Inductive Inference and Reverse Mathematics*

Rupert Hölzl[1], Sanjay Jain[2], and Frank Stephan[3]

1 Department of Mathematics, National University of Singapore, S17, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore, r@hoelzl.fr
2 Department of Computer Science, National University of Singapore, COM2, 15 Computing Drive, Singapore 117417, Republic of Singapore, sanjay@comp.nus.edu.sg
3 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

Abstract

The present work investigates inductive inference from the perspective of reverse mathematics. Reverse mathematics is a framework which relates the proof strength of theorems and axioms throughout many areas of mathematics in an interdisciplinary way. The present work looks at basic notions of learnability including Angluin's tell-tale condition and its variants for learning in the limit and for conservative learning. Furthermore, the more general criterion of partial learning is investigated. 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 domination and induction strength.

1998 ACM Subject Classification F.4.1 Mathematical Logic.

Keywords and phrases reverse mathematics, recursion theory, inductive inference, learning from positive data

Digital Object Identifier 10.4230/LIPIcs.STACS.2015.420

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 easily, but not vice versa. In the 1970s, Friedman [11] proposed a framework that formalises this intuition and allows gauging the different strengths of theorems that can be 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 with regards to logical implication, then we need to look in 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 subject has

* R. Hölzl was fully and S. Jain and F. Stephan partially supported by NUS/MOE grant R146-000-184-112 (MOE2013-T2-1-062); furthermore, S. Jain is partially supported by NUS grant C252-000-087-001.

32nd Symposium on Theoretical Aspects of Computer Science (STACS 2015).
Editors: Ernst W. Mayr and Nicolas Ollinger; pp. 420–433
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

developed well since its inception, in particular thanks to many substantial contributions made by Simpson and his students [19]. 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 [14] and Simpson [19] which are convenient resources for the topic and give many references.

In the practice of reverse mathematics we will look at proper subsets of the axioms of second order arithmetic and will investigate the properties of possible models of these axiom sets. Such a model will be 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 join and Turing reduction. Furthermore, the axioms ensure that the system satisfies Σ_1-induction with parameters from $\mathcal{S}$; in particular, even in nonstandard models of RCA_0 all numbers of the form $\max_{i<n} f(i)$ exist for all functions $f \in \mathcal{S}$.

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 has 0 as the 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 number x which is not equal to $y + 1$ for some other number y;
- The addition $+$ is inductively defined from the successor by $x + 0 = x$ and $x + (y + 1) = (x + y) + 1$;
- The ordering $<$ is definable from $+$ by $x < y \Leftrightarrow \exists z\ [x + z + 1 = y]$;
- The multiplication is inductively defined from the addition 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 satisfies for all e the implication $[\forall d < e\ (d \in I)] \Rightarrow e \in I$ then I is equal to M;
- The set $\mathcal{S}$ contains $\emptyset$ and all sets which are recursive in the model $(M, +, <, \cdot, 0, 1)$;
- The second order model is a Turing ideal, that is, if $I, J \in \mathcal{S}$ then $I \oplus J = \{i+i : i \in I\} \cup \{j+j+1 : j \in J\}$ is also a member of $\mathcal{S}$ and, furthermore, if $I \in \mathcal{S}$ and J can be obtained from I by both a Σ_1-definition and a Π_1-definition then $J \in \mathcal{S}$.

Note that the last statement ensures that $\mathcal{S}$ is closed under join and Turing reducibility.

The model which 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. In particular if M is the standard model of natural numbers, then there is also the model where $\mathcal{S}$ is the power set of M; for nonstandard models, the power-set of M cannot be a model as it fails the induction axiom. There also exist many intermediate models between those two extremes. When M is the standard model of the natural numbers, then $(M, +, \cdot, <, 0, 1, \mathcal{S})$ is called an ω-model and due to their well-behavedness (compared to nonstandard models), they are better understood than nonstandard models in reverse mathematics. However, various complicated results in reverse mathematics were only obtained through the use of nonstandard models [8, 9].

As we will show in this article, many results in inductive inference relate to the following three axioms from reverse mathematics:

- The axiom DOM which says that for every weakly represented family of functions in $\mathcal{S}$ (defined below) there exists a function in $\mathcal{S}$ growing faster than all members of the family;
- The axiom ACA_0 which says that the class $\mathcal{S}$ is closed under Turing jump;

- The axiom $\mathrm{I}\Sigma_2$ which postulates that every Σ_2 set I definable using parameters from $\mathcal{S}$ satisfies the induction axiom: if $\forall e\ [[\forall d < e\ [d \in I]] \Rightarrow e \in I]$ then $I = M$.

