Learnability of Co-r.e. Classes

Ziyuan Gao[1] and Frank Stephan[2⋆]

[1] Department of Mathematics,
National University of Singapore, Singapore 117543, Republic of Singapore.
Email: ziyuan84@yahoo.com
[2] Department of Mathematics and Department of Computer Science,
National University of Singapore, Singapore 117543, Republic of Singapore.
Email: fstephan@comp.nus.edu.sg

Abstract. The object of investigation in this paper is the learnability of co-recursively enumerable (co-r.e.) languages based on Gold's [14] original model of inductive inference. In particular, the following learning models are studied: finite learning, explanatory learning, vacillatory learning and behaviourally correct learning. The relative effects of imposing further learning constraints, such as conservativeness and prudence on these various learning models are also investigated. Moreover, an extension of Angluin's [2] characterisation of identifiable indexed families of recursive languages to families of conservatively learnable co-r.e. classes is presented. In this connection, the paper considers the learnability of indexed families of recursive languages, uniformly co-r.e. classes as well as other general classes of co-r.e. languages. A containment hierarchy of co-r.e. learning models is thereby established; while this hierarchy is quite similar to its r.e. analogue, there are some surprising collapses when using a co-r.e. hypothesis space; for example vacillatory learning collapses to explanatory learning.

1 Introduction

The model of learning adopted in the present paper is built on that pioneered by Gold [14]. In this setting, a language is conceived as a set of strings over some fixed finite alphabet; a learner is modelled as an algorithmic machine that reads step by step example strings from a pre-assigned language and at each step, based on what it has read so far, outputs a conjecture on the language using a specified nomenclature. The selection of a suitable nomenclature, known formally as a hypothesis space, depends to some extent on the nature of the languages considered: if, for example, a language L is recursive, then a programme for a decision procedure on whether or not a given string belongs to L may be an adequate hypothesis for describing L. The learner is said to have successfully learnt a target language if it outputs a correct identification of the language at some point and thenceforth never returns to an incorrect conjecture. The success of a learner in identifying a language L may be contingent on the sort of string

⋆ Work is supported in part by NUS grant R252-000-420-112.

examples presented to it at each step, and also on the order of presentation; the ensuing discussion is confined to the case that only positive data is presented in the form of a text - an infinite sequence of strings that contains all strings of L and perhaps some pause symbols.

Gold's original model has since been generalised in various directions to encompass a large class of learnability notions and systems of learning. One may consider in particular the learnability of co-recursively enumerable classes in the limit. Accordingly, a language is modelled as a co-recursively enumerable class of strings, and a learner is represented by a recursive function mapping the set of finite sequences of natural numbers together with a pause symbol (#) into the set of natural numbers. As a motivation for studying learnability under this new scheme, it may be noted that several psychology experiments [11] have revealed that subjects tasked to identify certain conjunctive concepts generally tend to perform better when receiving positive instances rather than negative instances, even when both types of instances carry the same amount of information. At the same time, it has been observed [16] that the performance of subjects in *disjunctive* concept identification tasks is initially superior when they receive negative instances to that when they receive positive instances.

If the human mind is modelled as a recursive learning machine, and the extension of a concept is always assumed to be r.e., it is possible to interpret the learnability of languages with respect to a co-r.e. hypothesis space as a model for concept identification based on negative examples. On this view, it is hoped that the results in this paper may offer some insight on whether the human processing of negative instances differs from positive instances with regard to learnability of concepts, and whether there are precise settings under which the two kinds of data presentations yield identical learnability results. As was pointed out earlier, the choice of a hypothesis space may be a decisive factor in determining the learnability of a class. We shall take a fixed universal numbering of all co-recursively enumerable sets, whose existence is well-established [25], as a natural hypothesis space for the type of classes considered. This work sits between the setting of uniformly recursive languages considered by Angluin, Lange and Zeugmann [2, 18, 21, 22] and the inference using limiting recursive programs by Case, Jain and Sharma [10] or correction grammars by Carlucci, Case and Jain [8].

The present paper proceeds along two main lines of investigation: the classification of learning criteria according to their relative strengths, as well as the intrinsic characterisation of classes that are learnable under particular constraints. In regard to the latter objective, the paper focusses on characterising classes that are conservatively learnable with respect to a co-r.e. hypothesis space. The main contribution, Theorem 12, formulates this characterisation based on the notion of a *tell-tale set*, originally introduced by Angluin [2]: a finite set D is a tell-tale set for some language L if D is a subset of L, and there are no languages L' that are properly contained in L and which also contain D. The purpose of a tell-tale set is to provide sufficient evidence for a plausible conjecture based on a finite number of instances of the target language. As a model for learning scientific theories, the restriction to a co-r.e. hypothesis space of the learner is a natural realisation of Popper's [24] proposed criterion that a statement of an empirically scientific nature should be falsifiable. In fact, conservative learners using co-enumerable hypotheses spaces explicitly follow Popper's principle of scientific discovery: they revise hypotheses when they are proven to be false and they issue hypotheses such that they can be refuted whenever they are wrong, at least as long as the learner has only to deal with target sets from the class to be learnt.

More generally, learnability with respect to a co-r.e. hypothesis space appears to be quite a stringent criterion for a scientific law to be learnt recursively, as a learner is required to formulate in the limit a general rule enumerating counterexamples based on the positive data it has seen so far. The subsequent results obtained provide evidence that under fairly natural conditions, learnability using co-enumerable indices is indeed a sharper criterion than learnability using enumerable indices.

Notation 1. The notation and terminology from recursion theory adopted in this paper follow [25] in the main. The abbreviation r.e. shall be used for the term "recursively enumerable." Note that it suffices to consider r.e. and co-r.e. languages of natural numbers, since Cantor's pairing function permits a coding of all strings over some fixed finite or countable alphabet into the natural numbers; hence the r.e. languages range over subsets of the set $\mathbb{N} = \{0, 1, 2, \ldots\}$. A universal numbering of all partial-recursive functions is fixed as $\varphi_0, \varphi_1, \varphi_2, \ldots$. Given a set S, $\overline{S}$ denotes the complement of S, and S^* denotes the set of all finite sequences of elements from S. $\overline{W}_0, \overline{W}_1, \overline{W}_2, \ldots$ is a universal numbering of all co-r.e. sets, where W_e is the domain of φ_e. $\langle x, y\rangle$ denotes Cantor's pairing function, given by $\langle x, y\rangle = \frac{1}{2}(x + y)(x + y + 1) + y$. $W_{e,s}$ is an approximation to W_e; without loss of generality, $W_{e,s} \subseteq \{0, 1, \ldots, s\}$, and $\{\langle e, x, s\rangle : x \in W_{e,s}\}$ is primitive recursive; $W_e = \bigcup_s W_{e,s}$ and $W_{e,s} \subseteq W_{e,s+1}$ for all s. $\varphi_e(x) \uparrow$ means that $\varphi_e(x)$ remains undefined; $\varphi_{e,s}(x) \downarrow$ means that $\varphi_e(x)$ is defined, and that the computation of $\varphi_e(x)$ halts within s steps. $\mathbb{K}$ denotes the diagonal halting problem, $\{e : \varphi_e(e) \downarrow\}$. For any two sets A and B, $A \oplus B = \{2x : x \in A\} \cup \{2y + 1 : y \in B\}$. Analogously, $A \oplus B \oplus C = \{3x : x \in A\} \cup \{3y + 1 : y \in B\} \cup \{3z + 2 : z \in C\}$.

For any $\sigma, \tau \in (\mathbb{N} \cup \{\#\})^*, \sigma \preceq \tau$ if and only if $\sigma = \tau$ or τ is an extension of σ, $\sigma \prec \tau$ if and only if σ is a proper prefix of τ, and $\sigma(n)$ denotes the element in the nth position of σ, starting from $n = 0$. $Pref(\sigma)$ denotes the sequence of all prefixes, in lexicographic ordering, of σ. Given a number a and some fixed $n \geq 1$, denote by a^n the finite sequence $a \ldots a$, where a occurs n times. a^0 denotes the empty string. The concatenation of two strings σ and τ shall be denoted by $\sigma\tau$, and occasionally by $\sigma \circ \tau$.

Definition 2. Let $\{L_i\}_{i\in\mathbb{N}}$ be a family of languages.

i. $\{L_i\}_{i\in\mathbb{N}}$ is *uniformly recursive* if the set $\{\langle i, x\rangle : x \in L_i\}$ is recursive.
ii. $\{L_i\}_{i\in\mathbb{N}}$ is *uniformly recursively enumerable* if the set $\{\langle i, x\rangle : x \in L_i\}$ is recursively enumerable.
iii. $\{L_i\}_{i\in\mathbb{N}}$ is *uniformly co-recursively enumerable* if the set $\{\langle i, x\rangle : x \notin L_i\}$ is recursively enumerable.

Definition 3. A class $\mathcal{L}$ of recursive languages is said to be an *indexed family of recursive languages* if there exists a uniformly recursive family $\{L_i : i \in \mathbb{N}\}$ such that $\mathcal{L} = \{L_i : i \in \mathbb{N}\}$. $\mathcal{L}$ is said to be an *indexed family of r.e. (co-r.e.) languages* if there exists a uniformly r.e. (co-r.e.) family $\{L_i : i \in \mathbb{N}\}$ such that $\mathcal{L} = \{L_i : i \in \mathbb{N}\}$.

