e(y) < f(y)];$$
- ACA_0 states that every set arithmetically definable with parameters in $\mathcal{S}$ is also in $\mathcal{S}$ and, in particular, $\mathcal{S}$ is closed under Turing jump;
- $\mathrm{I}\Sigma_2$; where $\mathrm{I}\Lambda$ states that every set $I \subseteq M$ which is definable via a Λ-formula using parameters from $\mathcal{S}$ satisfies the induction axiom
$$\text{"if } \forall e \ [(\forall d < e \ [d \in I]) \Rightarrow e \in I], \text{ then } I = M."$$

2.4. **Learners.** We proceed with defining the relevant concepts from learning theory. The general scenario is that one learning target drawn from a family of sets is presented to the learner as an infinite sequence of data and the learner has to identify which of the possible targets the data is from. Such a data presentation is called a text. We define the notion of text in a way that is compatible with reverse mathematics, that is, in such a way that when M is equal to the standard natural numbers the definition coincides with the traditional one, but that for nonstandard models the definitions may differ.

Definition 1. *Let $T \in \mathcal{S}$ be a function from M to $M \cup \{\#\}$. We call "#" the* pause symbol *and write* $\mathrm{range}(T) = \{T(n) \colon n \in M \wedge T(n) \neq \#\}$. *Then T is called a* text *for $A \in \mathcal{S}$ if* $\mathrm{range}(T) = A$. *Without loss of generality we assume that $T(n) \in \{0, 1, \ldots, n\} \cup \{\#\}$ for all $n \in M$.*

The pause symbol "#" is a padding symbol that carries no information and is mainly used for giving a text for the empty set.

Note that here a set A in the second order model is called the range of a text (or function) T iff the formula
$$\forall x \in M \, [x \in A \Leftrightarrow \exists n \in M \, [x = T(n)]]$$
is true. Given $A \subseteq M$ we write A^* for the set of finite sequences over $A \cup \{\#\}$, where the domain of the finite sequence is of the form $\{y \in M : y < x\}$ for some $x \in M$. Thus the set M^* can be thought of as the prefixes of texts. Note that the word "finite" needs to be understood in the reverse mathematics sense discussed on page 4. For $\sigma \in M^*$ let $|\sigma|$ be its length and define
$$\mathrm{range}(\sigma) = \{\sigma(n) \colon \ n < |\sigma|\} \setminus \{\#\}.$$
M^* can be represented by a canonical indexing, where each finite sequence σ is represented by the canonical index of the set
$$\{\langle n, 0 \rangle \colon \sigma(n) = \#\} \cup \{\langle n, x+1 \rangle \colon \sigma(n) = x\}.$$
One can prove by induction over a text in $\mathcal{S}$ that such canonical indices exist for all of its prefixes.

The remaining definitions of this subsection will all be formulated for weakly represented families as this is the most general setting, but hold analogously for uniformly represented families by identifying them in the obvious way with weakly represented families that have $D = M$ or $E = M$, respectively.

Definition 2 (Angluin [2]; Gold [13]; Osherson, Stob and Weinstein [23]). *Let* $\{A_e\}_{e\in D}$ *and* $\{B_e\}_{e\in E}$ *be weakly represented families such that* $\{A_e\}_{e\in D} \subseteq \{B_e\}_{e\in E}$.

A learner *is a function* $L\colon M^* \to M$, *where* $L \in \mathcal{S}$ *and where the elements of* M^* *as inputs are represented by canonical indices.*

A learner L learns $\{A_e\}_{e\in D}$ with hypothesis space $\{B_e\}_{e\in E}$ in the limit *if for every* $e \in D$ *and every text* T *for* A_e *we have that* L *outputs a sequence of hypotheses*

$$e_n = L(T(0)\dots T(n))$$

such that, for some n, *for all* $m \geq n$, $e_m = e_{m+1}$ *and* $e_m \in E$ *and* $B_{e_m} = A_e$.

We say that $\{A_e\}_{e\in D}$ is learnable in the limit with hypothesis space $\{B_e\}_{e\in E}$ *if there exists a learner that learns* $\{A_e\}_{e\in D}$ *with hypothesis space* $\{B_e\}_{e\in E}$ *in the limit.*

When a learner updates its hypothesis, that is, when $e_m \neq e_{m+1}$, we call that a *mind change*.

Note that often $B_e = A_e$ for all e and $E = D$, that is, the learning target itself is used as hypothesis space. In these cases we will just say that L *learns* $\{A_e\}_{e\in D}$ *in the limit* and that $\{A_e\}_{e\in D}$ *is learnable in the limit*, respectively.

The intuition for learning in the limit is that when a learner learns a family its output should converge to an index of the member of the family that the given text corresponds to.

As we are working in a reverse mathematics setting we of course only work with weakly represented families of functions $\{F_e\}_{e\in D}$ whose representation set F is in $\mathcal{S}$ for the given model $(M, +, \cdot, <, 0, 1, \mathcal{S})$ of second order arithmetic. Then every F_e is in $\mathcal{S}$ as well. But note the important fact that in general the index set D need not be in $\mathcal{S}$. For this reason, we need to be careful when constructing learners for weakly represented families of sets $\{A_e\}_{e\in E}$, because a learner may conjecture elements of M that are not elements of E, as at the time of the conjecture it cannot know whether a particular element of M is a valid index or not. But note that a learner is allowed to output invalid indices when it is given a text that does not belong to a set in the family to be learnt.

Definition 3 (Angluin [2]). *A learner is called* conservative *if for* $n < m$ *and* $e_n \neq e_m$ *either* $e_n \notin E$ *or there exists some* $k \leq m$ *with* $T(k) \in M \setminus B_{e_n}$.

A learner conservatively learns $\{A_e\}_{e\in D}$ with hypothesis space $\{B_e\}_{e\in E}$ *if it is conservative and learns* $\{A_e\}_{e\in D}$ *with hypothesis space* $\{B_e\}_{e\in E}$ *in the limit.*

We say that $\{A_e\}_{e\in D}$ is conservatively learnable with hypothesis space $\{B_e\}_{e\in E}$ *if there exists a learner that conservatively learns* $\{A_e\}_{e\in D}$ *with hypothesis space* $\{B_e\}_{e\in E}$.

Intuitively, a conservative learner never makes an unjustified mind change. Again we simply say that a learner *conservatively learns* $\{A_e\}_{e\in D}$ and that $\{A_e\}_{e\in D}$ *is conservatively learnable*, respectively, in case $B_e = A_e$ for all $e \in M$ and $E = D$.

Definition 4 (Osherson, Stob and Weinstein [23]). *A learner* partially learns $\{A_e\}_{e\in D}$ with hypothesis space $\{B_e\}_{e\in E}$ *if for every* $e \in D$ *and every text* T *for* A_e *the learner outputs a sequence of hypotheses as above such that there is exactly one* d *with the property* $\forall m\, \exists n > m\, [d = e_n]$ *and this* d *satisfies* $d \in E$ *and* $B_d = A_e$.

We say that $\{A_e\}_{e\in D}$ is partially learnable with hypothesis space $\{B_e\}_{e\in E}$ *if there exists a learner that partially learns* $\{A_e\}_{e\in D}$ *with hypothesis space* $\{B_e\}_{e\in E}$.

Again we simply say that a learner *partially learns* $\{A_e\}_{e\in D}$ and that $\{A_e\}_{e\in D}$ *is partially learnable*, respectively, in case $B_e = A_e$ for all $e \in M$ and $E = D$.

Partial learning is a more general learning notion which, in the classical setting, allows learning the family of all r.e. sets.

2.5. **Angluin's tell-tale criterion.** Angluin [2] studied the learnability of *indexed families* of sets. These are families $\{A_e\}_{e\in\mathbb{N}}$ of sets for which there exists a computable two-place function A such that for all $e \in \mathbb{N}$ we have $A(e,x) = 1$ if $x \in A_e$ and $A(e,x) = 0$ otherwise. Indexed families are a special case of the uniformly represented families studied in this article; they are *actually* uniformly recursive, while uniformly represented families are in general only uniformly recursive relative to their representation set A. This is in analogy to the paradigm described above that in reverse mathematics often all sets in $\mathcal{S}$ are treated *as if* they were recursive.

Angluin's *tell-tale criterion* (or *tell-tale condition*) established in [2] states that an indexed family $\{A_e\}_{e\in\mathbb{N}}$ of sets is learnable in the limit if and only if uniformly in $e \in \mathbb{N}$ one can enumerate a finite *tell-tale set* $B_e \subseteq A_e$; that is, a set B_e such that there is no $d \in \mathbb{N}$ with $B_e \subseteq A_d \subset A_e$. It will turn out that to obtain analogous results for uniformly and weakly represented families of sets it is sufficient to consider bounds b_e for each A_e such that there is no A_d with $A_e \cap \{0,1,\dots,b_e\} \subseteq A_d \subset A_e$. We will call these bounds *tell-tale bounds* in the following. The next definition assigns names to different levels of effectivity for obtaining such tell-tale bounds.

Definition 5. *Let a weakly represented family $\{A_e\}_{e\in D}$ with index set D be given.*

- $\{A_e\}_{e\in D}$ satisfies the tell-tale criterion in general *iff for each $e \in D$ there is a b_e such that there is no $d \in D$ with*
$$A_e \cap \{0,1,\dots,b_e\} \subseteq A_d \subset A_e;$$
- $\{A_e\}_{e\in D}$ satisfies the tell-tale criterion in the limit *iff there is a two-place function $g \in \mathcal{S}$ such that for every $e \in D$ the limit $b_e = \lim_s g(e,s)$ exists and there is no $d \in D$ with*
$$A_e \cap \{0,1,\dots,b_e\} \subseteq A_d \subset A_e;$$
- $\{A_e\}_{e\in D}$ satisfies the tell-tale criterion effectively *iff there is a function $g \in \mathcal{S}$ such that for all $e \in D$ we have that there is no $d \in D$ with*
$$A_e \cap \{0,1,\dots,g(e)\} \subseteq A_d \subset A_e.$$

For uniformly represented families these definitions hold with $D = M$.