Note that $\mathrm{I}\Sigma_2$ is satisfied for all standard models of the natural numbers; however, if M is a nonstandard model, assuming $\mathrm{I}\Sigma_2$ is a nontrivial constraint. By identifying a function with its graph we can also informally talk about the functions from M to M that exist in the model (that is, whose graphs are 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 classical 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 exists in a model, or what additional axiom – say, for example, comprehension for Σ_2^0 formulas – is needed to ensure its existence. We are then allowed to apply these additional axioms relative to any object X already existing in the model, no matter if X is recursive in the classical sense. That is, as soon as we know that X is guaranteed to exist in $\mathcal{S}$, we are allowed to take advantage of it, so for our purposes it is as good as recursive.

In this article we propose to apply the methodology of reverse mathematics to the field of inductive inference. We would like to point out that articles by de Brecht and Yamamoto [5] and by Hayashi [13] pursue the same idea, but in ways that differ from our approach and from each other. We proceed with defining central notions and analysing basic results of the field of inductive inference [1, 2, 3, 6, 7, 12, 15, 17, 21]. We will in particular study Angluin's tell-tale condition for learnability and related results. In this context the notion of finiteness of a set is of high importance as Angluin's tell-tale sets are finite. We 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 a subset of M to mean that the subset of M has an upper bound and "infinite" to mean that no such bound exists. Furthermore, it should be noted that "finite sets" are always considered to be "finite sets contained in $\mathcal{S}$" and that they are precisely those sets E for which there is a member $e \in M$ with $e = \sum_{d \in E} 2^d$.

Furthermore, we use Cantor's pairing function $\langle x, y\rangle = (x + y) \cdot (x + y + 1)/2 + y$ and extend it appropriately to triples and quadruples and so on. Now we code a family of sets $\{A_e\}_{e \in M}$ using a single set A by defining $x \in A_e \Leftrightarrow \langle e, x\rangle \in A$. Such families of sets are called uniformly represented families. Similarly one can define a uniformly represented family of functions by $F_e(x) = F(\langle e, x\rangle)$ using one representation function. This notion was generalised to the notion of weakly represented families of functions as follows [18]. Assume that a representation set $A \in \mathcal{S}$ satisfies the following conditions on quadruples:

- For all e, x, y, z, y', z': If $\langle e, x, y, z\rangle, \langle e, x, y', z'\rangle \in A$ then $y = y'$ and $z = z'$;
- If $\langle e, x, y, z\rangle \in A$ and $x' < x$ then there exist y' and z' such that $\langle e, x', y', z'\rangle \in A$ and $\langle e, x', y', z'\rangle < \langle e, x, y, z\rangle$.

The intention behind the second condition is to ensure that the coding quadruples for each function appear in the family in order ascending in the function argument x, even if the function is not monotone; for this purpose we use the fourth component of the quadruples as padding parameter. The first condition ensures that for each e, x there is at most one code defining $F_e(x)$. An index is invalid if there is some x where $F_e(x)$ is not defined. Hence the set $D = \{e \colon \forall x\, \exists y, z\, [\langle e, x, y, z\rangle \in A]\}$ is the index set of functions in the weakly represented family and for $e \in D$, $F_e(x)$ is the unique y such that $\langle e, x, y, z\rangle \in A$ for some z. The family $\{F_e\}_{e \in D}$ is called the weakly represented family defined by A and every function F_e, $e \in D$, is a function in the given second order model $(M, +, \cdot, <, 0, 1, \mathcal{S})$. Furthermore, in the case that all functions F_e in the family are $\{0, 1\}$-valued, they can also be viewed as the characteristic functions of a weakly represented family of sets and might be denoted with A_e rather than F_e. Note that Dzhafarov and Mummert [10] have considered the more general concept of *enumerated families* of sets.

In some of the proofs in this article we will make statements of the form "$A_e(x)$ can be retrieved from A in time less than s." By this we mean that the code number $c = \langle e, x, A_e(x), z\rangle \in A$ is bounded by s. The intuition of time is explained by the fact that $A_e(x)$ or the padding parameter z could be very large, so that it will depend on c how far we need to search in A to determine the function value $A_e(x)$ for a given x.