2 Learnability

Let $\mathcal{C}$ be a class of recursive, recursively enumerable or co-recursively enumerable sets. A *text* T_L for some L in $\mathcal{C}$ is a map $T_L : \mathbb{N} \to L \cup \{\#\}$, whose range contains every element of L. $T[n]$ denotes the string $T(0)T(1)\ldots T(n)$. A learner is a recursive function $M : (\mathbb{N} \cup \{\#\})^* \to \mathbb{N}$. The two main learning criteria studied in this paper are *explanatory* and *behaviourally correct* learning. In general, the subscript of the symbol for a learning criterion indicates the nature of the hypothesis space used; "r.e." and "co-r.e." would refer to r.e. and co-r.e. hypothesis spaces respectively, while "rec" would refer to a recursive hypothesis space.

i. [14] M is said to *explanatorily* ($Ex_{co-r.e.}$) learn $\mathcal{C}$ if, for each L in $\mathcal{C}$, and any corresponding text T_L for L, there is a number n for which $L = \overline{W}_{M(T_L[j])}$ whenever $j \geq n$, and for any $k \geq j$, $M(T_L[k]) = M(T_L[j])$.
ii. [6] M is said to *behaviourally correctly* ($BC_{co-r.e.}$) learn $\mathcal{C}$ if, for each L in $\mathcal{C}$, and any corresponding text T_L for L, there is a number n for which $L = \overline{W}_{M(T_L[j])}$ whenever $j \geq n$.

The next three definitions restrict the way the learner handles hypotheses.

Definition 4. i. [23] A recursive learner M is said to be *prudent* ($Prud$) if it learns the class $\{\overline{W}_{M(\sigma)} : \sigma \in (\mathbb{N} \cup \{\#\})^*\}$. In other words, a prudent learner M learns every set it conjectures.
ii. [2] A recursive learner M is said to *conservatively* ($Cnsv_{co-r.e.}$) learn $\mathcal{C}$ if it $Ex_{co-r.e.}$ learns $\mathcal{C}$ and if, given any two finite sequences $\sigma, \tau \in (\mathbb{N} \cup \{\#\})^*$ such that $M(\sigma) \neq M(\sigma\tau)$, there is a number x with $x \in \mathrm{range}(\sigma\tau) - \overline{W}_{M(\sigma)}$.
iii. [21] Let $\{L_i\}_{i \in \mathbb{N}}$ be a family of languages. $\{L_i\}_{i \in \mathbb{N}}$ is said to be *class-preservingly learnable* if there is a recursive learner M that learns $\{L_i\}_{i \in \mathbb{N}}$ such that $\{L_i\}_{i \in \mathbb{N}} = \{\overline{W}_{M(\sigma)} : \sigma \in (\mathbb{N} \cup \{\#\})^*\}$.

Remark 5. Throughout this paper, we often consider combinations of various learning contraints with different criteria; whenever there is no risk of ambiguity or contradiction, the relevant notations for the criteria are concatenated. For example, the criterion of prudent and conservative learning with respect to the hypothesis space $\{\overline{W}_0, \overline{W}_1, \overline{W}_2, \ldots\}$ is denoted by $PrudCnsv_{co-r.e.}$.

3 Characterisations of conservative learning

Angluin's [2] theorem furnishes a necessary and sufficient condition for an indexed family of recursive languages to be explanatorily learnable in the limit.

This is formally stated in terms of the notion of a finite *tell-tale set*, the main role of which is to aid a learner in steering clear of overgeneralisation; that is, conjecturing a set that properly contains the target language. Overgeneralising at some stage during a sequence of guesses is a potential pitfall for the learner, as it will not observe any counter-examples to its current conjecture. However, by ensuring that the data revealed eventually contain every member of the tell-tale set corresponding to its last conjecture, the learner can avoid being caught in such a situation.

Definition 6 (Angluin [2]). Let $\{L_i\}_{i\in\mathbb{N}}$ be a family of languages. A uniformly r.e. family of finite sets $\{H_i\}_{i\in\mathbb{N}}$ is said to be a *family of tell-tale sets* for $\{L_i\}_{i\in\mathbb{N}}$ if, for every $i, j \in \mathbb{N}$,

- $H_i \subseteq L_i$;
- if $H_i \subseteq L_j \subseteq L_i$, then $L_j = L_i$.

For each language L_i, H_i is said to be a *tell-tale set* for L_i. A class of languages for which a uniformly r.e. family of tell-tale sets exists is said to satisfy *Angluin's tell-tale condition*.

The necessity of Angluin's tell-tale condition for the learnability of indexed families follows from a more general property that learnable languages possess - the existence of locking sequences. This is stated below for the explanatory learning of co-r.e. sets, but the definition extends in a natural way to other kinds of learning criteria, as well as to r.e. sets.

Definition 7 (Blum and Blum [4]). Let M be a recursive learner and L a co-r.e. set learned by M. A finite sequence σ in $(L\cup\{\#\})^*$ is said to be a *locking sequence* for L if $\overline{W}_{M(\sigma)} = L$ and for all $\tau \in (L \cup \{\#\})^*, M(\sigma\tau) = M(\sigma)$.

Using the concept of a locking sequence, one can prove Angluin's theorem for indexed families of recursive languages:

Theorem 8 (Angluin [2]). *An indexed family of recursive languages $\mathcal{L} = \{L_i\}_{i\in\mathbb{N}}$ is explanatorily learnable if and only it it satisfies Angluin's tell-tale condition.*

In seeking to bridge the disconnect between two research areas, one in indexed families of recursive languages, and the other in general classes of r.e. languages, de Jongh and Kanazawa [9] later generalised Angluin's tell-tale condition to characterise indexed families of r.e. languages. In [2], Angluin also introduced the notion of conservativeness and observed that it is an important learnability property of indexed families. Here a learner is called *conservative* if and only if $M(\sigma\tau) \neq M(\sigma)$ only for such σ, τ where the range of $\sigma\tau$ contains some elements outside the hypothesis issued by $M(\sigma)$. Note that with this definition, every conservative learner is automatically explanatory. Zeugmann, Lange and Kapur [19] showed how to generalise Angluin's tell-tale condition to characterise conservative learnability of indexed families of recursive languages; this was later extended by de Jongh and Kanazawa [9] to the case of indexed families of r.e. languages. A family of finite sets $\{H_i\}_{i\in\mathbb{N}}$ is *uniformly recursively generable* if there is a total effective procedure g, which, on every input j, generates all elements of H_j and stops. Conservative learnability of an indexed family of recursive languages $\mathcal{L} = \{L_i\}_{i\in\mathbb{N}}$ can be realised if there exists a family of tell-tale sets $\{H_i\}_{i\in\mathbb{N}}$ for $\mathcal{L}$ which is uniformly recursively generable. More precisely, one has the following theorem.

Theorem 9 (Lange, Zeugmann and Kapur [19]). *Let $\mathcal{L}$ be an indexed family of non-empty recursive languages. Then $\mathcal{L}$ is conservatively learnable if and only if there is a hypothesis space $\mathcal{G} = \{G_i\}_{i\in\mathbb{N}}$ and a uniformly recursively generable family $\{H_i\}_{i\in\mathbb{N}}$ of finite non-empty sets such that*

- *range($\mathcal{L}$) $\subseteq$ range($\mathcal{G}$);*
- *for all $i \in \mathbb{N}$, $H_i \subseteq G_i$;*
- *for all $i, j \in \mathbb{N}$ with $H_i \subseteq G_j$ it holds that $G_j \not\subset G_i$.*

Theorem 9 has a natural analogue when the learner is restricted to a co-r.e. hypothesis space, as will be demonstrated later. The next series of results give characterisations of conservative co-r.e. learning under various learning constraints. The first theorem states that for $Ex_{co-r.e.}$-learning, Angluin's tell-tale characterisation has a counterpart with respect to class-preserving hypothesis spaces.

Theorem 10. *Assume that a class $\mathcal{C}$ is uniformly co-r.e. and contains the empty set. Then $\mathcal{C}$ is conservatively $Ex_{co-r.e.}$ learnable using a class-preserving hypothesis space if and only if it has a uniformly co-r.e. indexing $L_0, L_1, \ldots$ and there is a uniformly r.e. family $H_0, H_1, \ldots$ of finite sets such that for all i, j with $H_i \cap L_i \subseteq L_j \subseteq L_i$ it holds that $L_i = L_j$.*

Proof. Let M be a conservative of the given class using a class-preserving hypothesis space, that is, outputting only such indices e such that $\overline{W_e}$ is in the class. Let ϑ be a recursive one-one translation from sequences to natural numbers. Now define

$$L_{\vartheta(\sigma)} = \begin{cases} \overline{W_{M(\sigma)}} & \text{if } W_{M(\sigma)} \cap \text{range}(\sigma) = \emptyset; \\ \emptyset & \text{if } W_{M(\sigma)} \cap \text{range}(\sigma) \neq \emptyset. \end{cases}$$

Furthermore, let $H_{\vartheta(\sigma)} = \text{range}(\sigma)$. As $\emptyset$ is in the given class, all languages in the new numbering are in the class. Now consider any $G \in C$. There is a σ with $\overline{W_{M(\sigma)}} = G$ and range$(\sigma) \subseteq G$. Then $L_{\vartheta(\sigma)} = G$. So G is in the numbering and the given numbering is a co-r.e. class-preserving hypothesis space. If $L_{\vartheta(\sigma)} = \emptyset$ then $\emptyset$ is the only set G with $L_{\vartheta(\sigma)} \cap H_{\vartheta(\sigma)} \subseteq G \subseteq L_{\vartheta(\sigma)}$. If $L_{\vartheta(\sigma)} \neq \emptyset$ then M conjectures $L_{\vartheta(\sigma)}$ on σ and, by the conservativeness of M, M makes only a mind change when seeing some τ with data elements outside $L_{\vartheta(\sigma)}$. It follows that M does not learn any set G with range$(\sigma) \subseteq G \subset L_{\vartheta(\sigma)}$; thus such a set G is not in the class. Hence the numbering satisfies the tell-tale condition given in the theorem.