To avoid confusion we point out the informal use of the word "effectively" in the third item, which needs to be understood as "$g \in \mathcal{S}$."

Blum and Blum [4] introduced the notion of locking sequences, and established their existence. Their argument can easily be modified to carry over to the reverse mathematics setting as follows.

Given a weakly represented family $\{A_e\}_{e\in D}$, an $e \in D$ and a learner L, we say that $\sigma \in M^*$ is a *locking sequence for L on A_e* if *(i)* $L(\sigma\tau) = L(\sigma)$ for all $\tau \in (A_e)^*$ and *(ii)* $A_{L(\sigma)} = A_e$. That is, L will stick with the correct hypothesis for A_e it has reached after having processed σ, unless it subsequently sees data that is not drawn from $A_e \cup \{\#\}$. For uniformly represented families this definition holds with $D = M$.

Theorem 6. RCA_0 *proves the following: Suppose a learner L learns a weakly represented family $\{A_e\}_{e\in D}$ in the limit. Then there is a procedure $P \in \mathcal{S}$ which on all inputs $e \in M$ converges in the limit to a $\sigma \in M^*$ (represented by canonical indices as in Subsection 2.4); in case that $e \in D$, σ is a locking sequence for L on A_e.*

For uniformly represented families the statement holds with $D = M$.

Proof. We prove the theorem for weakly represented families; the argument applies analogously for uniformly represented families by identifying them in the obvious way with weakly represented families that have $D = M$.

For the proof we need an enumeration $\tau_0, \tau_1, \dots$ of the finite sequences over $M \cup \{\#\}$ that can be extended to texts in $\mathcal{S}$. To obtain such an enumeration, we recall that there is a coding of all finite *sets* and identify each finite sequence $\tau(0)\tau(1)\dots\tau(n)$ with the finite set $\{\langle 0, \tau(0)\rangle, \langle 1, \tau(1)\rangle, \dots, \langle n, \tau(n)\rangle\}$. Now if the i-th finite set is of this form for some sequence $\tau = \tau(0)\dots\tau(n)$ where for all i and j we have $\tau_i(j) \in \{\#, 0, 1, \dots, j\}$ then we let $\tau_i = \tau$, else we let τ_i be the empty sequence (the second half of the condition is needed to ensure extensibility of finite sequences into texts, as we had stipulated an analogous convention for texts in Definition 1). Now to see that all τ_i's can be extended to texts in $\mathcal{S}$, note that a text $T \in \mathcal{S}$ extending a given τ_i can be defined via $T(x) = y$ if $\langle x, y\rangle \in \tau_i$ and $T(x) = \#$ otherwise.

So let A be the representation set of $\{A_e\}_{e \in D}$. For $d \in M$ we let $P(d)$ enumerate a sequence $\sigma_{-1}, \sigma_0, \sigma_1, \dots \in M^*$, where σ_{-1} is the empty string and where σ_k, if it exists, is an extension of σ_{k-1} for all $k \geq 0$. Roughly speaking, once σ_{k-1} is defined, $P(d)$ searches in a greedy way for sequences σ_k which could cause another mind change of L, until it cannot find any further such σ_k that are consistent with A_d. We will show that for all inputs $d \in M$ only finitely many σ_k become defined in this way and that if we even have $d \in D$ then the last σ_k that becomes defined is a locking sequence for L on A_d.

More formally, let $d \in M$ be given and assume that at time s the last sequence enumerated by $P(d)$ was σ_{k-1}. Then consider the following process until σ_k becomes defined. At time $s + \langle i, j\rangle + 1$, $P(d)$ checks whether τ_i extends σ_{k-1}, whether $L(\tau_i) \neq L(\sigma_{k-1})$, whether for each x occurring in τ_i there is a code $\langle d, x, 1, \cdot\rangle \leq s + j$ in A and whether for all $x \leq k$ there is a code $\langle d, x, 0, \cdot\rangle \leq s + j$ or a code $\langle d, x, 1, \cdot\rangle \leq s + j$ in A. If the answers to all questions are "Yes" and $k \in A_d$ then let $\sigma_k = \tau_i\, k$; if the answers to all questions are "Yes" and $k \notin A_d$ then let $\sigma_k = \tau_i\, \#$; if the answer to at least one question is "No" then do not define σ_k at this stage.

If all σ_k become defined eventually we can define a text T via $T(k) = \sigma_k(k)$ for all k. In this case d must also be in D as the construction only defines σ_k once it has for all $x \leq k$ found information in A that determines whether $x \in A_d$ or not. Then T is a text for A_d by construction. As L was assumed to be learning $\{A_e\}_{e \in D}$ in the limit, L can make only finitely many mind changes on input T; this contradicts that we only define σ_{k+1} if L makes a mind change from σ_k to $\sigma_{k+1}(0)\dots\sigma_{k+1}(|\sigma_{k+1}| - 2)$. Therefore only finitely many σ_k become defined for all $d \in M$.

Now assume in addition that $d \in D$ and let k be largest such that σ_k becomes defined when running $P(d)$. Then we claim that σ_k is a locking sequence for L on A_d. To see this, note that for no non-empty $\tau \in (A_d)^*$ do we have $L(\sigma_k) \neq L(\sigma_k \tau)$, as otherwise σ_{k+1} would become defined. Therefore, if T is a text for A_d, then $\lim_\ell L(\sigma_k T(0)T(1)\dots T(\ell)) = L(\sigma_k)$. By the assumption that L learns $\{A_e\}_{e \in D}$ in the limit, we have

$$\lim_\ell L(\sigma_k T(0)T(1)\dots T(\ell)) = d'$$

with $A_{d'} = A_d$. Therefore σ_k is a locking sequence for L on A_d. □

From the existence of locking sequences follows that every weakly represented family which is learnable in the limit must satisfy the tell-tale criterion in general: the bound is simply the largest element contained in the locking sequence. The next section will address the question of which reverse mathematics axiom is needed for each of the above variants of the tell-tale criterion to ensure learnability of all weakly represented families that satisfy it. The axiom DOM will be identified as necessary and sufficient for all three versions.

3. Learnability of Weakly Represented Families

As mentioned above, the reverse mathematics counterpart of the indexed families studied by Angluin are the uniformly represented families. So it should not come as a surprise that to prove results similar to those of Angluin for families that are represented in a less accessible way, such as weakly represented families, we need some additional assumptions on the second order model. This assumption will turn out to be the axiom DOM which we will show is equivalent to the statement that every weakly represented family that satisfies the tell-tale criterion is learnable in the limit.

Note that Adleman and Blum [1] showed that one can learn all families of graphs of recursive functions (which all satisfy the tell-tale criterion) iff one has access to a dominating function as an oracle. The axiom DOM on the other hand ensures that for every weakly represented family of functions there is such a dominating function $f \in \mathcal{S}$, and the learner can then take advantage of f's existence. The following result can therefore be thought of as a generalization of the results by Adleman and Blum.

Theorem 7. *Over* RCA_0*, the following statements are equivalent:*

(1) DOM*;*
(2) The index set of every weakly represented family of functions can be approximated in the limit;
(3) Every weakly represented family that satisfies the tell-tale criterion effectively is learnable in the limit;
(4) Every weakly represented family that satisfies the tell-tale criterion in the limit is learnable in the limit;
(5) Every weakly represented family that satisfies the tell-tale criterion in general is learnable in the limit.

Proof. ($1 \Rightarrow 2$) Let $\{F_e\}_{e\in D}$ be a weakly represented family of functions with representation set F. For $e \in M$ let G_e be the function which, if it exists, assigns to x the minimum tuple $\langle e, x, y, z\rangle \in F$. Then G_e is total iff $e \in D$ and the functions form a weakly represented family $\{G_e\}_{e\in D}$ having the same index set as $\{F_e\}_{e\in D}$.

By assumption there is a function f dominating all $\{G_e\}_{e\in D}$. Now we claim that $e \in D$ if and only if, for all but finitely many numbers x, there are pairwise distinct elements

$$\langle e, 0, y_0, z_0\rangle, \dots, \langle e, x, y_x, z_x\rangle \in F \cap \{0, 1, \dots, f(x)\}.$$

This is because if, on one hand, $e \in D$, then the existence of such elements below $f(x)$ follows from the fact that f dominates G_e. On the other hand, if $e \notin D$, then there exists an x' such that F does not contain *any* element of the form $\langle e, x', \cdot, \cdot\rangle$; so in particular there exist only finitely many x such that $\langle e, 0, y_0, z_0\rangle, \dots, \langle e, x, y_x, z_x\rangle \in F \cap \{0, 1, \dots, f(x)\}$.

Define a function g by letting $g(e, x) = 1$ iff there are pairwise distinct elements

$$\langle e, 0, y_0, z_0\rangle, \dots, \langle e, x, y_x, z_x\rangle \in F \cap \{0, 1, \dots, f(x)\}$$

and $g(e, x) = 0$ otherwise. Then we have that $g(e, x)$ converges to 1 exactly when $e \in D$ and $g(e, x)$ converges to 0 exactly when $e \notin D$, so g is as needed.

($2 \Rightarrow 1$) Assume that a weakly represented family $\{F_e\}_{e\in D}$ of functions has representation set F and an index set D which is approximated by g in the limit. Then we define a function f via

$$f(x) = \min\{t\colon\ \forall e \le x\ [\exists u, y, z \le t\, (g(e, u + x) = 0 \vee \langle e, x, y, z\rangle \in F)]\}.$$

This f is total, as eventually for all indices e either a stage $u + x$ with $g(e, u + x) = 0$ or some value $\langle e, x, y, z\rangle \in F$ is found.

Furthermore, the minimum is taken over only finitely many conditions (in the square brackets) and for every condition individually the minimal t making it true can be computed from e and x (relative to the parameter $F \in \mathcal{S}$). Therefore, using Σ_1-induction, $f(x)$ exists as the maximum over the t's that are minimal for the individual conditions.