In the reverse mathematics setting, we will of course only work with representation sets A as above that exist in the given model of second order arithmetic $\mathcal{S}$. In the case of uniformly represented families then also their index set D as above must exist in $\mathcal{S}$. But note the important fact that for families that are only weakly represented, this will typically not be the case, that is, D is usually not required to exist in $\mathcal{S}$, only A always exists in $\mathcal{S}$. For this reason, we need to be careful in this article when working with families of functions, because a learner (which has to be a function in $\mathcal{S}$ from finite sequences of elements of $M \cup \{\#\}$ to M) may conjecture members of M that are not members of D, as at the time of the conjecture it cannot know whether a particular member of M is a valid index or not. Note that weakly represented families can be much more general than uniformly represented families; for example, for a fixed member $A \in \mathcal{S}$, the family of all A-recursive functions is weakly representable but in general not uniformly representable.

We now turn to the more formal notations from learning theory. Note that the families defined above correspond to the classes of possible learning targets in learning theory. The general scenario is that one possible learning target is presented to the learner in 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 a 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 in the case of nonstandard models they may differ.

▶ **Definition 1.** *A* text *for a set $A \in \mathcal{S}$ is a function $T\colon M \to M \cup \{\#\}$ in $\mathcal{S}$ such that $\{T(n)\colon n \in M \wedge T(n) \neq \#\} = A$. We call $\#$ the pause symbol. Without loss of generality we assume that $T(x) \in \{0, 1, \ldots, x\} \cup \{\#\}$.*

The pause symbol "#" is a padding symbol that carries no information and is useful to give a text for empty set. Again as usual we will write M^* for the set of finite sequences over $M \cup \{\#\}$. These can be thought of as the prefixes of texts. Take note that the word "finite" needs to be understood in the reverse mathematics sense discussed above. M^* can be represented by some canonical indexing, where each finite sequence σ is represented by the canonical index of the set $\{\langle x, 0\rangle\colon \sigma(x) = \#\} \cup \{\langle x, y+1\rangle\colon \sigma(x) = y\}$. One can prove by induction over a text that such canonical indices exist for every prefix of a text.

▶ **Definition 2** (Angluin [2]; Gold [12]; Osherson, Stob and Weinstein [17])**.** *Let $\{A_e\}_{e\in D}$ be a uniformly or weakly represented family and let $\{B_e\}_{e\in E}$ be a hypothesis space such that $\{A_e\}_{e\in D} \subseteq \{B_e\}_{e\in E}$. A* learner *is a function $L\colon M^* \to M$, where the elements of M^* are represented by canonical indices.*

A learner L learns in the limit *a family $\{A_e\}_{e\in D}$ if for every $e \in D$ and every text T for A_e the learner outputs a sequence of hypotheses $e_n = L(T(0)\ldots T(n))$ such that, for some n, for all $m \geq n$, each hypothesis e_m is equal to e_{m+1} and $e_m \in E$ and $B_{e_m} = A_e$.*

A conservative *learner never makes an unjustified mind change. So if $n < m$ and $e_n \neq e_m$ then either $e_n \notin E$ or there exists $k \leq m$ with $T(k) \in M - B_{e_n}$. Conservative learning then requires learning in the limit by a conservative learner.*

A learner partially learns *the family* $\{A_e\}_{e \in D}$ *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* $\forall m\, \exists n > m\, [d = e_n]$ *and, furthermore, this* d *satisfies* $d \in E$ *and* $B_d = A_e$.

Note that often $B_e = A_e$ for all e and $E = D$, that is, the original family is used as hypothesis space. 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. If invalid texts are presented to the learner, it may output numbers that are not actually in the set D of valid indices. Partial learning is a more general learning notion which, in the classical setting, allows learning the family of all r.e. sets.

Due to space constraints the proofs of most of the results in this article have been omitted.

2 Angluin's Condition

Angluin [2] gave a fundamental condition for the learnability of so-called indexed families of sets. These are families of sets such that there exists a computable two-place function F which on input (e, x) outputs 1 if $x \in L_e$ and 0 if $x \notin L_e$. As F works for all e, the closest equivalent to indexed families in the area of reverse mathematics are uniformly represented families. Angluin's condition (also called Angluin's tell-tale condition/criterion) says that one can learn an indexed family from positive data in the limit if and only if one can enumerate for each member A_e of the family a finite tell-tale subset B_e of A_e such that there is no other member A_d of the family with $B_e \subseteq A_d \subset A_e$. In reverse mathematics, it is difficult to handle finite sets, therefore one mostly represents them by canonical indices. However, for the tell-tale sets it is sufficient to consider bounds (called tell-tale bounds) b_e for each A_e such that there is no A_d with $A_e \cap \{0, 1, \ldots, b_e\} \subseteq A_d \subset A_e$.