For the other way round, assume that a co-r.e. indexing $L_0, L_1, \ldots$ of the class and the corresponding finite tell-tale sets $H_0, H_1, \ldots$ of the class are given.

Now one can make a class-preserving learner which starts with a hypothesis for the empty set and which revises a hypothesis if and only if the previous one has enumerated some data observed into its complement; the new hypothesis formed on current input σ is then obtained by searching for the least e such that $H_{e,|\sigma|} \cap L_{e,|\sigma|} \subseteq \text{range}(\sigma) \subseteq L_{e,|\sigma|}$ where $L_{e,|\sigma|}$ is a cofinite set for which finitely many elements have been enumerated into the complement so far. If such an e is not found in time $|\sigma|$ then the learner conjectures $\emptyset$ until a better hypothesis is available. If e is found then the hypothesis issued is a co-r.e. set G such that

$$G = \begin{cases} L_e \text{ if } H_{e,|\sigma|} = H_e; \\ \emptyset \quad \text{if } H_{e,|\sigma|} \subset H_e, \end{cases}$$

This is to reflect that the hypothesis is chosen under the assumption that all elements of H_e are seen; if new elements follow then the hypothesis can be withdrawn and this is made possible by making G to become the empty set (which is then inconsistent with the data observed). Note that the hypotheses issued by the learner are all from the class C. ▮

One has indeed to take a new numbering for Angluin's tell-tale condition to be implemented. The following example is class-preservingly conservatively learnable but does not satisfy the tell-tale condition.

Example 11. Consider the following indexing:

$$\begin{aligned} L_0 &= \emptyset, \\ L_{2n-1} &= \{0, 1, 2, \ldots, n\} \oplus \{n\} \oplus \emptyset, \\ L_{2n} &= \{0, 1, 2, \ldots, n\} \oplus \{n\} \oplus \{x : x > |W_n|\}. \end{aligned}$$

This family is conservatively learnable, but no finite family of $H_0, H_1, \ldots$ as in Theorem 10 exists; hence, the hypothesis space has to be adequately chosen such as to achieve the existence of the tell-tale family.

Proof. A conservative learner of this family first conjectures L_0; if all elements of L_{2n-1} have been observed then it conjectures L_{2n-1}; if a further element has been observed then it conjectures L_{2n}.

But if the family $\{H_0, H_1, \ldots\}$ would exist for this given numbering then the maximum of H_{2n} would be larger than $|W_n|$ whenever W_n is finite. This would then permit to check with oracle $\mathbb{K}$ whether W_n is finite or infinite. But such a check does not exist. ▮

It may be reasonable to inquire next whether all $Cnsv_{co-r.e.}$ learnable, but not necessarily class-preservingly learnable classes have a similar tell-tale characterisation. This is the main content of the following result, which pieces together some presently known equivalent characterisations of $Cnsv_{co-r.e.}$ learnable classes. It shows, moreover, that a conservative co-r.e. learner must also necessarily be prudent. $\{L_0, L_1, L_2, \ldots\}$ *covers* $\mathcal{C}$ if and only if $\mathcal{C} \subseteq \{L_0, L_1, L_2, \ldots\}$.

Theorem 12. *Let $\mathcal{C}$ be a class of co-r.e. sets. The following statements are equivalent:*

(i) $\mathcal{C}$ is $Cnsv_{co-r.e.}$ learnable;
(ii) $\mathcal{C}$ is $PrudCnsv_{co-r.e.}$ learnable;
(iii) There is a uniformly co-r.e. family $L_0, L_1, L_2, \ldots$ covering $\mathcal{C}$ and a uniformly recursively generable family $H_0, H_1, H_2, \ldots$ of the finite sets such that for all n, m it holds that $H_n \subseteq L_n$ and $H_n \subseteq L_m \subseteq L_n \Rightarrow L_n = L_m$;
(iv) There is a uniformly co-r.e. family $L_0, L_1, L_2, \ldots$ covering $\mathcal{C}$ and a uniformly r.e. family of finite sets $H_0, H_1, H_2, \ldots$ such that for all n, m it holds that $H_n \subseteq L_n$ and $H_n \subseteq L_m \subseteq L_n \Rightarrow L_n = L_m$.

Proof. The implications (iii)⇒(iv) and (ii)⇒(i) are immediate, hence only the implications (i)⇒(iii) and (iv)⇒(ii) have to be shown. Among these two, the proof of the implication (i)⇒(iii) shall first be demonstrated.

Suppose that M is a $Cnsv_{co-r.e.}$ learner of $\mathcal{C}$ and that M has as initial hypothesis on the empty sequence a hypothesis for the emptyset. Let $\tau_0, \tau_1, \tau_2, \ldots$ be an enumeration of all sequences in $(\mathbb{N} \cup \{\#\})^*$ which satisfy the following conditions:

– A number x occurs in τ_n if and only if it occurs in τ_n exactly in position x and either τ_n is empty or τ_n ends with a number;
– For all $x < |\tau_n|$ such that $x \notin \text{range}(\tau_n)$, $x \in W_{M(\tau_n)}$;
– For all $\sigma \prec \tau_n$ with either σ being empty or σ ending with a number, $M(\sigma) \neq M(\tau_n)$.

Now define a uniformly co-r.e. family of sets $\{L_n\}_{n\in\mathbb{N}}$ by

$$L_n = \begin{cases} \overline{W}_{M(\tau_n)} & \text{if } W_{M(\tau_n)} \cap \text{range}(\tau_n) = \emptyset; \\ \text{range}(\tau_n) - \{\#\} & \text{if } W_{M(\tau_n)} \cap \text{range}(\tau_n) \neq \emptyset. \end{cases}$$

Furthermore, put $H_n = \text{range}(\tau_n) - \{\#\}$ for all n; by this definition, $\{H_n\}_{n\in\mathbb{N}}$ is a uniformly recursive family of finite sets. For each $L \in \mathcal{C}$, let T_L be the text for L such that for all n, $T_L(n) = n$ if $n \in L$, and $T_L(n) = \#$ otherwise. Since M learns L when fed with the text T_L, there is a least n for which $\overline{W}_{M(T_L[n])} = L$ and $W_{M(T_L[n])} \cap \text{range}(T_L[n]) = \emptyset$; moreover, by the construction of T_L, $T_L[n] \in \{\tau_0, \tau_1, \tau_2, \ldots\}$. Hence $\mathcal{C} \subseteq \{L_n\}_{n\in\mathbb{N}}$. If $\text{range}(\tau_n) \cap W_{M(\tau_n)} = \emptyset$, then $L_n = \overline{W}_{M(\tau_n)}$; so $H_n = \text{range}(\tau_n) - \{\#\} \subseteq L_n$. If $\text{range}(\tau_n) \cap W_{M(\tau_n)} \neq \emptyset$, then $L_n = \text{range}(\tau_n) - \{\#\}$; therefore $H_n \subseteq L_n$ still holds.

It remains to prove the tell-tale characterisation of $\mathcal{C}$. Suppose that for some $i, j, H_i = \text{range}(\tau_i) - \{\#\} \subseteq L_j \subseteq L_i$. It shall be shown that $\tau_i = \tau_j$, and hence $L_i = L_j$, by eliminating in turn the following cases (a), (b) and (c).

(a) Suppose that $\tau_i \prec \tau_j$. By the third condition in the construction of $\{\tau_n\}_{n\in\mathbb{N}}$, $M(\tau_i) \neq M(\tau_j)$, and so by the conservativeness of M, there is an $x \in \text{range}(\tau_j) \cap W_{M(\tau_i)}$. If $L_i = \overline{W_{M(\tau_i)}}$ then $\text{range}(\tau_j) - \{\#\} \subseteq L_j$ and this implies that $L_j \not\subseteq L_i$, if $L_i = range(\tau_i) - \{\#\}$, then the last number occuring in τ_j is in $L_j - L_i$. This gives the desired contradiction.

(b) Suppose that $\tau_j \prec \tau_i$. Now $H_i \subseteq L_j \subseteq L_i$ can only happen when $L_j \neq H_j$. Furthermore, $M(\tau_i) \neq M(\tau_j)$. So by the conservativeness of M, there is an $x \in \text{range}(\tau_i) \cap W_{M(\tau_j)}$. Hence $\text{range}(\tau_i) - \{\#\} \not\subseteq L_j$, a contradiction.

(c) Suppose that $\tau_i(x) \neq \tau_j(x)$ for some $x < \min\{|\tau_i|, |\tau_j|\}$. Let e be the least such x. If $\tau_i(e) = \#$ and $\tau_j(e) = e$, then by the second condition in the construction of $\{\tau_n\}_{n\in\mathbb{N}}$, $e \in W_{M(\tau_i)}$. Further, if $L_i = \text{range}(\tau_i) - \{\#\}$, then one has $e \in L_j - L_i$, contrary to the condition that $L_j \subseteq L_i$. If $L_i = \overline{W}_{M(\tau_i)}$, then again $e \in L_j - L_i$, giving rise to the same contradiction. On the other hand, if $\tau_i(e) = e$ and $\tau_j(e) = \#$, then $e \in \text{range}(\tau_i) - L_j$, contradicting the condition that $\text{range}(\tau_i) - \{\#\} \subseteq L_j$. This completes the case distinction and establishes the required implication.