Note that the "$+x$" in the definition of f ensures that, for each e, wrong behaviour of $g(e, \cdot)$ during the first finitely many approximation stages is ignored in the limit. As a result the function f dominates each function F_e with $e \in D$, as for each such function there is a large enough $x \geq e$ with $g(e, u + x) = 1$ for all u and therefore $f(x') \geq F_e(x')$ for all $x' \geq x$.

(1 & 2 ⇒ 5) Let $\{A_e\}_{e \in D}$ be a weakly represented family that satisfies the tell-tale criterion in general. By (2) there is a function $g \in \mathcal{S}$ such that if $e \in D$ then $\lim_x g(e, x) = 1$, else $\lim_x g(e, x) = 0$. Now define for each $(e, b) \in M \times M$ a function $G_{\langle e,b\rangle}$ such that $G_{\langle e,b\rangle}(x)$ is the least $t \geq x$ found such that for each $d \leq x$ at least one of the following three conditions applies:

- We have $g(d, u + x) = 0$ or $g(e, u + x) = 0$ for some $u \leq t$;
- There is a number $x' \leq t$ such that $A_d(x')$ and $A_e(x')$ can be retrieved in time at most t and either $x' \in A_d \setminus A_e$ or $[x' \in A_e \setminus A_d \wedge x' \leq b]$;
- All values $A_d(x')$ and $A_e(x')$ for $x' \leq x$ can be retrieved in time at most t and $A_d(x') = A_e(x')$ for all $x' \leq x$.

These three conditions search for either e being an invalid index, or d being an invalid index, or x' witnessing that A_d is not a subset of A_e, or $x' \leq b$ witnessing that A_e is not a subset of A_d, or A_d being equal to A_e up to x. Note that when $e \in D$ the function $G_{\langle e,b\rangle}$ is total for exactly those b which are correct tell-tale bounds for A_e; thus the index set D' of the weakly represented family $\{G_{\langle e,b\rangle}\}_{\langle e,b\rangle \in D'}$ of functions is the set of all $\langle e, b\rangle$ such that either $e \notin D$ or such that b is a correct tell-tale bound for A_e. Now by (1) there is a function f dominating all functions in $\{G_{\langle e,b\rangle}\}_{\langle e,b\rangle \in D'}$. Note that whenever $\langle e, b\rangle \in D'$ then $\langle e, b + 1\rangle \in D'$ and $G_{\langle e,b\rangle}(x) \geq G_{\langle e,b+1\rangle}(x)$ for all x.

Let $(e_0, b_0), (e_1, b_1), \ldots$ be a sequence of elements of $M \times M$ in which each pair appears infinitely often. We construct a learner L as follows: L has an internal counter c with initial value 0. Throughout the entire construction L will always conjecture e_c for the current value of c.

Assume that L processes a text T and that at stage s the counter has reached some value c, so that L conjectures e_c. To determine whether an update to c is needed, L checks whether all of the following conditions are satisfied:

(a) We have $g(e_c, u + s) = 1$ for all $u \leq f(s)$ and all values $A_{e_c}(x)$ for $x \leq s$ can be retrieved in time at most $f(s)$;
(b) $G_{\langle e_c,b_c\rangle}(s) \leq f(s)$, which can be effectively verified by checking for each $t \leq f(s)$ whether it meets the conditions above that define $G_{\langle e_c,b_c\rangle}(s)$;
(c) $\{x\colon x \leq b_c \wedge x \in A_{e_c}\} \subseteq \{T(i)\colon i \leq s\}$;
(d) $\{x\colon x \notin A_{e_c}\} \cap \{T(i)\colon i \leq s\} = \emptyset$.

If all conditions are satisfied then L keeps c unchanged and continues to conjecture e_c; else L increments c by 1 and accordingly conjectures e_{c+1}.

Now fix $e \in D$ and assume L processes a text T for A_e. As a first possible outcome consider the case that c converges to an n such that e_n is an incorrect hypothesis or such that b_n is an incorrect tell-tale bound for A_{e_n}. The following cases are possible:

- If $e_n \notin D$ let t_0 be large enough so that $f(t_1) \geq G_{\langle e_n,b_n\rangle}(t_1)$ for all $t_1 \geq t_0$. Then it holds for large enough $t \geq t_0$ that $g(e_n, u + t) = 0$ for some $u \leq f(t)$. This will be detected by L when it checks Condition (a) at stage t.

- If $e_n \in D$ but b_n is an incorrect tell-tale bound for A_{e_n}, then for large enough t the value $G_{\langle e_n,b_n\rangle}(t)$ is undefined, and this will be detected by L when it checks Condition (b) at stage t.
- If $e_n \in D$ and b_n is a correct tell-tale bound for A_{e_n} but $A_{e_n} \neq A_e$, then for large enough t either

$$\{x\colon\ x \leq b_n \wedge x \in A_{e_n}\} \not\subseteq \{T(i)\colon\ i \leq t\}$$

or

$$\{x\colon\ x \notin A_{e_n}\} \cap \{T(i)\colon\ i \leq t\} \neq \emptyset.$$

This will be detected by L when it checks Conditions (c) and (d) at stage t.

In each case, by construction, c will be incremented to $n+1$ at a large enough stage t, contrary to the assumption.

The second possible outcome is that L infinitely often has $c = n$ for an n such that (e_n, b_n) is some fixed correct pair (e, b), that is, $e_n \in D$ and b_n is a correct tell-tale bound for A_{e_n} and $A_{e_n} = A_e$. As the function f dominates $G_{\langle e,b\rangle}$, it holds for all sufficiently large t where the current (e_n, b_n) is equal to (e, b) that all Conditions (a)–(d) are satisfied and that therefore c will not be incremented any further. So in this case L indeed converges to the correct hypothesis e.

As c is nondecreasing and grows in each step by at most 1, it either eventually converges or it equals every element of M at some stage; hence the above two outcomes are exhaustive and L must be correct.

($5 \Rightarrow 4$ and $4 \Rightarrow 3$) This holds by definition.

($3 \Rightarrow 2$) Let $\{F_e\}_{e\in D}$ be any weakly represented family of functions with representation set $F \in \mathcal{S}$. Define a new weakly represented family $\{A_{\langle e,s\rangle}\}_{\langle e,s\rangle\in E}$ of sets as follows.

- $A_{\langle e,0\rangle} = \{\langle e, x\rangle\colon x \in M\}$ if $e \in D$;
- otherwise $\langle e, 0\rangle$ is an invalid index;
- $A_{\langle e,s+1\rangle} = \{\langle e, x\rangle\colon x \leq s\}$ in case

$$s = \max\{\langle e, u, y, z\rangle \in F\colon\ u, y, z \in M\};$$

- otherwise $\langle e, s+1\rangle$ is an invalid index.

Note that for each e, there is exactly one s such that $\langle e, s\rangle$ is a valid index.

For every $e \in M$ let T_e be the text defined as follows. To define $T_e(s)$, let x be least such that $\langle e, x\rangle \notin \{T_e(s') : s' < s\}$. $T_e(s) = \langle e, x\rangle$ if for some $\langle e, u, y, z\rangle \leq s$ we have $x \leq \langle e, u, y, z\rangle$ and $\langle e, u, y, z\rangle \in F$; and $T_e(s) = \#$ otherwise. Then T_e is a text for $A_{\langle e,s\rangle}$ for the unique s such that $\langle e, s\rangle$ is a valid index. Note that T_e can be produced uniformly in e relative to the parameter $F \in \mathcal{S}$. Assume now that a learner L learns $\{A_{\langle e,s\rangle}\}_{\langle e,s\rangle\in E}$ in the limit. Thus, for every $e \in M$, L converges on T_e to some index d.

We define a function $g \in \mathcal{S}$ as follows: $g(e, t)$ searches for a $t' > t$ such that

- either $A_{L(T_e(0)\ldots T_e(t'))}(\langle e, x\rangle) = 0$ for some x with $\langle e, x\rangle < t$,
- or $A_{L(T_e(0)\ldots T_e(t'))}(\langle e, x\rangle) = 1$ for all x with $\langle e, x\rangle < t$.

In the first case let $g(e, t) = 0$, in the second case let $g(e, t) = 1$. Then we have

- $\lim_{t\to\infty} g(e, t) = 0$ in case that L converges on T_e to an index d with $\langle e, x\rangle \notin A_d$ for some $x \in M$ and
- $\lim_{t\to\infty} g(e, t) = 1$ in case that L converges on T_e to an index d with $\langle e, x\rangle \in A_d$ for every $x \in M$.

Note that the first case coincides with $e \notin D$ and the second with $e \in D$. Thus g is as needed. □

One might ask whether the necessity of DOM in this context is due to the difficulty of finding indices in weakly represented families rather than the difficulty of learning the languages contained in them. Therefore one might be inclined to choose a more comprehensive hypothesis space, hoping that this would simplify learning. However, in the proof of Theorem 7 ($3 \Rightarrow 2$) we only check whether the learner converges to an index of a set not containing some pair $\langle e, x\rangle$. As this is a property of the set, and not of its index, the choice of hypothesis space is not crucial for the proof.

Hölzl, Raghavan, Stephan and Zhang [17] investigate the reverse mathematics strength of DOM. They show that under RCA_0 and $\mathrm{I}\Sigma_2$, DOM implies COH but not vice versa. Here the axiom COH states that for every uniformly represented family $\{A_e\}_{e\in M}$ of sets there exists an infinite set $B \in \mathcal{S}$ such that for every $e \in M$ there is an $a \in M$ such that either A_e contains all $b \in B$ with $b > a$ or A_e does not contain any $b \in B$ with $b > a$. The described relationship between DOM and COH is a counterpart to the recursion-theoretic result that every high Turing degree contains a cohesive set, but that some cohesive sets do not have high Turing degree [19, 22].