▶ **Definition 3.** *Let a weakly represented family* $\{A_e\}_{e \in D}$ *with index set* D *be given:*
1. *The family* satisfies Angluin's condition in general *iff for each* $e \in D$ *there is a bound* b_e *such that there is no* $d \in D$ *with* $A_e \cap \{0, 1, \ldots, b_e\} \subseteq A_d \subset A_e$*;*
2. *The family* satisfies Angluin's condition in the limit *iff there is a two-place function* $g \in \mathcal{S}$ *such that for every* $e \in D$ *the values* $g(\langle e, 0\rangle), g(\langle e, 1\rangle), \ldots$ *approximate from below a bound* b_e *such that there is no* $d \in D$ *with* $A_e \cap \{0, 1, \ldots, b_e\} \subseteq A_d \subset A_e$*;*
3. *The family* satisfies Angluin's condition 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* $A_e \cap \{0, 1, \ldots, g(e)\} \subseteq A_d \subset A_e$.

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 [3] established the existence of so-called locking-sequences; that is, whenever a learner learns a language X there is a finite sequence of elements in X such that, after having processed this sequence, the learner conjectures a hypothesis which will not be changed on any subsequent data drawn from $X \cup \{\#\}$. Blum and Blum's proof can easily be modified to carry over to the reverse mathematics setting; it then proves the following statement.

▶ **Theorem 4.** *RCA_0 proves the following: Suppose a weakly represented family* $\{A_e\}_{e \in D}$ *and a learner* L *are given such that for every* $e \in D$ *and every text* T *for* A_e*,* L *converges on* T *to an index* $d \in D$ *with* $A_d = A_e$*. Then, there is a procedure which for every index* $e \in M$ *converges in the limit to a finite sequence (represented by a code); in the case that* $e \in D$*, this sequence is a locking sequence for* A_e.

This existence of locking-sequences then shows that every weakly represented family which is learnable in the limit must satisfy Angluin's tell-tale condition with a general bound: the bound is simply the largest element contained in the locking sequence. Section 3 will address the question of which of the above variants of Angluin's tell-tale condition is sufficient for learning all weakly represented families satisfying it. 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 uniformly represented families which, as mentioned before, are the closest equivalent in reverse mathematics to the indexed families that Angluin studied. The difference is that Angluin's families are *actually* uniformly recursive, while our uniformly represented families are only uniformly recursive relative to the parameter A representing them. This corresponds to the paradigm described above that in the reverse mathematics context often all sets in $\mathcal{S}$ are treated *as if* they were recursive. As we will show, for uniformly represented families, the degree of effectiveness of the bound in Angluin's condition is crucial. Section 5 then looks at sufficient criteria for learning from the classical theory and shows that in reverse mathematics they work for uniformly represented families as well. However, for weakly represented families we will again require the axiom DOM. In Section 6 we will study partial learning.

3 Learnability of Weakly Represented Families

As mentioned above, the counterpart of the indexed families studied by Angluin are the uniformly represented families in reverse mathematics. So it is not surprising that to prove similar results for families that are represented in a less accessible way, such as weakly represented families, we will need an additional assumption on the second order model. This assumption is the axiom DOM which will turn out to be equivalent to saying that every weakly represented family satisfying Angluin's condition is learnable in the limit. The axiom DOM says that every weakly represented family of functions is dominated by a single function in $\mathcal{S}$. Note that Adleman and Blum [1] showed that one can learn all classes of graphs of recursive functions (which all satisfy Angluin's condition) iff one has access to a dominating function as an oracle. The axiom DOM now enforces that for every weakly represented family of functions there is such a dominating function in $\mathcal{S}$; this function can therefore be used by the learner (which also has to be an object in $\mathcal{S}$).

▶ **Theorem 5.** *Over RCA_0, the following conditions are equivalent:*

1. *The axiom DOM holds, that is, for every weakly represented family $\{F_e\}_{e\in D}$ of functions there is a function $f \in \mathcal{S}$ dominating this family in the sense that $\forall e \in D\ \exists x\ \forall y > x\ [F_e(y) < f(y)]$;*
2. *The index set of every weakly represented family can be approximated in the limit;*
3. *Every weakly represented family satisfying Angluin's condition effectively can be learnt in the limit;*
4. *Every weakly represented family satisfying Angluin's condition in the limit can be learnt in the limit;*
5. *Every weakly represented family satisfying Angluin's condition generally can be learnt in the limit.*

Proof. $1 \Rightarrow 2$: Let $\{F_e\}_{e\in D}$ be a weakly represented family with representation set A. Then the set of the functions G_e for $e \in M$ which assign to x the minimum tuple (if it exists) $\langle e, x, y, z\rangle \in A$ also forms a weakly represented family, and G_e is total iff $e \in D$. Thus the index set of this weakly represented family is also D.