Next, the implication (iv)⇒(ii) shall be shown. Let $\{L_e\}_{e\in\mathbb{N}}$ be the uniformly co-r.e. superclass of C and $\{H_e\}_{e\in\mathbb{N}}$ the corresponding uniformly r.e. class of finite sets with the prescribed properties. Denote the approximation to $\overline{L_e}$ at step n by $J_{e,n}$. A conservative and prudent learner M may be defined as follows: upon reading the input σ, M searches for the least index $e \leq n$ so that

$$H_{e,n} \subseteq \text{range}(\sigma) - \{\#\} \subseteq \overline{J_{e,n}}.$$

If no such index is found, let M output co-r.e. index for $\emptyset$; otherwise, set M to conjecture an index c with

$$\overline{W_c} = \begin{cases} \emptyset & \text{if } H_e \neq H_{e,n}; \\ L_e & \text{if } H_e = H_{e,n}. \end{cases}$$

Since M only issues hypotheses for $\emptyset$ and members of $\{L_e\}_{e\in\mathbb{N}}$, it suffices to show that M $Cnsv_{co-r.e.}$ learns the class $\{L_e\}_{e\in\mathbb{N}} \cup \{\emptyset\}$, for it would follow that M is also prudent. If M is fed with an empty text, it must eventually converge to a hypothesis for some L in $\{L_0, L_1, L_2, \ldots\}$ such that $L = \emptyset$, or to a hypothesis for the empty set added explicitly to $\{L_e\}_{e\in\mathbb{N}}$. Suppose that M is fed with a text $T = a_0a_1a_2\ldots$ for some $L \in \{L_0, L_1, L_2, \ldots\}$, and let e be the minimum index such that $L_e = L$. If the learner intermediate conjectures some index based on a set $H_{d,n}$ observed such that $L_d \neq L_e$, then either $L_d \not\supseteq L_e$ and the learner will eventually see a counterexample or $H_{d,n} \neq H_d$ and the conjecture is indeed a co-r.e. index for $\emptyset$. In both cases the learner will revise the hypothesis later. Furthermore, it will not happen that M outputs infinitely often a wrong index since, from some stage n onwards, e is the least index satisfying $H_{e,n} = H_e$ and all elements of H_e having appeared in T before step n and no counter examples to the current conjecture have been observed so far; all $d < e$ fail to have this property from stage n onwards. Hence the learner will then settle for e at the next revision, unless it has already taken some other correct index. The required implication thus follows. ∎

To complete the picture, the following example shows that in general, the uniformly co-r.e. family $\{L_0, L_1, L_2, \ldots\}$ covering $\mathcal{C}$ in statements (iii) and (iv) of Theorem 12 may not be equal to $\mathcal{C}$, and so an exact Angluin tell-tale condition cannot be obtained.

Example 13. Consider the class $\mathcal{C} = \{D : D \text{ is finite and } D \cap \mathbb{K} = \emptyset\}$. The class $\mathcal{C}$ is $PrudCnsv_{co-r.e.}$ learnable, but there is no uniformly co-r.e. family

$L_0, L_1, L_2, \ldots$ and no uniformly r.e. family of finite sets $H_0, H_1, H_2, \ldots$ with $\mathcal{C} = \{L_0, L_1, L_2, \ldots\}$ such that for all n, m, $H_n \subseteq L_n$, and whenever $H_n \subseteq L_m \subseteq L_n$, it holds that $L_m = L_n$.

Proof. Define a learner M by putting $\overline{W}_{M(\sigma)} = \overline{\mathrm{range}(\sigma) - \mathbb{K}}$ for all finite sequences σ. Since each D in $\mathcal{C}$ is finite, M will conservatively learn D when fed with any text for D. Further, as M always conjectures a finite set which does not intersect $\mathbb{K}$, it follows that $\{\overline{W}_{M(\sigma)} : \sigma \in (\mathbb{N} \cup \{\#\})^*\} = \mathcal{C}$, and so M is also prudent. If, however, Angluin's tell-tale condition were satisfied, then the uniformly r.e. tell-tale sets $\{H_e\}_{e \in \mathbb{N}}$ are precisely the members of $\mathcal{C}$, and thus $\bigcup_n H_n = \bigcup \{D : D \text{ is finite and } D \cap \mathbb{K} = \emptyset\} = \overline{\mathbb{K}}$, a contradiction since $\overline{\mathbb{K}}$ is not r.e. ∎

Baliga, Case and Jain [5] proved that an indexed family is behaviourally correctly learnable using some other uniformly r.e. hypothesis space if and only if for every set L_e in the hypothesis space there is a finite set $H_e \subseteq L_e$ such that if $H_e \subseteq L_d \subseteq L_e$ for some index d, then $L_d = L_e$. By contrast, the following example shows that some weak analogue of this result fails; Theorem 15 below implies that also the direct analogue of this result (using uniformly recursive families) fails.

Example 14. Let $L_e = \{\langle x, y\rangle : \varphi_e(x) = y \vee \exists z \leq x[\varphi_e(z)\uparrow]\}$ and let $\mathcal{C}$ be the class of all sets L_e. Then $\mathcal{C}$ is uniformly co-r.e. and satisfies Angluin's tell-tale condition non-effectively, but it is not $BC_{co-r.e.}$ learnable.

Proof. $\mathcal{C}$ is not behaviourally correctly learnable using co-r.e. indices as otherwise the class of all recursive functions would be behaviourally correctly learnable. Now it is shown that $\mathcal{C}$ satisfies Angluin's tell-tale condition non-effectively. One can take

$$H_e = \begin{cases} \emptyset & \text{if } \varphi_e \text{ is total;} \\ \{\langle x, 0\rangle, \langle x, 1\rangle\} & \text{if } x \text{ is the least value where } \varphi_e(x) \text{ is undefined.} \end{cases}$$

Assume now that $H_e \subseteq L_d \subseteq L_e$. If L_e is the graph of a recursive function then $L_d = L_e$ as every set L_d is either the graph of a recursive function or a set containing $\langle x, 0\rangle, \langle x, 1\rangle$ for almost all x.

If $H_e = \{\langle x, 0\rangle, \langle x, 1\rangle\}$ then L_d contains these two elements and x is also the smallest value where $\varphi_d(x)$ is undefined. Now $\varphi_d(x') = \varphi_e(x')$ for $x' < x$ as otherwise L_d would not be a subset of L_e and for each x' there is at least one y with $\langle x', y\rangle \in L_d$. So $L_d = \{\langle x', \varphi_e(x')\rangle : x' < x\} \cup \{\langle x', y\rangle : x' \geq x\} = L_e$ and this completes the proof of the criterion. ∎

4 Indexed families of recursive sets

The present section establishes some results on the learnability of families of recursive sets with either a uniformly r.e. numbering or a uniformly co-r.e. numbering. First, it is shown that when a given class of recursive sets has a uniformly r.e. numbering and is $BC_{co-r.e.}$ learnable, then it is already Ex_{rec} learnable. Moreover, Theorem 24 in the next section will establish the existence of a uniformly r.e. class of recursive sets which is finitely learnable using r.e. indices but not $BC_{co-r.e.}$ learnable.

Theorem 15. *Let $\mathcal{C}$ be a uniformly r.e. class of recursive sets. If $\mathcal{C}$ is $BC_{co-r.e.}$ learnable, then it is Ex_{rec} learnable.*

Proof. Let $A_0, A_1, ...$ be a uniformly recursive enumeration of the class and let M be a $BC_{co-r.e.}$ learner for it. For each number n, let $\sigma_{n,0}, \sigma_{n,1}, \sigma_{n,2}, \ldots$ be an enumeration of all finite sequences in A_n^*. Then one can find for every n in the limit a sequence $\sigma_n = \lim_{t\to\infty} \sigma_{n,t}$ such that $\sigma_n \in A_n^*$ and M outputs on every sequence $\sigma_n\tau$ with $\tau \in A_n^*$ an index $M(\sigma_n\tau)$ of a co-r.e. set such that $W_{M(\sigma_n\tau)} \subseteq \overline{A}_n$. One can find a recursive function f such that $f(n,t)$ is a recursive index for A_n if $\sigma_{n,t} = \sigma_n$, and $f(n,t)$ is a recursive index for a finite set if $\sigma_{n,u} \neq \sigma_{n,t}$ for some $u > t$.

Now, on input with range D and length s, the new learner N searches for the first pair (n,t) such that (1) $\sigma_{n,t} = \sigma_{n,t+s}$, (2) range$(\sigma_{n,t}) \subseteq D$, and (3) the set described by $f(n,t)$ contains D, and N conjectures $f(n,t)$ if this pair is found in time s and $f(s,s)$ otherwise. Note that condition (3) can be checked as $f(n,t)$ produces a recursive index. Furthermore, if the learner converges to $f(n,t)$, then (n,t) satisfies $\sigma_{n,t} = \sigma_n$ as otherwise (n,t) would not have fulfilled requirement (1) of the above list when s is sufficiently large. Hence the data must contain range(σ_n) and furthermore the data must be contained in A_n. ∎

It is an immediate consequence of Theorem 15 that an indexed family which is $BC_{co-r.e.}$ learnable must also be $Ex_{co-r.e.}$ and $Ex_{r.e.}$ learnable; this, in turn, implies by Theorem 8 that a uniformly recursive family of sets is $BC_{co-r.e.}$ learnable if and only if it satisfies Angluin's tell-tale condition. The next theorem sums up these observations, and asserts in addition that Angluin's criterion implies $Cnsv_{co-r.e.}$ learnability. In the context of concept identification, this result seems to affirm empirical findings that human learning by means of positive as opposed to negative instances is indeed superior when the extensions of concepts to be learnt are uniformly recursive.