The axiom DOM also has connections to set-theoretically motivated axioms for ω-models. For example, DOM is true iff MAD is false [17]. Here MAD is the statement that there exists a *maximal almost disjoint family*, that is, a weakly represented family $\{A_e\}_{e\in D}$ of sets such that *(i)* for all $d, e \in D$ with $d \neq e$, $A_d \cap A_e$ is finite and *(ii)* for every infinite $B \in \mathcal{S}$ there is an e such that $B \cap A_e$ is infinite. It is also known that DOM does not imply WKL_0, the statement that every infinite binary tree in $\mathcal{S}$ has an infinite branch in $\mathcal{S}$.

4. Learnability of Uniformly Represented Families

We now show that Angluin's classical result also applies to uniformly represented families in the framework of reverse mathematics.

Theorem 8. *Over* RCA_0*, a uniformly represented family is learnable in the limit if and only if it satisfies the tell-tale criterion in the limit.*

Proof. Let L be a learner that learns a uniformly represented family $\{A_e\}_{e\in M}$ in the limit. By Theorem 6 there is a procedure which on input e enumerates finitely many elements $\sigma_{e,0}, \sigma_{e,1}, \ldots \sigma_{e,\ell_e} \in M^*$ such that σ_{e,ℓ_e} is a locking sequence for L on A_e. Write $\ell_{e,s}$ for the number of $\sigma_{e,i}$'s that have been enumerated by stage s and define

$$g(e, s) = \max(\{\max(\mathrm{range}(\sigma_{e,i}))\colon\ 0 \leq i < \ell_{e,s}\}).$$

It is easy to verify that g witnesses that $\{A_e\}_{e\in M}$ satisfies the tell-tale criterion in the limit.

For the other direction let $e_0, e_1, \ldots$ be an infinite sequence in $\mathcal{S}$ in which every element of M occurs infinitely often. Let a uniformly represented family $\{A_e\}_{e\in M}$ and a function $g \in \mathcal{S}$ be given such that for every $e \in M$ the limit $b_e = \lim_s g(e, s)$ exists and there is no $d \in M$ with

$$A_e \cap \{0, 1, \ldots, b_e\} \subseteq A_d \subset A_e;$$

Then we define a learner L as follows: L has an internal counter c which is initialised with value 0. Throughout the entire construction, L always conjectures e_c for the current value of c. Write T for L's input. At stage s, L's counter has reached some value c and L checks whether

- $A_{e_c} \cap \{0, 1, \ldots, g(e_c, s)\} \subset \{T(i)\colon i \leq s\}$ and
- $\{T(i)\colon i \leq s\} \setminus A_{e_c} \subseteq \{\#\}$.

If both conditions hold, L keeps c constant; otherwise c is incremented by 1, implying a mind change of L to the new hypothesis e_{c+1}.

Assume now that L is processing a text T of a given set A_e: In case that c converges to a value n, then all elements of A_{e_n} below $b_n = \lim_{t\to\infty} g(e_n, t)$ are contained in a sufficiently long initial segment of T and no non-element of A_{e_m} is contained in T. Hence

$$A_{e_n} \cap \{0, 1, \dots, b_n\} \subseteq A_e \subseteq A_{e_n}.$$

Then it follows from the tell-tale criterion that $A_{e_n} = A_e$.

In case that c does not converge, c takes every element of M as its value at some time. Therefore infinitely often c satisfies $e_c = e$. Let s be large enough so that $g(e, s)$ has reached its limit and such that $A_e \cap \{0, 1, \dots, \lim_s g(e, s)\} \subseteq \{T(i) \colon i \leq s\}$. Then at the next stage $t > s$ where c satisfies $e_c = e$ the two update conditions above will be satisfied forever and c will never be updated again; thus this second case does not occur and L is correct. ☐

It is a natural next step to ask under what conditions a learner for a family exists if we replace the approximable tell-tale bounds in the statement of the previous theorem by general tell-tale bounds.

Theorem 9. *Over* RCA_0*,* DOM *holds iff every uniformly represented family that satisfies the tell-tale criterion in general is learnable in the limit.*

Proof. If DOM holds then learnability in the limit follows by Theorem 7, as every uniformly represented family also trivially has a weak representation.

So assume that a weakly represented family $\{F_e\}_{e\in D}$ of functions with representation set F is given. We construct the following uniformly represented family $\{A_{\langle e,s\rangle}\}_{e,s\in M}$ of sets:

- $A_{\langle e,0\rangle} = \{\langle e, x\rangle : x \in M\}$;
- $A_{\langle e,s+1\rangle} = \{\langle e, i\rangle \colon i \leq s\} \cup \{\langle e + 1, s\rangle\}$ if s is not of the form $\langle e, \cdot, \cdot, \cdot\rangle$ or $s \notin F$ (meaning informally that at position s no new value of F_e is coded into F);
- $A_{\langle e,s+1\rangle} = \{\langle e, i\rangle \colon i \leq s\}$ if s is the largest element in F which is of the form $\langle e, \cdot, \cdot, \cdot\rangle$ (that is, s is the last position at which a value of F_e is coded into F);
- $A_{\langle e,s+1\rangle} = \{\langle e, i\rangle \colon i \leq s\} \cup \{\langle e + 1, t\rangle\}$ if $s, t \in F$ and s, t are of the form $\langle e, \cdot, \cdot, \cdot\rangle$ and there is no $u \in F$ of the form $\langle e, \cdot, \cdot, \cdot\rangle$ such that $s < u < t$ (that is, at position s and at position $t > s$ some values of F_e are coded into F, but no such value is coded strictly between s and t).

Then by assumption $\{A_{\langle e,s\rangle}\}_{e,s\in M}$ is learnable in the limit, so by Theorem 8 there is a function $g \in \mathcal{S}$ such that for all $e, s \in M$ we have that $\lim_t g(\langle e, s\rangle, t)$ is a tell-tale bound for $A_{\langle e,s\rangle}$.

For a given $e \in M$ let $s = \lim_t g(\langle e, 0\rangle, t)$. If F_e is total, then there trivially exists a $u \in F$ such that $u > s$ and u is of the form $\langle e, \cdot, \cdot, \cdot\rangle$. Conversely, if for a given $e \in M$ we have $\lim_t g(\langle e, 0\rangle, t) = s$ and there is a $u \in F$ such that $u > s$ and u is of the form $\langle e, \cdot, \cdot, \cdot\rangle$, then we claim that F_e must be total. Assume otherwise and let $v \in F$ be largest of the form $\langle e, \cdot, \cdot, \cdot\rangle$. Then

$$A_{\langle e,v+1\rangle} = \{\langle e, i\rangle \colon i \leq v\}.$$

But then s cannot be a valid tell-tale bound for $A_{\langle e,0\rangle}$, as we have

$$A_{\langle e,0\rangle} \cap \{0, 1, \dots, s\} \subseteq A_{\langle e,v+1\rangle} \subset A_{\langle e,0\rangle},$$

contradiction.

It follows that F_e is total if and only if for $\lim_t g(\langle e, 0\rangle, t) = s$ there is a $u \in F$ such that $u > s$ and u is of the form $\langle e, \cdot, \cdot, \cdot\rangle$. Hence the index set D meets the requirements of Condition 2 of Theorem 7. As the above construction is possible for *any* weakly represented family of functions, DOM follows using Theorem 7. □

Recall from Definition 3 that conservative learning, as introduced by Angluin [2], requires that a learner only makes a mind change if some datum observed so far is not contained in the previously conjectured set or if the current conjecture is an invalid index. Conservative learners therefore never overgeneralise the language to be learnt. Thus before a conservative learner conjectures some language X it needs to ensure that there is no proper subset of X in the family being learnt which could explain the data observed so far. As the following theorem shows, conservative learning enforces the effective version of the tell-tale criterion and may require that the learner use a strictly larger hypothesis space than the family to be learnt.

Theorem 10. *Over* RCA_0*, a uniformly represented family* $\{C_e\}_{e\in M}$ *is conservatively learnable with a uniformly represented hypothesis space* $\{A_e\}_{e\in M}$ *if and only if* $\{C_e\}_{e\in M}$ *is contained in a uniformly represented family* $\{B_e\}_{e\in M}$ *(possibly different from* $\{A_e\}_{e\in M}$*) which satisfies the tell-tale criterion effectively.*

Proof. Let a learner L that conservatively learns $\{C_e\}_{e\in M}$ with hypothesis space $\{A_e\}_{e\in M}$ be given. Then one can effectively enumerate a sequence $P = ((e_k, b_k))_{k\in M}$ such that a pair (e_k, b_k) appears in P if and only if $L(a_0 a_1 \dots a_{b_k}) = e_k$ where

$$a_x = \begin{cases} x & \text{if } x \in A_{e_k}, \\ \# & \text{if } x \notin A_{e_k}. \end{cases}$$

This defines a new uniformly represented family $\{B_k\}_{k\in M}$ with $B_k = A_{e_k}$. Then, for each k, b_k is an effective tell-tale bound for B_k (in the family $\{B_k\}_{k\in M}$) as L cannot make a mind change to any proper subset of A_{e_k} once it has seen the data in $A_{e_k} \cap \{0, 1, \dots, b_k\}$ and conjectured e_k; hence for every $k' \in M$ we have that $B_k \cap \{0, 1, \dots, b_k\} \subseteq B_{k'} \subseteq B_k$ implies $B_{k'} = B_k$. It is easy to verify that for every language C_e, $e \in M$, some pair (e_k, b_k) with $A_{e_k} = C_e$ appears in P, implying that there is a B_k with $B_k = C_e$; hence $\{C_e\}_{e\in M} \subseteq \{B_k\}_{k\in M}$ (though $\{A_e\}_{e\in M}$ need not be contained in $\{B_k\}_{k\in M}$ in general).