By assumption there is a function f dominating all $\{G_e\}_{e \in D}$. Now it holds that $e \in D$ if and only if, for almost all numbers x, there are pairwise distinct elements of the form $\langle e, 0, y_0, z_0 \rangle, \ldots, \langle e, x, y_x, z_x \rangle \in A \cap \{0, 1, \ldots, f(x)\}$. This is because on one hand, if $e \in D$, the existence of these 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 A does not contain *any* element of the form $\langle e, x, \cdot, \cdot \rangle$, so in particular there is no sequence as above.

Now one defines a function g by letting $g(e, x) = 1$ iff there are elements of the form $\langle e, 0, y_0, z_0 \rangle, \ldots, \langle e, x, y_x, z_x \rangle \in A \cap \{0, 1, \ldots, f(x)\}$, 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 ⇒ 1: Assume that a weakly represented family $\{F_e\}_{e \in D}$ has an index set D which is approximated by g in the limit and has the representation set A. Then one can construct the following function f:

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

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

The minimum is taken over only finitely many conditions (in the square brackets) and for every condition individually the minimal t can be computed from e and x (using the same parameter set in the second order model as for the computation of A). Therefore, using Σ_1-induction, $f(x)$ exists as the maximum over the t's that are minimal for the individual conditions (for each $e \leq x$). Note that the "$+x$" in the definition of f ensures that wrong behaviour of g during the first finitely many approximation stages is ignored in the limit.

The function f dominates each function F_e with $e \in D$, as for that 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 and 2 ⇒ 5: Let $\{A_e\}_{e \in D}$ be a weakly represented family satisfying Angluin's tell-tale condition generally. Furthermore, by the second condition 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 and bound b a function $G_{e,b}$ such that $G_{e,b}(x)$ is the first $t \geq x$ 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 from the representation set within time t and either $x' \in A_d - A_e$ or $x' \in A_e - A_d \wedge x' \leq b$;
- The values of A_d and A_e up to x have been retrieved from the representation set within time t and $A_d(x') = A_e(x')$ for all $x' \leq x$.

These three conditions search for either e not being a valid index, or d not being a valid index, or x' witnessing that A_d is not a subset of A_e, or x' being an element of the tell-tale set of A_e that is not in A_d, or A_d being equal to A_e up to x. Note that the function $G_{e,b}$ is total for those b which are valid bounds for F_e; thus the index set of the family $\{G_{e,b}\}_{(e,b) \in D'}$ is the set of all (e, b) such that either $e \notin D$ or b is a valid general bound for Angluin's condition with respect to A_e. Now there is a function f dominating all the $G_{e,b}$ in the weakly represented family. Note that whenever $G_{e,b}$ is in this family then so is $G_{e,b+1}$ and $G_{e,b}(x) \geq G_{e,b+1}(x)$ for all x.

Without loss of generality one can assume that any number x does not appear in the text earlier than at stage x – this is achieved by inserting pause symbols into the text at all places where needed. Let $(e_0, b_0), (e_1, b_1), \ldots$ be a sequence of pairs in which each pair of index and bound appears infinitely often. The learner has the initial counter value 0, the initial hypothesis e_0 and initial bound b_0. Assume that after processing s items, the learner has the counter n, the previous hypothesis e_n and the bound b_n. To determine whether an update to these parameters is needed, the learner now checks whether they satisfy all of the following conditions:

- We have $g(e_n, u+s) = 1$ for all $u \leq f(s)$ and all values $A_{e_n}(x)$ for $x \leq s$ can be retrieved from the representation set within time $f(s)$;
- It holds that $G_{e_n,b_n}(s)$ is defined within $f(s)$ steps;
- All data x with $x \leq b_n \wedge x \in A_{e_n}$ have been observed in the text so far;
- No datum x with $x \notin A_{e_n}$ has been observed in the text so far.

If (e_n, b_n) satisfies all these conditions then the learner keeps the counter n, hypothesis e_n and the bound b_n, else the learner changes the counter to $n+1$, the hypothesis to e_{n+1} and the bound to b_{n+1}. Assume that the learner converges to an incorrect hypothesis e_n or a hypothesis with an incorrect bound b_n, then one of the following happens at some future stage s eventually:

- It holds that $g(e_n, u+s) = 0$ for some $u \leq f(s)$ (in the case that e_n is not a valid index);
- $G_{e_n,b_n}(s)$ is not defined (in the case that the bound b_n is invalid and that there is a $d \leq s$ inside D discovered with $A_e \cap \{0, 1, \ldots, b_n\} \subseteq A_d \subset A_e$);
- Not all data in $A_e \cap \{0, 1, \ldots, b_n\}$ have shown up in the text or some datum outside A_e has shown up in the text (in the case that the index and the bound are valid but that the hypothesis is not the correct one).