Theorem 16. *Let $\mathcal{C} = \{L_e\}_{e\in\mathbb{N}}$ be a uniformly recursive family of sets. Then the following conditions are equivalent:*

(i) $\mathcal{C}$ is $Cnsv_{co-r.e.}$ learnable;
(ii) $\mathcal{C}$ is $BC_{co-r.e.}$ learnable;
(iii) $\mathcal{C}$ is $Ex_{r.e.}$ learnable;
(iv) $\mathcal{C}$ satisfies Angluin's tell-tale condition.

Proof. It is an immediate consequence of Theorem 15 that if $\mathcal{C}$ were $BC_{co-r.e.}$ learnable, then it must also be $Ex_{co-r.e.}$ and $Ex_{r.e.}$ learnable. The equivalence of (iii) and (iv) follows from Theorem 8. To show that $Ex_{co-r.e.}$ learnability implies $Cnsv_{co-r.e.}$ learnability, let M be an $Ex_{co-r.e.}$ learner of C. By the necessity of Angluin's criterion, there is a uniformly r.e. family $H_0, H_1, H_2, \ldots$ of finite sets such that for all i, j, $H_i \subseteq L_i$, and if $H_i \subseteq L_j \subseteq L_i$, then $L_i = L_j$. Now define a learner N as follows: on reading the data $\sigma = a_0 a_1 \ldots a_n$, N searches for the least $e \leq n$ so that

$$H_{e,n} \subseteq \mathrm{range}(\sigma) - \{\#\} \subseteq L_e.$$

Since $\{L_e\}_{e \in \mathbb{N}}$ is uniformly recursive, the search can be carried out effectively. If no such e is found, let N conjecture an index for $\emptyset$. If the least such e is found, N conjectures an index i for which

$$\overline{W_i} = \begin{cases} \emptyset & \text{if } \exists x \in H_e - H_{e,n}; \\ L_e & \text{if } H_{e,n} = H_e. \end{cases}$$

Hence if $H_{e,n} \neq H_e$, then $\overline{W_i} = \emptyset$; and if $H_{e,n} = H_e$, then $\overline{W_i} = L_e$. As there are no i, j with $H_i \subseteq \mathrm{range}(\sigma) - \{\#\} \subseteq L_j \subsetneq L_i$ for any input σ, it follows that N is a conservative learner of $\mathcal{C}$. ∎

5 General classes of r.e. sets, co-r.e. sets and separations

This section summarizes results dealing with the two main learning criteria, namely $Ex_{co-r.e.}$ and $BC_{co-r.e.}$ learning. On the one hand, prudent, explanatory and behaviourally correct learnability form a strict hierarchy of learning notions; on the other hand, vacillatory and explanatory learnability using co-r.e. indices turn out to be equivalent. The following theorem states that *vacillatory learning*, in which the learner is permitted to vacillate between finitely many different indices, is still just as powerful as $Ex_{co-r.e.}$ learning.

Definition 17 (Case [7]). M is said to *vacillatorily* ($Vac_{co-r.e.}$) learn $\mathcal{C}$ if it $BC_{co-r.e.}$ learns $\mathcal{C}$ and outputs on every text T_L for each L in $\mathcal{C}$ only finitely many different indices $i_0, i_1, \ldots, i_n$.

Theorem 18. *If a class is vacillatorily learnable using co-r.e. indices, then the class is explanatorily learnable using co-r.e. indices.*

Proof. Given a vacillatory learner M, consider a new learner N such that N on input σ conjectures $\overline{W}_{u(E)}$ where $u(E)$ is the index of $\cup_{e \in E} W_e$ and $E = \{M(\tau) : \tau \preceq \sigma \wedge \mathrm{range}(\sigma) \cap W_{M(\tau)} = \emptyset\}$.

Now let $L \in \mathcal{C}$ and T be a text for L. Let $c_0, c_1, \ldots, c_k$ be the hypotheses output by M on T. For each $i \in \{0, 1, \ldots, k\}$ let n_i be so large that M has output c_i within processing the first n_i data-items of the text and that, whenever $x_i = \min(L \cap W_{c_i})$ exists, x_i occurs within the first n_i data-items of T and $x_i \in W_{c_i, n_i} \cap L$.

On any $\sigma \preceq T$ of length $n > \max\{n_0, n_1, \ldots, n_k\}$, the set E in the definition of $N(\sigma)$ contains all c_i with $W_{c_i} \cap L = \emptyset$ and no c_i with $W_{c_i} \cap L \neq \emptyset$. Hence E is a fixed set which is the same for all $n > \max\{n_1, n_2, \ldots, n_k\}$ and N converges on T syntactically to an index of the set $\overline{\cup_{e \in E} W_e}$. Furthermore, all W_e with $e \in E$ are subsets of the complement of L and at least one W_e with $e \in E$ is the complement of L; hence N converges to a co-r.e. index of L. This establishes that N is an $Ex_{co-r.e.}$ learner for $\mathcal{C}$. ▮

However, for uniformly co-r.e. families of sets, $PrudBC_{co-r.e.}$ learnability is less restrictive than $Ex_{co-r.e.}$ learnability. This result is parallel to the usual case of learning using r.e. indices.

Theorem 19. *There is a uniformly co-r.e. class $\mathcal{C}$ which is $PrudBC_{co-r.e.}$ learnable but not $Ex_{co-r.e.}$ learnable.*

Proof. In the following proof, the set of all finite binary strings $\{0,1\}^*$ shall be taken as the universal class in any applicable context. Following this convention, the condition $\sigma \notin T$ is equivalent to $\sigma \in \{0,1\}^* - T$. Let $M_0, M_1, M_2, \dots$ be a numbering of all partial-recursive learners. A co-r.e. tree T is constructed by first defining a two-place partial-recursive function $g(n,x)$. The purpose of $g(n,x)$ is to select one or two infinite branches after the sequence $1^n \circ 0$ in such a way that the learner M_n is forced to either make infinitely many mind changes along one branch, or to converge in the limit to the same hypothesis along two distinct branches. Further, an auxiliary r.e. binary string A_n is constructed in tandem with $g(n,x)$ to track the position of the least number x for which $g(n,x) \uparrow$ at each stage. The partial-recursive function $g(n,x)$ is built in stages, and proceeds by the following algorithm. First, set $A_{n,0} = \emptyset$. At stage $s+1$,

1°: If $M_{n,s}(\overbrace{1^n \circ 0 \circ A_{n,s}}) \downarrow \neq M_{n,s}(\overbrace{1^n \circ 0 \circ A_{n,s} \circ 0 \circ 0^m}) \downarrow$ for some $m \leq s$, put $g(n, |A_{n,s}|) = 0$ and $A_{n,s+1} = g(n,0)g(n,1)\dots g(n,s)$. Update $s+1 = s+2$. Otherwise, proceed to 2°.
2°: If $M_{n,s}(\overbrace{1^n \circ 0 \circ A_{n,s}}) \downarrow \neq M_{n,s}(\overbrace{1^n \circ 0 \circ A_{n,s} \circ 1 \circ 0^m}) \downarrow$ for some $m \leq s$, put $g(n, |A_{n,s}|) = 1$ and $A_{n,s+1} = g(n,0)g(n,1)\dots g(n,s)$. Update $s+1 = s+2$. Otherwise, proceed to 3°.
3°: Put $g(n, s+1) = 0$ and $A_{n,s+1} = A_{n,s}$. Update $s+1 = s+2$.

It may be verified that the above algorithm well-defines the partial-recursive function g and the auxiliary binary string A_n. Let f be the recursive function defined by $\varphi_{f(n)}(x) = g(n,x)$. The desired co-r.e. tree T is given by

$$\begin{aligned} T = \{\sigma \in \{0,1\}^* : \exists n \geq 1 \exists \tau \in \{0,1\}^* \forall x \in \mathrm{dom}(\tau) \cap \mathrm{dom}(\varphi_{f(n)}) \\ [\tau(x) = \varphi_{f(n)}(x) \wedge \sigma \preceq 1^n \circ 0 \circ \tau]\} \cup \{\emptyset\}. \end{aligned}$$

Set the class $L = \{\{\tau \in T : \tau \preceq \sigma \text{ or } \sigma \preceq \tau\} : \sigma \in \{0,1\}^* \wedge |\sigma| > 0\} \cup \{\emptyset\}$. It shall be shown that L is $PrudBC_{co-r.e.}$ learnable, but not $Ex_{co-r.e.}$ learnable.