For the converse direction assume that $\{C_e\}_{e\in M}$ is included in $\{B_e\}_{e\in M}$ which satisfies the tell-tale criterion effectively as witnessed by the function $e \mapsto b_e$ in $\mathcal{S}$. Assume without loss of generality that $B_0 = \emptyset$. Now a conservative learner L (using $\{A_e\}_{e\in M} = \{B_e\}_{e\in M}$ as hypothesis space) can be constructed as follows: L initially conjectures 0 and updates an old conjecture d to a new conjecture e at stage s when

- $\{T(i)\colon\ i \le s\} \setminus \{\#\} \not\subseteq B_d$,
- $\{T(i)\colon\ i \le s\} \setminus \{\#\} \subseteq B_e$ and
- $B_e \cap \{0, 1, \dots, b_e\} \subseteq \{T(i)\colon\ i \le s\}$;

should several such e exists, then L chooses the smallest one. It is easy to see that this learner is conservative and can be realised by a function in $\mathcal{S}$.

It remains to show that L is correct. To see this, let $U_{e,d}$ for $e, d \in M$ consist of the least element in $B_e \setminus B_d$ in case that this exists and let $U_{e,d}$ be empty in case that $B_e \subseteq B_d$. By Σ_1-induction, there exists a $\widetilde{b}_e > \max\left(\bigcup_{d<e} U_{e,d}\right)$.

Now assume that L is presented with a text T for a B_e with $B_e \neq B_{e'}$ for all $e' < e$. Further assume that L has read an initial segment of T sufficiently long to contain all elements in $\{0, 1, \dots, b_e + \widetilde{b}_e\} \cap B_e$ and that L's current hypothesis is d. It cannot be that $B_e \subset B_d$, because L is conservative and is reading a text for B_e. That leaves three possible cases:

- $B_d \subset B_e$. Since L has seen all of $\{0, 1, \ldots, b_e + \widetilde{b}_e\} \cap B_e$ and since b_e is a tell-tale bound for B_e, L must have seen a counter-example against B_d and must eventually make a mind change to reflect this.
- B_d is incomparable with B_e. Since in the construction we always choose minimal indices, we must have $d < e$. Then by choice of $\widetilde{b}_e$ the set $\{0, 1, \ldots, b_e + \widetilde{b}_e\} \cap B_e$ contains a counter-example against B_d; and since L has seen all elements of this set it must eventually make a mind change reflecting this.
- $B_d = B_e$. Since in the construction we always choose minimal indices, we must have $d \le e$; and then by the assumed minimality of e even $d = e$.

In each of these cases, L's next hypothesis will be e, as required. □

Corollary 11. *Over* RCA_0*, the following statements are equivalent:*
(1) ACA_0*;*
(2) Every uniformly represented family that satisfies the tell-tale criterion in general is conservatively learnable;
(3) Every weakly represented family that satisfies the tell-tale criterion in general is conservatively learnable.

Proof. ($1 \Rightarrow 2$) Let $\{B_e\}_{e\in M}$ be a uniformly represented family that satisfies the tell-tale criterion in general. Over ACA_0 all general bounds are effective bounds, so $\{B_e\}_{e\in M}$ also satisfies the tell-tale criterion effectively. Note that the proof of the backwards direction of Theorem 10 did not require a change of the hypothesis space (that is, it worked with $\{A_e\}_{e\in M} = \{B_e\}_{e\in M}$). So it follows with the same argument as there that $\{B_e\}_{e\in M}$ is conservatively learnable.

($2 \Rightarrow 1$) Given arbitrary $X \in \mathcal{S}$, let K^X denote the halting problem relative to X. We will show that $K^X \in \mathcal{S}$, implying that ACA_0 holds.

Define the following uniformly represented family $\{A_{\langle e,s\rangle}\}_{\langle e,s\rangle\in M}$.

- $A_{\langle e,0\rangle} = \{\langle e, x\rangle : x \in M\}$;
- $A_{\langle e,s+1\rangle} = \{\langle e, x\rangle \colon x \le s\}$ in case that $e \in K^X[s + 1] \setminus K[s]$, where $K^X[t]$ denotes the approximation of K^X relative to X at stage t;
- $A_{\langle e,s+1\rangle} = \emptyset$ otherwise.

Now assume that a learner L conservatively learns $\{A_{\langle e,s\rangle}\}_{\langle e,s\rangle\in M}$. Then we can decide whether $e \in K^X$ as follows: Search for an s such that

$$L(\langle e, 0\rangle \ldots \langle e, s\rangle) = \langle e, 0\rangle.$$

Such an s must be found, as $A_{\langle e,0\rangle}$ is in the family for every e and L learns this family.

Then we claim that $e \in K^X \Leftrightarrow e \in K^X[s + 1]$. Assume otherwise; then there exists a $t > s$ such that $e \in K^X[t + 1] \setminus K^X[t]$. Then $A_{\langle e,t+1\rangle} = \{\langle e, x\rangle \colon x \le t\}$ is a member of $\{A_{\langle e,s\rangle}\}_{\langle e,s\rangle\in M}$. Then L, being conservative, could not have output $\langle e, 0\rangle$ on input $\langle e, 0\rangle \ldots \langle e, s\rangle$ as all elements of that input are contained both in $A_{\langle e,0\rangle}$ and in its proper subset $A_{\langle e,t+1\rangle}$; contradiction.

($3 \Rightarrow 2$) This holds by definition.

(1 & $2 \Rightarrow 3$) Over ACA_0, the index set D of any weakly represented family is in $\mathcal{S}$, so that every weakly represented family also has a uniform representation. □

5. Sufficient Criteria for Learning

Besides the tell-tale criterion which characterises learnability, Angluin [2] discovered several other criteria that are only sufficient for learnability of indexed families. We will prove that these criteria are also sufficient over RCA_0 to ensure learnability of uniformly represented families. Furthermore we will show that the additional axiom DOM is again necessary and sufficient to extend these results to weakly represented families.

The first criterion that we consider is finite thickness.

Theorem 12. *Say that a family $\{A_e\}_{e\in D}$ of sets has finite thickness if and only if every $x \in M$ is contained in only finitely many A_e, that is, for every $x \in M$ there is a bound b such that for all $e > b$, either $e \notin D$ or $x \notin A_e$ or $A_e = A_d$ for some $d \leq b$.*

(1) *Over* RCA_0*, every uniformly represented family which has finite thickness is learnable in the limit.*
(2) *Over* RCA_0*,* DOM *is equivalent to the statement that every weakly represented family which has finite thickness is learnable in the limit.*

Proof. (1) We first define $h(x) = \lim_s h(x,s)$, for $x \in M$, as follows: $h(x,0) = 0$ and $h(x,s+1) = s+1$ if there exists an e with $h(x,s) \leq e \leq s+1$ such that $x \in A_e$ and $A_e \cap \{y : y \leq s+1\} \neq A_d \cap \{y : y \leq s+1\}$ for each $d < e$. Note that $h(x)$ gives an upper bound on the least indices of all sets in $\{A_e\}_{e\in M}$ that contain x.

To construct a function g that approximates the tell-tale bounds for $\{A_e\}_{e\in M}$, we start with $g(e,0) = 0$ and keep $g(e,s) = 0$ for all $s \leq \min(A_e)$. For $s \geq \min(A_e)$ we let $g(e,s+1) = s+1$ if there exists a $d \leq h(\min(A_e),s)$ such that

$$A_d \cap \{y : y \leq g(e,s)\} = A_e \cap \{y : y \leq g(e,s)\}$$

and there is an $x \in A_e \setminus A_d$ with $x \leq s+1$.

We claim that $\lim_s g(e,s)$ exists for every e. Assume otherwise; then A_e is not empty and there exist infinitely many sets A_d containing $\min(A_e)$ that cause updates to $g(e,\cdot)$. By assumption and definition of h, for each such update one can chose the witnessing set A_d to satisfy $d \leq h(\min(A_e))$. This allows defining a uniformly Σ_1 sequence of sets $(U_d)_{d\in M}$ such that for all d it holds that U_d contains the first stage s where A_d forces $g(e,s+1)$ to take value $s+1$, if such an s exists, and such that U_d is empty otherwise. As singletons, all of the sets U_d are trivially bounded and therefore by Σ_1-induction there is a common upper bound $s > \max\left(\bigcup_{d\leq h(\min(A_e))} U_d\right)$. Since for every set in $\{A_e\}_{e\in D}$ only its smallest index can be the cause of an update, we have for $s' \geq s$ that $g(e,s'+1) = g(e,s')$. As a consequence, $g(e) = \lim_s g(e,s)$ exists and there is no set A_d such that $A_d \cap \{y : y \leq g(e)\} = A_e \cap \{y : y \leq g(e)\}$ and such that A_d fails to contain some element of A_e. Hence $g(e)$ is a correct tell-tale bound for A_e and by Theorem 8 the family $\{A_e\}_{e\in D}$ is learnable in the limit.

(2) For the sufficiency of DOM let $\{A_e\}_{e\in D}$ be any given weakly represented family and A its representation set. Then we define a sequence of sets $(U_{e,d})_{e,d\in M}$ as follows: If it exists, $U_{e,d}$ contains the least $\langle e,x,y,z\rangle \in A$ such that $y = 1$ (implying that $x \in A_e$) and such that there exists a z' with $\langle d,x,0,z'\rangle \in A$ (implying $x \notin A_d$); otherwise $U_{e,d}$ is empty. Note that for given e, the sets $U_{e,d}$ are Σ_1 uniformly in d. Furthermore, by finite thickness, there is a bound b such that every set in $\{A_e\}_{e\in D}$ which contains $\min(A_e)$ has an index $d \leq b$. By Σ_1-induction there exists a b' with $\bigcup_{d\leq b} U_{e,d} < b'$. Then b' satisfies for all $d \in D$ that

$$A_e \cap \{0,1,\ldots,b'\} \subseteq A_d \subseteq A_e \Rightarrow A_d = A_e.$$

Therefore $\{A_e\}_{e\in D}$ satisfies the tell-tale criterion in general, and, by Theorem 7, the axiom DOM implies that it is learnable in the limit.