All these conditions imply that the hypothesis will be updated to e_{n+1} (and the bound to b_{n+1}) in contradiction to the assumption. The next possibility is that the learner would infinitely often have a counter value n such that (e_n, b_n) is some fixed correct pair (e, b). As the function f dominates $G_{e,b}$, it holds for all sufficiently large s where the current (e_n, b_n) is equal to (e, b) that all four conditions from the above update test are satisfied and that therefore the current (e_n, b_n) will be kept and n will not be incremented. So the learner indeed converges to the correct hypothesis e. As the function h from s to the n currently processed is increasing and grows each step at most by one and is a member of $\mathcal{S}$, this function h is either eventually constant or has range M; hence the above two cases (converging to a wrong hypothesis or taking one correct hypothesis infinitely often) are exhaustive and the learner is correct.

$5 \Rightarrow 4 \Rightarrow 3$. This follows from the definition.

$3 \Rightarrow 2$. Let $\{F_e\}_{e \in D}$ be any weakly represented family of functions (as represented by the set $F \in \mathcal{S}$). Now define a new weakly represented family $A_{\langle e,s\rangle}$ of sets such that $A_{\langle e,0\rangle} = \{\langle e, x\rangle \colon x \in M\}$ in case that $e \in D$ and let $\langle e, 0\rangle$ be an invalid index in case that $e \notin D$. Let $A_{\langle e,s+1\rangle} = \{\langle e, x\rangle \colon x \leq s\}$ in case that $s = \max\{\langle e, u, y, z\rangle \in F \colon u, y, z \in M\}$ and let $\langle e, s+1\rangle$ be an invalid index otherwise. Note that for each e, there is a unique s such that $\langle e, s\rangle$ is a valid index: we denote the corresponding unique $A_{\langle e,s\rangle}$ as A_e.

Assume now that this weakly represented family is learnable in the limit. Then, uniformly in e, there is a text T_e which contains all the pairs $\langle e, x\rangle$ such that for some $\langle e, u, y, z\rangle \geq x$, $\langle e, u, y, z\rangle \in F$. This text T_e is a text for A_e. The learner converges on T_e to some index d in the limit. By simulating the learner one can make a function g such that

- $\lim_{t\to\infty} g(e, t)$ converges to 0 in the case that the learner converges on the text T_e to an index d for a set which does not contain $\langle e, x\rangle$ for some $x \in M$,
- $\lim_{t\to\infty} g(e, t)$ converges to 1 in the case that the learner converges on the text T_e to an index d for a set containing $\langle e, x\rangle$ for each $x \in M$.

The first case occurs iff $A_e = A_{\langle e,s+1\rangle}$ for some s and the second case occurs iff $A_e = A_{\langle e,0\rangle}$. Here the first case coincides with $e \notin D$ and the second with $e \in D$. Thus g is correct. ◀

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. Therefore one might be inclined to choose a more comprehensive but somehow easier hypothesis space. However, in the proof of Theorem 5 ($3 \Rightarrow 2$) 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.

Raghavan, Stephan and Zhang [18] investigate the strength of DOM. They show that under RCA_0 and $I\Sigma_2$, DOM implies COH but not vice versa. This result is the counterpart to the recursion-theoretic result that every high Turing degree contains a cohesive set. Furthermore, for ω-models, there are also connections to set-theoretically motivated axioms. For example, DOM is true iff MAD is false [18]. Here MAD is the statement that there exists a *maximal almost disjoint family*, that is, a weakly represented family of sets $\{A_e\}_{e\in D}$ 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 Uniformly Represented Families

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

▶ **Theorem 6.** *Over RCA_0, a uniformly represented family is learnable in the limit if and only if it satisfies Angluin's condition in the limit.*

One might ask when a learner exists in the case of general bounds in place of limit bounds.

▶ **Theorem 7.** *Over RCA_0, DOM holds iff every uniformly represented family satisfying Angluin's condition with a general bound is learnable in the limit.*

Angluin [2] introduced the notion of conservative learning by requiring that a conservative learner only makes a mind change (that is, updates its hypothesis) if some datum observed so far is not contained in the previously conjectured set. Conservative learners do, 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 that could explain the data observed so far. This requirement enforces the effective version of Angluin's condition and may require that the learner use a different hypothesis space than the family to be learnt. Such a hypothesis space is itself a family which needs to contain all sets from the family to be learnt but possibly also other sets.