A regular procedure to produce a $BC_{co-r.e.}$ learner M is as follows. On receiving the input data $\overbrace{\sigma} = \sigma_0\sigma_1\dots\sigma_n$, M distinguishes two cases:

- There is no x such that $\sigma_i(x) \downarrow \neq \sigma_j(x) \downarrow$ for some $i, j \leq n$. M then identifies the longest string of $\overbrace{\sigma}$, say τ, and conjectures the index e with

$$\overline{W}_e = \{\sigma \in T : \sigma \preceq \tau \vee \tau \preceq \sigma\}.$$

– There is an x such that $\sigma_i(x) \neq \sigma_j(x)$ for some $i, j \leq n$. M then chooses the shortest string τ so that $\tau \circ 0 \preceq \sigma_i$ and $\tau \circ 1 \preceq \sigma_j$ for some $i, j \leq n$, and again conjectures the index e with

$$\overline{W}_e = \{\sigma \in T : \sigma \preceq \tau \vee \tau \preceq \sigma\}.$$

Suppose that M is fed with a text for $L_\sigma = \{\tau \in T : \tau \preceq \sigma$ or $\sigma \preceq \tau\}$. If $\sigma \notin T$, then there is no extension τ of σ so that $\tau \in T$, and thus every string in the text must be a proper prefix of σ. Hence the first case always holds, and M must eventually converge in the limit to an index for $\{0,1\}^* - L_\sigma$. If, however, $\sigma \in T$, then all prefixes of σ are also in T, and by the construction of T, there are at most two branches extending σ. If a single branch extends σ, then the first case always holds; if two branches extend σ, then at some point the second case will hold. In both instances, M always converges in the limit to an index for $\{0,1\}^* - L_\sigma$. Together with the fact that M only outputs co-r.e. indices for sets in L, this establishes that M is a $PrudBC_{co-r.e.}$ learner, as was required to be shown. On the other hand, as the construction of T is directly diagonalised against every possible explanatory learner of L, L is not $Ex_{co-r.e.}$ learnable, as claimed. ∎

While Fulk [13] showed that every $Ex_{r.e.}$ learnable class is prudently $Ex_{r.e.}$ learnable and Jain, Stephan and Ye [15] showed that every $BC_{r.e.}$ learnable class is prudently $BC_{r.e.}$ learnable, the next result shows that prudence does not need to hold for co-r.e. learning, even if every member of the class is recursive.

Theorem 20. *There is a class $\mathcal{C}$ of recursive sets which is $Ex_{co-r.e.}$ learnable but not $PrudBC_{co-r.e.}$ learnable.*

Sketch. For each partial-recursive $\{0,1\}$-valued function f, let $L_f = \{\langle x, y\rangle : y \in \{0,1\} \wedge f(x)\downarrow \Rightarrow y = f(x)\}$. Next, define the class

$$\mathcal{C} = \{L_f : \exists e(\forall n > e)[f(e)\uparrow \wedge \varphi_e(n)\downarrow = f(n)\downarrow \in \{0,1\}]\}.$$

An $Ex_{co-r.e.}$ learner M of $\mathcal{C}$ is built as follows. Suppose that M has observed the data $\sigma = a_0a_1a_2\ldots a_n$. M then identifies the maximum number e such that $\{\langle e, 0\rangle, \langle e, 1\rangle\} \subseteq \mathrm{range}(\sigma)$. If no such e exists, M conjectures an index for $\emptyset$; otherwise, it conjectures the index i with

$$\overline{W}_i = \{\langle x, y\rangle : y \geq 2 \vee (x \leq e \wedge y \leq 1 \wedge \langle x, y\rangle \in \mathrm{range}(\sigma)) \vee (x > e \wedge \varphi_e(x)\downarrow \leq 1 \Rightarrow y = 1 - \varphi_e(x))\}.$$

Following this algorithm, when M is fed with a text for $L_f \in \mathcal{C}$, it eventually identifies the maximum e for which f is undefined. Since, for all $x > e$, $\varphi_e(x)\downarrow = f(x) \leq 1$ and $\langle x, f(x)\rangle \in L_f$, it follows that M will converge to an index i for $\overline{L}_f$ in the limit.

Next, assume by way of contradiction that there were a prudent $BC_{co-r.e.}$ learner M of $\mathcal{C}$. It shall be shown that there is a sequence $\sigma \in ((\mathbb{N}\times\{0,1\})\cup\{\#\})^*$ such that $\mathrm{range}(\sigma) - \overline{W}_{M(\sigma)} = \emptyset$, and $\{x : y \in \{0,1\} \wedge \langle x, y\rangle \in W_{M(\sigma)}\}$ is finite. To this end, define the class

$$\Psi_\sigma = \{\langle x, y\rangle : y \in \{0,1\} \wedge (\text{range}(\sigma) - \overline{W}_{M(\sigma)} \neq \emptyset \vee \langle x, 1-y\rangle \notin \overline{W}_{M(\sigma)})\}.$$

Claim: There is a string $\sigma \in ((\mathbb{N}\times\{0,1\})\cup\{\#\})^*$ such that Ψ_σ has finite domain.

Suppose towards a contradiction that for all $\sigma \in ((\mathbb{N}\times\{0,1\})\cup\{\#\})^*$, Ψ_σ has an infinite domain. Now define a recursive $\{0,1\}$-valued function g as follows: firstly, let

$$\sigma_0, \sigma_1, \sigma_2, \ldots$$

be an enumeration of all finite sequences in $(\mathbb{N} \times \{0,1\}) \cup \{\#\}$ such that each sequence occurs infinitely often. Secondly, generate an infinite list of pairs

$$\langle u_0, v_0\rangle, \langle u_1, v_1\rangle, \langle u_2, v_2\rangle, \ldots$$

such that $\langle u_0, v_0\rangle$ is the first pair found in Ψ_{σ_0} with $u_0 \geq 0$, and for $i \geq 1$, $\langle u_i, v_i\rangle$ is the first pair found in Ψ_{σ_i} with $u_i > \max\{i-1, u_{i-1}\}$. By the initial assumption, Ψ_σ has an infinite domain for all sequences σ, implying that the above list may be effectively generated. Lastly, let

$$g(x) = \begin{cases} 0 & \text{if } x \notin \{u_0, u_1, u_2, \ldots\}; \\ 1 - v_i & \text{if } x = u_i. \end{cases}$$

Since $u_0 < u_1 < u_2 < \ldots$, it can be decided recursively whether or not $x \in \{u_0, u_1, u_2, \ldots\}$, by running a search through $u_0, u_1, u_2, \ldots$ in increasing order until the least $u_i \geq x$ is found. Hence g is total.

Let e be an index for the recursive function g defined above. There is a partial-recursive function f such that

$$f(x) = \begin{cases} g(x) & \text{if } x \neq e; \\ \uparrow & \text{if } x = e. \end{cases}$$

It follows that $L_f \in \mathcal{C}$. However, L_f differs from all the functions in the enumeration. Hence the assumption that all Ψ_σ in the enumeration have an infinite domain is wrong. This proves the claim.

Thus by the claim, one may find a $\sigma \in ((\mathbb{N}\times\{0,1\})\cup\{\#\})^*$ so that the domain of $W_{M(\sigma)} \cap (\mathbb{N} \times \{0,1\})$ is finite. Fix any such σ. Since M is a prudent learner, it must $BC_{co-r.e.}$ learn $L = \overline{W}_{M(\sigma)}$. In particular, there is a locking sequence $\tau \in (L\cup\{\#\})^*$ for L; in other words, for all $\eta \in (L\cup\{\#\})^*$, $\overline{W}_{M(\tau\eta)} = \overline{W}_{M(\tau)} = L$. Further, by the property that the domain of $W_{M(\sigma)} \cap (\mathbb{N} \times \{0,1\})$ is finite, there is a d such that for all $x \geq d$, $\{\langle x, 0\rangle, \langle x, 1\rangle\} \subset L$. Let h be any recursive $\{0,1\}$-valued function. By the Padding Lemma, one may choose an index e of h such that $e \geq d$. Define a partial-recursive function θ by

$$\theta(x) = \begin{cases} y & \text{if } x < e \wedge y \in \{0,1\} \wedge \langle x, y\rangle \in \text{range}(\tau); \\ \uparrow & \text{if } x = e, \text{ or if } x < e \text{ and } \forall y \in \{0,1\}[\langle x, y\rangle \notin \text{range}(\tau)]; \\ h(x) & \text{if } x > e. \end{cases}$$

By the above definition of θ, $L_\theta \in \mathcal{C}$. As τ is a locking sequence for L, and $\text{range}(\tau) \subset L_\theta \subset L$, $\overline{W}_{M(\tau\eta)} = \overline{W}_{M(\tau)} = L \neq L_\theta$ for all $\eta \in (L_\theta \cup \{\#\})^*$, implying that M cannot $BC_{co-r.e}$ learn L_θ, a contradiction. This establishes that $\mathcal{C}$ is not prudently learnable. ∎

A consequence of Theorem 12 is that every $Cnsv_{co-r.e.}$ learnable class is also $PrudEx_{co-r.e.}$ learnable; the next result shows that $Cnsv_{co-r.e.}$ learnability is in fact more restrictive than $PrudEx_{co-r.e.}$ learnability.

Theorem 21. *There exists a uniformly co-r.e. class $\mathcal{C}$ which is $PrudEx_{co-r.e.}$ learnable but not $Cnsv_{co-r.e.}$ learnable.*

Proof. Let $M_0, M_1, M_2, \ldots$ be a numbering of all partial-recursive learners, and $\sigma_0, \sigma_1, \sigma_2, \ldots$ be an enumeration of all elements of $\mathbb{N}^*$. Build a uniformly co-r.e. family $\{L_i\}_{i\in\mathbb{N}}$ as follows. With $\eta_0 = e$, set $\overline{L}_{2e,0} = \{\tau \in \mathbb{N}^* : |\tau| > 0 \wedge \tau(0) \neq \eta_0(0)\}$. At stage $s+1$, using η_s, search for the first $\sigma \in \mathbb{N}^*$ such that $\eta_s \prec \sigma$, and either $M_{e,s}(Pref(\eta_s)) \downarrow\neq M_{e,s}(Pref(\sigma))\downarrow$, or $\sigma \in W_{M_{e,s}(\eta_s),j}$ holds for some $j < s+1$. Let the computational time of this search be bounded by $s+1$. If such a σ is found, let $\eta_{s+1} = \sigma$ and set $\overline{L}_{2e,s+1} = \{\tau \in \mathbb{N}^* : \exists k[k < \max\{|\tau|, |\eta_{s+1}|\} \wedge \tau(k) \neq \eta_{s+1}(k)]\}$. Otherwise, if no such σ is found, set $\eta_{s+1} = \eta_s$ and $\overline{L}_{2e,s+1} = \overline{L}_{2e,s}$. Let $\overline{L}_{2e} = \bigcup_{s\in\mathbb{N}} \overline{L}_{2e,s}$. Since, at each stage $s+1$, $\overline{L}_{2e,s} \subseteq \overline{L}_{2e,s+1}$, the set L_{2e} is co-r.e., and consists of all strings which are comparable with the longest common prefix of the strings in $L_{2e,s}$ at some stage s. Further, set $L_{2e+1} = \{\eta \in \mathbb{N}^* : \eta \preceq \sigma_e \vee \sigma_e \preceq \eta\}$ for all e.