For the necessity of DOM, note that the sets $A_{\langle e,s\rangle}$ in the proof of $(3 \Rightarrow 2)$ of Theorem 7 satisfy that for each pair $\langle e,x\rangle$ there is at most one $(e,s) \in E \times M$ such that $A_{\langle e,s\rangle}$ contains $\langle e,x\rangle$ and that the sets $A_{\langle e,s\rangle}$, $(e,s) \in E \times M$, contain only elements of this form. Therefore, the weakly represented family defined in that proof has finite thickness and it follows with the same argument as there that learnability of all weakly represented families with finite thickness implies DOM. □

The property of finite thickness has been strengthened to finite elasticity [26]. Intuitively, finite elasticity of a family means that there is no text of which infinitely many pairwise distinct initial segments make a hypothesis inconsistent that was consistent with all shorter initial segments. One can also formulate this property the other way around: A family has finite elasticity if and only if for every text there is a prefix of that text such that every concept inconsistent with the full text is also inconsistent with this prefix.

Theorem 13. *Say that a family* $\{A_e\}_{e\in D}$ *of sets has finite elasticity if and only if for every* $T\colon M \to M \cup \{\#\}$ *in* $\mathcal{S}$ *there is a prefix* $\sigma \preceq T$ *such that for all* $e \in D$, $\mathrm{range}(\sigma) \subseteq A_e \Rightarrow \mathrm{range}(T) \subseteq A_e$.

(1) Over RCA_0*, every uniformly represented family which has finite elasticity is learnable in the limit.*
(2) Over RCA_0*,* DOM *is equivalent to the statement that every weakly represented family which has finite elasticity is learnable in the limit.*

Proof. (1) Like finite thickness, finite elasticity implies that the tell-tale criterion is satisfied in the limit. To see this, consider, for each A_e, the text T_e defined via $T_e(x) = x$ if $x \in A_e$, and $T_e(x) = \#$ otherwise. Now let $g(e, 0) = 0$ and in case that there is a $d \leq s$ with

$$A_e \cap \{0, 1, \ldots, g(e, s)\} \subseteq A_d \wedge A_e \cap \{0, 1, \ldots, s + 1\} \not\subseteq A_d$$

let $g(e, s + 1) = s + 1$; else let $g(e, s + 1) = g(e, s)$. By finite elasticity there is a bound b such that every A_d containing $\{T_e(i)\colon\ 0 \leq i \leq b\}$ also contains the full range of T_e. Since T_e has been built in such a way that

$$\{T_e(i)\colon\ 0 \leq i \leq b\} \setminus \{\#\} = A_e \cap \{0, 1, \ldots, b\}$$

we have that b also satisfies

$$A_e \cap \{0, 1, \ldots, b\} \subseteq A_d \Rightarrow A_e \subseteq A_d$$

for all $d \in M$ and therefore, once $g(e, s) \geq b$ for some s, it will hold that $g(e, t) = g(e, s)$ for all $t \geq s$. Thus $g(e) = \lim_s g(e, s)$ exists for all e and the function g witnesses that $\{A_e\}_{e\in D}$ satisfies the tell-tale criterion in the limit.

(2) For the sufficiency of DOM, observe that the argument for (1) also works for weakly represented families, but that we only obtain general bounds. However, we then meet Condition 5 in Theorem 7, which implies that every weakly represented family having finite elasticity is learnable over the axiom DOM.

For the necessity of DOM, note that the sets $A_{\langle e,s\rangle}$ in the proof of $(3 \Rightarrow 2)$ of Theorem 7 satisfy that for each pair $\langle e, x\rangle$ there is at most one $(e, s) \in E \times M$ such that $A_{\langle e,s\rangle}$ contains $\langle e, x\rangle$, and that the sets $A_{\langle e,s\rangle}$, $(e, s) \in E \times M$, contain only elements of this form. Therefore, the weakly represented family defined in that proof has finite elasticity and it follows with the same argument as there that learnability of all weakly represented families with finite elasticity implies DOM. □

Note that finite elasticity is only a sufficient criterion. For example the trivially learnable family of the sets $\{0, 1, \ldots, e\}$ with $e \in M$ does not have finite elasticity.

Angluin [3] and Kobayashi [5, 21] considered another sufficient learnability criterion which is a further strengthening of the property of finite elasticity: An indexed family is learnable in the limit if for every language A_e there is a finite subset E such that $E \subseteq A_d \Rightarrow A_e \subseteq A_d$ for all other languages A_d in the indexed family. This criterion was shown to hold without any effectivity requirement on finding E; a result which carries over to uniformly represented and weakly represented families in the following form.

Theorem 14. *Say that a family $\{A_e\}_{e\in D}$ of sets admits characteristic subsets if and only if for all $e \in D$ exists $b \in M$ such that for all $d \in D$ we have that $A_e \cap \{0, 1, \ldots, b\} \subseteq A_d$ implies $A_e \subseteq A_d$.*

(1) Over RCA_0*, every uniformly represented family which admits characteristic subsets is learnable in the limit.*

(2) Over RCA_0*,* DOM *is equivalent to the statement that every weakly represented family which admits characteristic subsets is learnable in the limit.*

Note that admitting characteristic subsets is a stronger property than satisfying the tell-tale criterion, as the first condition enforces

$$A_e \cap \{0, 1, \ldots, b\} \subseteq A_d \Rightarrow A_e \subseteq A_d,$$

while the tell-tale criterion merely enforces

$$A_e \cap \{0, 1, \ldots, b\} \subseteq A_d \Rightarrow A_d \not\subset A_e.$$

Proof. (1) We build a learner L that uses a list $e_0, e_1, \ldots$ of hypotheses such that each $e \in M$ occurs infinitely often in this list. L has a counter c initialised with 0. Throughout the entire construction, L always conjectures e_c with the current value of c. Write T for the text that L processes and assume that at stage s the counter has reached some value c. Then L checks whether there is an $e \leq s$ such that

- $\{T(i) \colon i \leq s\} \setminus \{\#\} \subseteq A_e$ and
- there is some $x \in A_{e_c} \setminus A_e$ with $x \leq s$.

If such an e exists then L increments c by 1, implying a mind change to e_{c+1}.

Now assume that L is processing a text T for a set A_e and that b is as in the statement of the theorem. Then there is a stage $s \geq e$ where $A_e \cap \{0, 1, \ldots, b\} \subseteq \{T(i) \colon i \leq s\}$. From that stage s onwards, L discards every current hypothesis A_{e_n} eventually whenever either A_{e_n} does not contain $A_e \cap \{0, 1, \ldots, b\}$ or there is an element in $A_{e_n} \setminus A_e$. Hence L will cycle through the hypotheses until it finds an n with $A_e \cap \{0, 1, \ldots, b\} \subseteq A_{e_n} \subseteq A_e$. By the assumption that the family admits characteristic subsets, the set A_{e_n} is equal to A_e and therefore L will converge to a correct hypothesis.

(2) For the necessity of DOM, note that the the sets $A_{\langle e,s\rangle}$ in the proof of $(3 \Rightarrow 2)$ of Theorem 7 satisfy that for each e there is at most one $(e, s) \in E \times M$ such that $A_{\langle e,s\rangle}$ contains $\langle e, 0\rangle$ and that every $A_{\langle e,s\rangle}$, $(e, s) \in E \times M$, contains $\langle e, 0\rangle$ for at most one e. Hence the family of sets defined in that proof has characteristic subsets, and it follows with the same argument as there that learnability of all weakly represented families that admit characteristic subsets implies DOM.

For the sufficiency of DOM, note that the b witnessing the characteristic subset property for A_e is in particular also a tell-tale bound for A_e. Thus, learnability follows from Condition 5 in Theorem 7. □

6. Partial Learning

Osherson, Stob and Weinstein [23] introduced the notion of partial learning where to be successful a learner is required to output one correct hypothesis infinitely often and all other hypotheses at most finitely often. This fundamental notion allows to learn the family of all r.e. languages, possibly after enlarging the hypothesis space. While proving the corresponding result for uniformly represented families in reverse mathematics is possible over RCA_0, our proofs for weakly represented families depend on the additional axiom $\mathrm{I}\Sigma_2$ which in particular proves that every set in a weakly represented family has a least index. It is an open question whether this is an inherent requirement for obtaining the results, or one that more involved arguments could dispense with.

We begin with the following important theorem.

Theorem 15. *Over* RCA_0*, a weakly represented family* $\{A_d\}_{d\in D}$ *is partially learnable if and only if there is a weakly represented family* $\{B_e\}_{e\in E}$ *such that*

- *for all* $d \in D$ *there is exactly one* $e \in E$ *with* $B_e = A_d$ *and*
- *all* $e \in E$ *are in* D *and satisfy* $A_e = B_e$*.*

That is, $\{B_e\}_{e\in E}$ is a trimmed version of $\{A_d\}_{d\in D}$ containing exactly one index for each set A_d, $d \in D$.

Proof. Note that the eventually 0 functions in $\mathcal{S}$ have a canonical indexing and that in the following constructions we can use such indices to keep track of how often a learner has output which hypotheses so far.

So assume that a learner L is given which partially learns the weakly represented family $\{A_d\}_{d\in D}$ with representation set A. We define a new weakly represented family $\{B_e\}_{e\in E}$ via its description set B as follows: B contains a pair $\langle e, x, y, z\rangle$ if and only if z is the least number such that the following three conditions are satisfied:

(1) For all $x' < x$ there is a $\langle e, x', y', z'\rangle \in B$ with $\langle e, x', y', z'\rangle < z$;
(2) We have $y \in \{0, 1\}$ and $\langle e, x, y, z'\rangle \in A$ for some $z' \le z$;
(3) There are $s \le z$ and
$$\langle e, 0, y_0, z_0\rangle, \langle e, 1, y_1, z_1\rangle, \ldots, \langle e, s, y_s, z_s\rangle \in A$$
such that there are at least $x + 1$ many $\ell \le s$ with
$$L(a_0a_1 \ldots a_\ell) = e$$
where, for $t \le s$, $a_t = t$ if $y_t = 1$, and $a_t = \#$ otherwise.