▶ **Theorem 8.** *Over RCA_0, a uniformly represented family $\{C_e\}_{e\in M}$ is conservatively learnable using some hypothesis space $\{A_e\}_{e\in M}$ if and only if $\{C_e\}_{e\in M}$ is contained in some uniformly represented family $\{B_e\}_{e\in M}$ (possibly different from $\{A_e\}_{e\in M}$) which satisfies Angluin's condition effectively.*

One might also ask in which cases every uniformly represented family satisfying Angluin's bound only in general is conservatively learnable. By Theorem 8 this only happens when for every uniformly represented family it is equivalent whether it satisfies Angluin's bound in general or effectively. This then allows coding the halting problem into such a family, and one obtains the following corollary. Finally, over ACA_0, the index set D of any weakly represented family is in $\mathcal{S}$; so the result carries over to weakly represented families.

▶ **Corollary 9.** *Over RCA_0, the following statements are equivalent:*

1. *ACA_0 (that is, every set arithmetically definable from parameters in $\mathcal{S}$ is also in $\mathcal{S}$ and, in particular, $\mathcal{S}$ is closed under the Turing jump);*
2. *Every uniformly represented family satisfying Angluin's tell-tale condition with a general bound is conservatively learnable;*
3. *Every weakly represented family satisfying Angluin's tell-tale condition with a general bound is conservatively learnable.*

5 Sufficient Criteria

Angluin [2] looked at sufficient criteria for learning. In the reverse mathematics setting, all these criteria can be proven to be sufficient over RCA_0 for uniformly represented families; for weakly represented families, the additional axiom DOM is again needed and sufficient to build the learners. The first of these criteria considered is finite thickness.

▶ **Theorem 10.** *Say that a family $\{A_e\}_{e \in D}$ 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.*

The property of finite thickness has been strengthened to finite elasticity [20]. Finite elasticity mainly says that one cannot construct a text which in each step makes a concept inconsistent that was consistent before. Abstracting from the requirement that this happens in every step, one can also formulate this the other way round: A family has finite elasticity if and only if for every text there is a prefix of the text such that every concept inconstent with the full text is also inconsistent with this prefix.

▶ **Theorem 11.** *Say that a family $\{A_e\}_{e \in D}$ 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 $e \in D$, $range(\sigma) \subseteq A_e \Rightarrow range(T) \subseteq A_e$.*

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.*

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

Kobayashi [4, 16] considered another sufficient learnability criterion which is a further strengthening of the property of finite elasticity: A class is learnable 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 class. This learnability condition was proven in the context of indexed families and holds without any effectivity requirement on finding this finite subset. One can carry it over to uniformly represented and weakly represented families as follows.

▶ **Theorem 12.** *Say that a family $\{A_e\}_{e \in D}$ 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 $A_e \cap \{0, 1, \dots, b\} \subseteq A_d \Rightarrow A_e \subseteq A_d$.*

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 Angluin's tell-tale criterion, as the former condition enforces $A_e \cap \{0, 1, ..., b\} \subseteq A_d \Rightarrow A_e \subseteq A_d$ while Angluin's tell-tale criterion merely enforces $A_e \cap \{0, 1, ..., b\} \subseteq A_d \Rightarrow A_d \not\subset A_e$.

6 Partial Learning

Osherson, Stob and Weinstein [17] 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 concept allows to learn all classes of r.e. languages, provided that the hypothesis space permits padding. Our proofs of the corresponding results in reverse mathematics depend on the axiom $I\Sigma_2$ which, for example, proves that every set in a weakly represented family has a least index. It is unknown whether this is an inherent requirement for obtaining the statements, or one more involved arguments could dispense with.

▶ **Theorem 13.** *Over RCA_0, a weakly represented family $\{A_d\}_{d \in D}$ is partially learnable if and only if there is a further 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.

▶ **Theorem 14.** *Over RCA_0, every uniformly represented family is partially learnable.*

▶ **Theorem 15.** *Over RCA_0 and $I\Sigma_2$, for every weakly represented family $\{A_e\}_{e \in D}$, there is a partial learner using the weakly represented family $\{B_{\langle e,b \rangle}\}_{e \in D, b \in M}$ with $B_{\langle e,b \rangle} = A_e$ for all $e \in D$, $b \in M$ as hypothesis space.*