Suppose that, in the construction of L_{2e}, there are infinitely many stages $s+1$ at which a string $\sigma \in \mathbb{N}^*$ is found such that if η_s is the longest common prefix of the strings in $L_{2e,s}$, then $\eta_s \prec \sigma$, and either $M_{e,s}(\overbrace{\eta_s}) \downarrow\neq M_{e,s}(\overbrace{\sigma})\downarrow$, or $\sigma \in W_{M_{e,s}(\eta_s),j}$ holds for some $j < s+1$. Consequently, L_{2e} is an infinite chain

$$e \prec \tau_0 \prec \tau_1 \prec \tau_2 \prec \ldots$$

of comparable strings. If, on the other hand, there is a final stage $s+1$ at which such a string τ is found, then $L_{2e} = \{\eta \in \mathbb{N}^* : \eta \preceq \tau \vee \tau \preceq \eta\}$.

On reading the input $\tau = \tau_0\tau_1\ldots\tau_n$, a prudent explanatory learner N distinguishes two cases:

- For all $i, j \leq n$, τ_i and τ_j are comparable by string extension. N identifies the first number e in all the τ_i's, and issues a hypothesis for the set L_{2e}.
- There are $i, j \leq n$ with $\tau_i \not\preceq \tau_j$ and $\tau_j \not\preceq \tau_i$. N identifies the longest common prefix of the τ_i's, say σ, and outputs an index c such that

$$\overline{W}_c = \{\eta \in \mathbb{N}^* : \eta \preceq \sigma \vee \sigma \preceq \eta\}.$$

First, it shall be verified that N $PrudEx_{co-r.e.}$ learns $\{L_0, L_1, L_2, \ldots\}$. Given a text $T = \tau_0\tau_1\tau_2\ldots$ for the set $\{\eta \in \mathbb{N}^* : \eta \preceq \sigma \vee \sigma \preceq \eta\}$, where $\sigma \in \mathbb{N}^*$, there is an n_0 so large that whenever $n \geq n_0$, σ is the longest common prefix of $\tau_0, \tau_1, \ldots, \tau_n$, and there are strings τ_i, τ_j with $i, j \leq n$ which are incomparable by string extension. Thus the second of the above two cases will always hold whenever $n \geq n_0$, and N converges to a correct hypothesis for the set $\{\eta \in \mathbb{N}^* : \eta \preceq \sigma \vee \sigma \preceq \eta\}$ in the limit. If N is fed with a text for L_{2e}, where L_{2e} is an infinite chain $e \prec \tau_0 \prec \tau_1 \prec \tau_2 \prec \ldots$ of comparable strings, then the first case will always hold, and therefore N will still $Ex_{co-r.e.}$ learn L_{2e}. Moreover, since N only issues hypotheses for sets in $\{L_0, L_1, L_2, \ldots\}$, it is also prudent.

It is shown that no M_e is a conservative learner of $\{L_0, L_1, L_2, \ldots\}$. Suppose that L_{2e} is an infinite chain $e \prec \tau_0 \prec \tau_1 \prec \tau_2 \prec \ldots$ of comparable strings; it follows that M_e fails to learn L_{2e}, as there are infinitely many stages at which M_e will either make a mind change, or conjecture a set whose complement enumerates elements of L_{2e}, when fed with the text T such that $T[i] = e \circ \tau_0 \circ \tau_1 \circ \tau_2 \ldots \circ \tau_{i-1}$ for all $i \geq 1$. By contrast, if $L_{2e} = \{\eta \in \mathbb{N}^* : \eta \preceq \sigma \vee \sigma \preceq \eta\}$ for some $\sigma \in \mathbb{N}^*$, then $L_{2e} \subseteq \overline{W}_{M_e(Pref(\sigma))}$. Further, if L_{2e} is not a proper subset of $\overline{W}_{M_e(Pref(\sigma))}$, then the set $H = \{\eta \in \mathbb{N}^* : \eta \preceq \sigma \circ 0 \vee \sigma \circ 0 \preceq \eta\}$ is a member of $\{L_0, L_1, L_2, \ldots\}$ that must be a proper subset of $\overline{W}_{M_e(Pref(\sigma))}$, and so M_e cannot conservatively learn H when fed with a text that starts with $Pref(\sigma)$. ∎

The final results separate the learnability of classes of recursive sets with respect to whether r.e. or co-r.e. hypothesis spaces are used. These results are a bit parallel to prior results on learning correction grammars [8].

Definition 22 (Bārzdiņš and Freivalds [3]). A learner M is said to *finitely* ($Fin_{co-r.e.}$) learn a class $\mathcal{C}$ if, for each L in $\mathcal{C}$ and any text T_L for L, there are an index i and a number n such that $\overline{W}_i = L$, $M(T_L[k]) = i$ for all $k \geq n$ and $M(T_L[k]) = ?$ for all $k < n$.

Theorem 23. *There exists a uniformly co-r.e. class of recursive sets which is $Fin_{co-r.e.}$ learnable but not $BC_{r.e.}$ learnable.*

Proof. Let $M_0, M_1, M_2, \ldots$ be a numbering of all partial-recursive learners. In brief, the example proceeds by constructing inductively, for each number n, a sequence of recursive functions $\{f_{n,t}\}_{t \in \mathbb{N}}$; as one uniformly enumerates the complements of the graphs of $\{f_{n,t}\}_{n,t}$ at a stage s, the outputs of M_n when fed with data from $f_{n,t}$ are observed in parallel; and if, on receiving the data $\langle 0, f_{n,t}(0)\rangle \circ \langle 1, f_{n,t}(1)\rangle \circ \langle 2, f_{n,t}(2)\rangle \circ \ldots \circ \langle x, f_{n,t}(x)\rangle$ for some x, M_n conjectures an index e for which $\langle x+1, f_{n,t}(x+1)\rangle \in W_e$, then the function $f_{n,t+1}$ is constructed so that $f_{n,t+1}(x+1) \neq f_{n,t}(x+1)$ and $\langle x+1, f_{n,t+1}(x+1)\rangle$ has not yet occurred in the enumeration of $\overline{\text{graph}(f_{n,i})}$ for all $i \leq t$. For each n, the pointwise limit of the recursive functions $f_{n,t}$ as $t \to \infty$ yields another recursive function f_n which is diagonalised against the learner M_n. The desired uniformly co-r.e. class $\mathcal{C}$ is then the set of all graphs of the f_n's.

The details of the construction are as follows. First, define, for each n,

$$f_{n,0}(x) = \begin{cases} n \text{ if } x = 0; \\ 0 \text{ if } x > 0. \end{cases}$$

At stage $s = \langle 0, n\rangle$, the next stage $s = s+1$ is updated. At stage $s = \langle t+1, n\rangle$, suppose that $f_{n,0}, f_{n,1}, \ldots, f_{n,t}$ have all been defined, and that $f_{n,t+1}$ has not been defined yet. Search for an $x > t_m$, where t_m is the least index with $f_{n,t} = f_{n,t_m}$, so that $\langle x+1, f_{n,t}(x+1)\rangle \in W_{M_n(\sigma),d}$ for some d, where $\sigma = \langle 0, f_{n,t}(0)\rangle \circ \langle 1, f_{n,t}(1)\rangle \circ \ldots \circ \langle x, f_{n,t}(x)\rangle$; let the search at this stage be bounded by the conditions $d < s$ and $x < t+1$.

– Suppose that such an x is found at this stage. Set

$$f_{n,t+1}(y) = \begin{cases} f_{n,t}(y) & \text{if } y \leq x; \\ t+1 & \text{if } y > x. \end{cases}$$

– Otherwise, suppose that no such x is found; set $f_{n,t+1}(y) = f_{n,t}(y)$ for all y.

Update $s = s+1$, and run through the algorithm again.

If $f_{n,0}, f_{n,1}, \ldots, f_{n,t}$ have not all been defined yet, or if $f_{n,t+1}$ has been defined at an earlier stage, the next stage $s = s+1$ is updated.

For all $n, y \in \mathbb{N}$, put $f_n(y) = \lim_{t\to\infty} f_{n,t}(y)$, and set $\mathcal{C} = \{\{\langle x, f_n(x)\rangle : x \in \mathbb{N}\} : n \in \mathbb{N}\}$.