It is clear from the definition that B is the representation set of a weakly represented family $\{B_e\}_{e\in E}$ and that $B_e = A_e$ whenever $e \in E$. Furthermore, given a member X of the family $\{A_d\}_{d\in D}$, and the canonical text T satisfying $T(x) = x$ for $x \in X$ and $T(x) = \#$ otherwise, L outputs on T one correct index e infinitely often and $e \in E$ and $B_e = A_e$. On the other hand, since we assumed that L partially learns $\{A_d\}_{d\in D}$, it follows from the definition of partial learning that for all other indices d of X the x for which Condition (3) above can be true are finitely bounded, which implies that $d \notin E$. Hence the family $\{B_e\}_{e\in E}$ is as required.

For the converse direction, assume that for a given weakly represented family $\{A_d\}_{d\in D}$ the corresponding family $\{B_e\}_{e\in E}$ exists and is weakly represented by B. Then we define a learner L which given some text T keeps track for all indices of how often they have been conjectured so far and outputs an index e at least n times iff

> there is an $s \ge n$ such that the $B_e(0)$, $B_e(1)$, …, $B_e(n)$ can be retrieved in time at most s and such that for all $x \le n$ we have $[x \in \{T(i) : i \le s\} \Leftrightarrow B_e(x) = 1]$.

It is clear that an index e is output infinitely often iff $e \in E$ and $B_e = \mathrm{range}(T)$. As each A_d with $d \in D$ has exactly one index $e \in E$ with $B_e = A_e$, it follows that L outputs exactly one index e infinitely often. □

Theorem 16. *Over* RCA_0*, every uniformly represented family is partially learnable.*

Proof. We first claim that every member of a uniformly represented family has a minimal index. To see this, assume that a set $X \in \mathcal{S}$ has no minimal index in a uniformly represented family $\{A_e\}_{e\in M}$. Consider the Σ_1 set
$$I = \{e \in M : \exists x\, [X(x) \ne A_e(x)]\}$$

of elements of M which are not indices of X. Since X has no minimal index, I must satisfy for all $e \in M$ that

$$(\forall d < e \;\; [d \in I]) \Rightarrow e \in I.$$

Then by Σ_1-induction $I = M$ and $X \notin \{A_e\}_{e \in M}$, a contradiction. Thus every member of $\{A_e\}_{e \in M}$ has a minimal index.

For every e let $b_e = \min\{b : \forall d < e \; \exists x < b \; [A_d(x) \neq A_e(x)]\}$, whenever this number is defined. First note that b_e cannot be defined for nonminimal indices e. And secondly note that for every minimal index e, b_e must indeed exist; this is because one can assign to every index $d < e$ the set $U_d = \{\min\{x : A_d(x) \neq A_e(x)\}\}$ and the maximum of this uniform sequence of finitely many Σ_1 singletons is in M by Σ_1-induction.

Note that when e is given such that b_e exists then one can effectively test whether a given number equals b_e, implying that given such an e one can effectively find b_e. Therefore we can define a weakly represented family via its representation set B as follows: B contains all $\langle e, 0, A_e(0), b_e\rangle$ where b_e is defined and, inductively, all $\langle e, x+1, A_e(x+1), u\rangle$ for which there is a value $u \in B$ of the form $\langle e, x, \cdot, \cdot\rangle$. The weakly represented family obtained in this way satisfies the properties required by the right-hand side of Theorem 15 and thus a partial learner for the original uniformly represented family $\{A_e\}_{e \in M}$ exists. □

Theorem 17. *Over* RCA_0 *and* $\mathrm{I}\Sigma_2$*, every weakly represented family* $\{A_e\}_{e \in D}$ *is partially learnable with hypothesis space* $\{B_{\langle e,b\rangle}\}_{e \in D, b \in M}$ *where* $B_{\langle e,b\rangle} = A_e$ *for all* $e \in D$ *and* $b \in M$*.*

Proof. Let a weakly represented family $\{A_e\}_{e \in D}$ be given, let A be its representation set and let $X \in \mathcal{S}$ be arbitrary. Now consider the Σ_2 set

$$I = \{e\colon \; \exists x \; \forall y, z \; [\langle e, x, y, z\rangle \notin A \vee y \neq X(x)]\}$$

consisting of the e's which are not indices of X in $\{A_e\}_{e \in D}$. In case that X does not have a minimal index in $\{A_e\}_{e \in D}$, the set I satisfies for all e that

$$(\forall d < e \, [d \in I]) \;\Rightarrow\; e \in I,$$

which by $\mathrm{I}\Sigma_2$ implies $I = M$ and therefore $X \notin \{A_e\}_{e \in D}$. It follows that for every $d \in D$ the set A_d has a minimal index in $\{A_e\}_{e \in D}$.

So let the minimal index e of a member of $\{A_e\}_{e \in D}$ be given and define for $d < e$ the uniformly Σ_2 sequence of singletons

$$U_d = \{\min\{x\colon \; A_d(x) \text{ is not defined or } A_d(x) \neq A_e(x)\}\}.$$

Let b_e be the least upper bound on all numbers appearing in some U_d, with $d < e$, which exists by another application of $\mathrm{I}\Sigma_2$.

Now define a partial learner L as follows: A hypothesis $\langle e, b\rangle$ is output at least n times if and only if there is an $s \geq n$ such that the following conditions are satisfied:

- $A_e(0), A_e(1), \ldots, A_e(n)$ can be retrieved in time at most s;
- There is no $d < e$ such that for all $x \leq b$ the values $A_d(x)$ and $A_e(x)$ can be retrieved in time at most s and such that $A_d(x) = A_e(x)$;
- For all $b' < b$ there is a $d < e$ such that for all $x \leq b'$ the values $A_d(x)$ and $A_e(x)$ can be retrieved in time at most s and such that $A_d(x) = A_e(x)$.

It is easy to verify that if L is processing a text for an $X \in \{A_e\}_{e \in D}$, then L outputs exactly one pair $\langle e, b\rangle$ infinitely often, namely the one where e is the least index of X and b is least such that all $d < e$ satisfy that either $A_d(b)$ is not defined or that there is an $x \leq b$ with $A_e(x) \neq A_d(x)$. Thus L partially learns $\{A_e\}_{e \in D}$ with hypothesis space $\{B_{\langle e,b\rangle}\}_{e \in D, b \in M}$. □

In the course of the last proof we established that, over RCA_0 and $\mathrm{I}\Sigma_2$, every member of a weakly represented family has a least index. In fact, over RCA_0, this latter assumption is even equivalent to $\mathrm{I}\Sigma_2$.

Proposition 18. *Over* RCA_0*, the axiom* $\mathrm{I}\Sigma_2$ *is equivalent to the statement that in every weakly represented family, all its members have a minimal index.*

Proof. The sufficiency of $\mathrm{I}\Sigma_2$ was already shown in the proof of Theorem 17. For the necessity assume that in every weakly represented family, all its members have a minimal index, but that $\mathrm{I}\Sigma_2$ is not satisfied. Then there is a Σ_2 set I which is a proper subset of M and which satisfies for all e that $(\forall d < e\,[d \in I]) \Rightarrow e \in I$, implying in particular that $M \setminus I$ cannot have a minimal element. As I is Σ_2, there is a ternary $\{0,1\}$-valued function $g \in \mathcal{S}$ such that

$$e \in I \Leftrightarrow \exists n \in M\ \forall m \in M\ [g(e,n,m) = 1].$$

Define a weakly represented family $\{A_d\}_{d\in D}$ via its representation set A as follows: For each e, inductively over n put into A the elements $\langle e, n, 1, z_{e,n}\rangle$ with $z_{e,0} = 0$ and $z_{e,n+1} = z_{e,n} + m$ where m is least such that $g(e,n,m) = 0$, if they exist.

Now fix an arbitrary e.

- If there is a least n such that $z_{e,n+1}$ is undefined then $g(e,n,m) = 1$ for all m. Thus, $e \in I$.
- If there is no least n with the property that $z_{e,n+1}$ is not defined then consider the Σ_1 set $J = \{n\colon z_{e,n+1}$ is defined$\}$ and use Σ_1-induction to show that $J = M$. It follows that all $z_{e,n}$ are defined and thus for all n exists an $m = z_{e,n+1} - z_{e,n}$ with $g(e,n,m) = 0$. Hence $e \notin I$.

It follows that $D = M \setminus I$. But note that by construction $\{A_d\}_{d\in D}$ contains only one unique member, that is, $A_d = M$ for every $d \in D$. Then, as $D = M \setminus I$ has no minimal element, M has no minimal index in $\{A_d\}_{d\in D}$, a contradiction. □

Theorem 17 showed that, over RCA_0 and $\mathrm{I}\Sigma_2$, weakly represented families are always partially learnable with a padded hypothesis space. The following result shows that the additional axiom DOM renders this padding unnecessary, that is, learning becomes possible with the weakly represented family itself as hypothesis space.

Theorem 19. *Over* RCA_0, $\mathrm{I}\Sigma_2$ *and* DOM, *every weakly represented family* $\{A_e\}_{e\in D}$ *is partially learnable.*

Proof. Given a weakly represented family $\{A_e\}_{e\in D}$ with representation set A, we consider the weakly represented family of functions $\{G_e\}_{e\in D}$ such that, for each $e \in D$ and $x \in M$, $G_e(x)$ is the unique tuple of the form $\langle e, x, y, z\rangle \in A$ defining $A_e(x)$. By DOM there is an increasing function $f \in \mathcal{S}$ dominating all G_e, $e \in D$.