Proof. Let a weakly represented family $\{A_e\}_{e \in D}$ be given, let A be its representation set and let $X \in \mathcal{S}$. Now consider the Σ_2 index 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 the case that X does not have a minimal index, the index set I satisfies for all e the property $(\forall d < e \, [d \in I] \Rightarrow e \in I)$ and then X does not have any index in the weakly represented family. Given the minimal index e of a member of the family, one can define for $d < e$ the uniform Σ_2 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 $d < e$. Now one defines the partial learner as follows: A hypothesis $\langle e, b \rangle$ is output at least n times if and only if there is $s \geq n$ such that the following conditions are satisfied:

- $A_e(0), A_e(1), \ldots, A_e(n)$ can be retrieved from A in time s;
- There is no $d < e$ such that for all $x \leq b$ the descriptions of $A_d(x)$ and $A_e(x)$ can be retrieved from A in time 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 descriptions of $A_d(x)$ and $A_e(x)$ can be retrieved from A in time s and such that $A_d(x) = A_e(x)$.

One can verify that on a text for a member X of the weakly represented family, exactly one pair $\langle e, b\rangle$ is output infinitely often and this is given by the least index e of X and the least bound b such that all $d < e$ satisfy that either $A_d(b)$ is not defined or there is $x \leq b$ with $A_e(x) \neq A_d(x)$. Thus the family is partially learnt by the given learner. ◀

The previous result establishes that, over RCA_0 and $\mathrm{I}\Sigma_2$, every member of a weakly represented family has a least index. This assumption is an essential ingredient of the learning algorithm and is equivalent to $\mathrm{I}\Sigma_2$ over RCA_0.

▶ **Proposition 16.** *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 Theorem 15. For the necessity assume that $\mathrm{I}\Sigma_2$ is not satisfied. Then there is a Σ_2 set I which is a proper subset of M and satisfies for all e that $[(\forall d < e\colon d \in I) \Rightarrow e \in I]$. As the set is Σ_2, there is a ternary $\{0,1\}$-valued function $g \in \mathcal{S}$ such that $e \in I \Leftrightarrow \exists n \in M\ \forall m \in M\ [g(e,n,m) = 1]$. Now define a weakly represented family such that every member of it is equal to M and its description A contains, for each e, inductively the pairs $\langle e, n, 1, z_n\rangle$ with $z_0 = 0$ and $z_{n+1} = z_n + m$ for the least m such that $g(e,n,m) = 0$. Now consider an arbitrary e.

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

It follows that the complement of I is the index set of the so constructed weakly represented family and this index set does not contain a minimal element by the choice of I. However, the index set contains only indices of the unique member M of the family, contradiction. ◀

▶ **Theorem 17.** *Over RCA_0, $\mathrm{I}\Sigma_2$ and DOM, every weakly uniform family can be partially learnt using the family itself as hypothesis space.*

Proof. Given a weakly represented family $\{A_e\}_{e\in D}$ with representation set A, one can 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 $A_e(x)$. The family of these G_e is dominated by some function $f \in S$. Now, the learner 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:

1. $A_e(0), A_e(1), \ldots, A_e(m)$ can be retrieved within time $f(m)$ from A;
2. There is no $d < e$ such that $A_d(0)$, $A_d(1)$, …, $A_d(m)$ can be retrieved within time $f(m)$ from A and such that $A_d(0) = A_e(0)$, $A_d(1) = A_e(1)$, …, $A_d(m) = A_e(m)$;
3. All numbers $x \leq m$ with $A_e(x) = 1$ have occurred within the first s members of the text;
4. No number $x \leq m$ with $A_e(x) = 0$ has occurred within the first s members of the text.

The assumptions are sufficient to prove that this partial learner indeed succeeds to partially learn the languages in the family; note that the first two conditions together with f being a dominating function enforce that only minimal indices – whose existence is ensured by $\mathrm{I}\Sigma_2$ – are output infinitely often and that the last two conditions enforce that a minimal index is output infinitely often iff it is correct. ◀

Acknowledgments. The authors would like to thank Chong Chi Tat and Yang Yue for interesting 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 Lenore Blum and Manuel Blum. Towards a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

4 Matthew de Brecht, Masanori Kobayashi, Hiroo Tokunaga, and Akihiro Yamamoto. Inferability of closed set systems from positive data. In *New Frontiers in Artificial Intelligence, JSAI 2006. Conference and Workshops, Tokyo, Japan, 5–9 June 2006, Revised Selected Papers*, volume 4384 of *Lecture Notes in Computer Science*, pages 265–275. Springer, 2007.

5 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.

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

7 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.

8 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.

9 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.

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

11 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.

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

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

14 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.

15 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.

16 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, 1996.

17 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.

18 Dilip Raghavan, Frank Stephan, and Jing Zhang. Weakly represented families in reverse mathematics. Manuscript, 2014.

19 Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Cambridge University Press, 2009.

20 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.

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