It shall be shown that each f_n is recursive, and that $\mathcal{C}$ is uniformly co-r.e. For the former assertion, fix an n, and suppose first that for all t, there is a $u > t$ such that $f_{n,t} \neq f_{n,u}$; in other words, one can effectively find an increasing sequence of numbers $x_1, x_2, x_3, \ldots$ corresponding respectively to recursive functions $f_{n,t_1}, f_{n,t_2}, f_{n,t_3}, \ldots$ as described in the algorithm. That each function f_{n,t_i} is indeed recursive may be observed from the following chain of implications: $f_{n,0}$ is recursive as $f_{n,0}(0) = n$ and $f_{n,0}(y) = 0$ for $y > 0$, so that by induction, since $f_{n,t_i+1}(y) = t_i + 1$ for $y > x_i$ and $f_{n,t_i+1} = f_{n,t_i}$ for $y \leq x_i$, f_{n,t_i+1} is also recursive. Thus to compute $f_n(y)$ for a given y, one may run a search for the first x_i with $y \leq x_i$; by assumption, this search would eventually terminate, whence, $f_n(y) = f_{n,t_i}(y)$. This proves that f_n is recursive. If, however, the search for an x fails whenever $t \geq t_m$, then $f_{n,u} = f_{n,t_m}$ for all $u \geq t_m$, and so $f_n = f_{n,t_m}$; that is, f_n is again recursive.

To prove the next assertion, the following observations are applied: (i) for all n, x, $f_n(x) = \max\{f_{n,t}(x) : t \in \mathbb{N}\}$, and (ii) for all n, x, $f_{n,t+1}(x) \in \{f_{n,t}(x), t+1\}$. Suppose that $f_n(x) = y$. (i) and (ii) imply that for all t such that $t > y$, it must follow that $f_{n,t+1}(x) = f_{n,t}(x)$. That is, for all $t > y$, $f_{n,t}(x) = f_{n,y}(x) = y$. Conversely, if, for all $t > y$, $f_{n,t}(x) = f_{n,y}(x) = y$, then $f_n(x) = y$. Let $A_{n,t} = \{\langle x, y\rangle : t \leq y \vee y = f_{n,t}(x)\}$, and put $A_n = \bigcap_{t\in\mathbb{N}} A_{n,t}$. By the s-m-n theorem, there is a recursive function f with $A_n = \overline{W_{f(n)}}$. Hence $\mathcal{C} = \{A_n\}_{n\in\mathbb{N}}$ is uniformly co-recursively enumerable.

A $Fin_{co-r.e.}$ learner M of $\mathcal{C}$ is constructed as follows: M outputs ? until it is fed with data $\sigma = a_0 a_1 \ldots a_n$ that contains the pair $\langle 0, n\rangle$ for some n. It then conjectures an index e for the set

$$\overline{W_e} = \{\langle x, y\rangle : \forall t > y[y = f_{n,t}(x)]\}.$$

Thus by the construction of each f_n, M $Fin_{co-r.e.}$ learns $\mathcal{C}$.

On the other hand, there is no $BC_{r.e.}$ learner of $\mathcal{C}$, as the graph of f_n is directly diagonalised against M_n if it is a $BC_{r.e.}$ learner, and this gives the required result. ∎

Theorem 24. *There is a uniformly r.e. class of recursive sets which is $Fin_{r.e.}$ learnable but not $BC_{co-r.e.}$ learnable.*

Proof. Let $L_n = \{n\} \oplus \{m : W_{n,m} \subset W_n \vee (m > 0 \wedge W_{n,m} = W_{n,m-1})\}$ and $\mathcal{C} = \{L_0, L_1, \ldots\}$; this family is uniformly r.e., finitely learnable with respect to r.e. indices, but is not $BC_{co-r.e.}$ learnable.

Now the class $\mathcal{C}$ is finitely learnable by conjecturing L_n for the first n such that $2n$ is observed in the input.

Assume now by way of contradiction that $\mathcal{C}$ is $BC_{co-r.e.}$ learnable via M. Then W_n is infinite iff M outputs on every text of L_n almost always sets which do not contain any number from $\{n\} \oplus \mathbb{N}$. Due to the usual locking sequence argument, this is equivalent to

$$|W_n| = \infty \Leftrightarrow \exists \sigma \in L_n^* \forall \tau \in L_n^* \forall s[(\{n\} \oplus \mathbb{N}) \cap W_{M(\sigma\tau),s} = \emptyset].$$

But that is impossible as the set of indices of infinite sets is not Σ_2^0; hence $\mathcal{C}$ is not $BC_{co-r.e.}$ learnable. ∎

Thus, while $BC_{r.e.}$ and $Fin_{co-r.e.}$ are incomparable learning notions, it follows from Theorem 15 that the class of all uniformly r.e. sets of recursive languages which are Ex_{rec} learnable strictly subsumes all uniformly r.e. families of $BC_{co-r.e.}$ learnable recursive languages. At present, however, it is still unknown whether or not there is a uniformly co-r.e. class of recursive languages which is $Ex_{r.e.}$ learnable but not $BC_{co-r.e.}$ learnable.

6 Conclusion

Co-r.e. languages are a natural counterpart to the r.e. ones and the current work studies the learnability of these languages. This framework of learning may be viewed as an approximate model for concept identification based on negative instances. Theorem 12 shows that for co-r.e. languages, Angluin's tell-tale condition characterises conservative learnability and not explanatory learnability. Furthermore, the characterisation works only with some class-comprising hypothesis space and not the given one. Later, Theorem 16 shows that for indexed families of recursive sets, the criteria Ex_{rec}, $Cnsv_{co-r.e.}$ and $BC_{co-r.e.}$ coincide. While for $Ex_{r.e.}$ and $BC_{r.e.}$ learning, prudence can be achieved, there is an $Ex_{co-r.e.}$ learnable class which does not even have a prudent $BC_{co-r.e.}$ learner.

References

1. Dana Angluin. Finding patterns common to a set of strings. *Journal of Computer and System Sciences* 21(1) (1980): 46-62.
2. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control* 45(2) (1980): 117-135.
3. Janis Bārzdiņš and Rūsiņš Freivalds. On the prediction of general recursive functions. *Soviet Mathematics Doklady*, 13 (1972): 1224 - 1228.
   Two theorems on the limiting synthesis of functions. *Theory of Algorithms and Programs*, Volume 1, pages 82–88, Latvian State University, Riga, Latvia, 1974.
4. Lenore Blum and Manuel Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28 (1975): 125–155.

5. Ganesh Baliga, John Case and Sanjay Jain. The synthesis of language learners. *Information and Computation* 152 (1999): 16-43.
6. John Case and Christopher Lynes. Machine inductive inference and language identification. *Proceedings of the 9th International Colloquium on Automata, Languages and Programming*, volume 140 of *Lecture Notes in Computer Science*, pages 107–115. Springer-Verlag, 1982.
7. John Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28 (1999): 1941-1969.
8. Lorenzo Carlucci, John Case and Sanjay Jain. Learning correction grammars. *The Journal of Symbolic Logic* 74 (2009): 489-516.
9. Dick de Jongh and Makoto Kanazawa. Angluin's theorem for indexed families of r.e. sets and applications. COLT 1996: 193-204.
10. John Case, Sanjay Jain and Arun Sharma. On learning limiting programs. *International Journal of Foundations of Computer Science* 3 (1992): 93-115.
11. Vaira Freibergs and Endel Tulving. The effect of practice on utilization of information from positive and negative instances in concept identification. *Canadian Journal of Psychology* 15(2) (1961) 101-106.
12. Rusins Freivalds, Efim Kinber and Rolf Wiehagen. Inductive inference and computable one-one numberings. *Zeitschrift fuer mathematische Logik und Grundlagen der Mathematik* 28 (1982): 463-479.
13. Mark A. Fulk. Prudence and other conditions on formal language learning. *Information and Computation* 85(1) (1990): 1-11.
14. E. Mark Gold. Language identification in the limit. *Information and Control* 10 (1967): 447-474.
15. Sanjay Jain, Frank Stephan and Nan Ye. Prescribed learning of r.e. classes. *Theoretical Computer Science* 410(19) (2009): 1796-1806.
16. Majorie J. Krebs and Eugene A. Lovelace. Disjunctive concept identification: stimulus complexity and positive versus negative instances. *Journal of Verbal Learning and Verbal Behaviour* 15(2) (1970), 653–657.
17. Steffen Lange and Thomas Zeugmann. Language learning in dependence on the space of hypotheses. COLT 1993: 127-136.
18. Steffen Lange and Thomas Zeugmann. A guided tour across the boundaries of learning recursive languages. GOSLER Final Report 1995: 190-258.
19. Steffen Lange, Thomas Zeugmann and Shyam Kapur. Characterizations of monotonic and dual monotonic language learning. *Information and Computation* 120(2) (1995): 155-173.
20. Steffen Lange, Thomas Zeugmann and Shyam Kapur. Monotonic and dual monotonic language learning. *Theoretical Computer Science* 155(2) (1996): 365-410.
21. Steffen Lange and Thomas Zeugmann. Set-driven and rearrangement-independent learning of recursive languages. *Mathematical Systems Theory* 29(6) (1996): 599-634.
22. Steffen Lange, Thomas Zeugmann and Sandra Zilles. Learning indexed families of recursive languages from positive data: a survey. *Theoretical Computer Science* 397(1-3) (2008): 194-232.
23. Daniel Osherson, Micheal Stob and Scott Weinstein. Ideal learning machines. *Cognitive Science*, 6:277–290, 1982.
24. Karl R. Popper. 1972. *Conjectures and refutations: the growth of scientific knowledge.* London: Routledge and Kegan Paul.
25. Hartley Rogers, Jr. 1987. *Theory of recursive functions and effective computability.* Cambridge, Massachusetts: MIT Press.
26. Frank Stephan. *Recursion theory.* Manuscript, 2009.