Now we define a learner L as follows: When L processes a text T, it outputs an index e at least n times iff there is an $m \geq n$ and an $s \geq f(m)$ such that the following conditions are met:

- $A_e(0), A_e(1), \ldots, A_e(m)$ can be retrieved in time at most $f(m)$;
- There is no $d < e$ such that $A_d(0)$, $A_d(1)$, $\ldots$, $A_d(m)$ can be retrieved in time at most $f(m)$ and such that
$$A_d(0) = A_e(0), A_d(1) = A_e(1), \ldots, A_d(m) = A_e(m);$$
- All numbers $x \leq m$ with $A_e(x) = 1$ are in $\{T(i)\colon i \leq s\}$;
- No number $x \leq m$ with $A_e(x) = 0$ is in $\{T(i)\colon i \leq s\}$.

These requirements are sufficient to prove that L indeed partially learns $\{A_e\}_{e\in D}$: note that, together with the fact that f dominates $\{G_e\}_{e\in D}$, the first two conditions ensure that only minimal indices (whose existence is ensured by $\mathrm{I}\Sigma_2$) are output infinitely often and that the last two conditions ensure that a minimal index is output infinitely often iff it is correct. □

One can show with easy recursion-theoretic constructions that there are weakly represented families in the minimal standard model of RCA_0 that do not have a partial learner. Hence it is clear that *some* additional assumption beyond RCA_0 and $\mathrm{I}\Sigma_2$ is needed to obtain the conclusion of Theorem 19, but it is an open question whether an assumption strictly weaker than DOM can suffice.

As the final result of this article we show that over DOM the arithmetic hierarchy collapses with regard to induction strength; that is, over DOM, $\mathrm{I}\Delta_2$ implies $\mathrm{I}\Sigma_n$ for all n. Note that $\mathrm{I}\Delta_2$ induction is equivalent to the better known criterion $\mathrm{B}\Sigma_2$ by a result of Slaman [25].

Theorem 20. *Over* RCA_0 *and* DOM *and* $\mathrm{I}\Delta_2$*, the axiom* $\mathrm{I}\Sigma_n$ *holds for all* n*.*

Proof. The proof is by induction over $n \geq 2$; we will show that every set E that is Σ_n over $\mathcal{S}$ is even Δ_2 over $\mathcal{S}$.

Case $n = 2$: Given E, there is a ternary $\{0,1\}$-valued function $R \in \mathcal{S}$ such that

$$e \in E \;\Leftrightarrow\; \exists x \;\forall y \;[R(e,x,y) = 1],$$

and we can define for each $e \notin E$ a function F_e such that $F_e(x)$ is the least y with $R(e,x,y) = 0$. Now $\{F_e\}_{e\in M\setminus E}$ forms a weakly represented family of functions and there is a function $f \in \mathcal{S}$ dominating all F_e, $e \in M \setminus E$. If for all $x' \leq x$ there is a

$$y' \leq f(0) + f(1) + \cdots + f(x) + x$$

such that $R(e,x',y') = 0$ then let $Q(e,x) = 0$; else let $Q(e,x) = 1$. Due to f dominating each F_e with $e \in M \setminus E$, Q has the following property:

$$\text{If } e \in E \quad \text{then } \forall^\infty x \;[Q(e,x) = 1] \quad \text{else } \forall^\infty x \;[Q(e,x) = 0].$$

Since $Q \in \mathcal{S}$, this is a Π_2 condition defining E, and consequentially E is Δ_2. Thus, over RCA_0 and DOM, $\mathrm{I}\Delta_2$ and $\mathrm{I}\Sigma_2$ are equivalent.

Case $n = 3$: Let E be a Σ_3 set. Then there is a quaternary relation $R \in \mathcal{S}$ such that

$$e \in E \;\Leftrightarrow\; \exists x \;\forall y \;\exists z \;[R(e,x,y,z) = 1],$$

and there is a weakly represented family of functions $F_{\langle e,x\rangle}$ such that for all e, x satisfying $\forall y \;\exists z \;[R(e,x,y,z) = 1]$ the function $F_{\langle e,x\rangle}$ is total and picks for each y the least z with $R(e,x,y,z) = 1$. Again there is a function f dominating all total functions $F_{\langle e,x\rangle}$ and we can define $Q \in \mathcal{S}$ as follows: $Q(e,x,y) = 1$ iff for all $y' \leq y$ there is a $z' \leq f(0) + f(1) + \cdots + f(y) + y$ such that $R(e,x,y',z') = 1$. Now

$$\forall y \;\exists z \;[R(e,x,y,z) = 1] \;\Leftrightarrow\; \exists y \;\forall y' > y \;[Q(e,x,y') = 1]$$

and so we obtain $e \in E \;\Leftrightarrow\; \exists x \;\exists y \;\forall y' > y \;[Q(e,x,y') = 1]$. This shows that E is in fact a Σ_2 set and thus a Δ_2 set.

Similarly, one can show for all even $n > 2$ that every Π_n set is in fact a Π_{n-1} set and for every odd $n > 2$ that every Σ_n set is a Σ_{n-1} set. Alternating between sets and their complements collapses the entire arithmetic hierarchy to Δ_2. Thus, over RCA_0 and DOM, $\mathrm{I}\Delta_2$ implies $\mathrm{I}\Sigma_n$ for all n. □

7. Acknowledgments

The authors would like to thank Chong Chi Tat and Yang Yue for helpful discussions on aspects of reverse mathematics related to this paper.

References

[1] Lenny Adleman and Manuel Blum. Inductive inference and unsolvability. *The Journal of Symbolic Logic*, 56:891–900, 1991.

[2] Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.

[3] Dana Angluin. Inference of reversible languages. *The Journal of the Association of Computing Machinery*, 29:741–765, 1982.

[4] Lenore Blum and Manuel Blum. Towards a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

[5] Matthew de Brecht, Masanori Kobayashi, Hiroo Tokunaga and Akihiro Yamamoto. Inferability of closed set systems from positive data. In *New Frontiers in Artificial Intelligence, Proceedings of the 20th Annual Conference of The Japanese Society for Artificial Intelligence, Revised Selected Papers*, volume 4384 of *Lecture Notes in Computer Science*, pages 265–275. Springer, June 2007.

[6] Matthew de Brecht and Akihiro Yamamoto. Mind change complexity of inferring unbounded unions of restricted pattern languages from positive data. *Theoretical Computer Science*, 411(7–9):976–985, 2010.

[7] Matthew de Brecht and Akihiro Yamamoto. Topological properties of concept spaces (full version). *Information and Computation*, 208(4):327–340, 2010.

[8] John Case and Chris Lynes. Machine inductive inference and language identification. In *Proceedings of the Nineth International Colloquium on Automata, Languages and Programming*, volume 140 of *Lecture Notes in Computer Science*, pages 107–115. Springer, 1982.

[9] Chi Tat Chong, Theodore A. Slaman and Yue Yang. Π^1_1-conservation of combinatorial principles weaker than Ramsey's Theorem for pairs. *Advances in Mathematics*, 230(3):1060–1077, 2012.

[10] Chi Tat Chong, Theodore A. Slaman and Yue Yang. The metamathematics of stable Ramsey's theorem for pairs. *The Journal of the American Mathematical Society*, 27:863–892, 2014.

[11] Damir D. Dzhafarov and Carl Mummert. On the strength of the finite intersection principle. *Israel Journal of Mathematics*, 196:345–361, 2013.

[12] Harvey Friedman. Some systems of second order arithmetic and their use. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1974*, volume 1, pages 235–242, 1974.

[13] E. Mark Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[14] Susumu Hayashi. Mathematics based on incremental learning — excluded middle and inductive inference. *Theoretical Computer Science*, 350:125–139, 2006.

[15] Dennis Hirschfeldt. *Slicing the Truth. On the Computable and Reverse Mathematics of Combinatorial Principles*, volume 28 of *Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*. World Scientific, 2014.

[16] Rupert Hölzl, Sanjay Jain and Frank Stephan. Inductive Inference and Reverse Mathematics. In *Proceedings of the 32nd International Symposium on Theoretical Aspects of Computer Science*, volume 30 of *Leibniz International Proceedings in Informatics*, pages 420–433. Schloss Dagstuhl — Leibniz-Zentrum für Informatik, 2015.

[17] Rupert Hölzl, Dilip Raghavan, Frank Stephan and Jing Zhang. Weakly Represented Families in the Context of Reverse Mathematics. Manuscript, May 2016.

[18] Sanjay Jain, Daniel Osherson, James Royer and Arun Sharma. *Systems That Learn: An Introduction to Learning Theory, Second Edition.* The MIT Press, Cambridge, Massachusetts, 1999.

[19] Carl Jockusch and Frank Stephan. A cohesive set which is not high. *Mathematical Logic Quarterly*, 39:515–530, 1993. Erratum in [20].

[20] Carl Jockusch and Frank Stephan. Correction to "A cohesive set which is not high". *Mathematical Logic Quarterly*, 43:569, 1997.

[21] Satoshi Kobayashi. Approximate identification, finite elasticity and lattice structure of hypothesis space. Technical Report CSIM 96–04, Department of Computer Science and Information Mathematics, University of Electro-Communications, Tokyo, 1996.

[22] Donald A. Martin. Classes of recursively enumerable sets and degrees of unsolvability. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 11:295–310, 1966.

[23] Daniel Osherson, Michael Stob and Scott Weinstein. *Systems That Learn. An Introduction to Learning Theory for Cognitive and Computer Scientists*. The MIT Press, Cambridge, Massachusetts, 1986.

[24] Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Cambridge University Press, second edition, 2009.

[25] Theodore A. Slaman. Σ_n-bounding and Δ_n-induction. *Proceedings of the American Mathematical Society*, 132(8):2449–2456, 2004.

[26] Keith Wright. Identification of unions of languages drawn from an identifiable class. In *Proceedings of the Second Annual Workshop on Computational Learning Theory*, pages 328–333. Morgan Kaufmann, 1989.

[27] Thomas Zeugmann, Steffen Lange and Shyam Kapur. Characterizations of monotonic and dual monotonic language learning. *Information and Computation*, 120(2):155–173, 1995.