Combining Models of Approximation with Partial Learning*

Ziyuan Gao[1], Frank Stephan[2] and Sandra Zilles[3]

[1] Department of Computer Science
University of Regina, Regina, SK, Canada S4S 0A2
Email: gao257@cs.uregina.ca
[2] Department of Mathematics and Department of Computer Science
National University of Singapore, Singapore 119076
Email: fstephan@comp.nus.edu.sg
[3] Department of Computer Science
University of Regina, Regina, SK, Canada S4S 0A2
Email: zilles@cs.uregina.ca

Abstract. This article considers the following general question in Gold's framework of inductive inference: If a class $\mathcal{C}$ can be learnt according to criterion A and also to criterion B, can $\mathcal{C}$ then also be learnt by a learner satisfying both criteria A and B at the same time? In particular, this paper will relate several notions of approximation in scientific inference to partial learning and answer one open question affirmatively of Mark Fulk and Sanjay Jain (Approximate inference and scientific method. *Information and Computation* 114(2):179–191, 1994). This question is whether there is a recursive learner that, when learning an arbitrary recursive function, outputs a sequence of hypotheses such that (a) the hypotheses approximate the function to be learnt in the way defined by Fulk and Jain and (b) almost all hypotheses differ from the target only on finitely many places.

1 Introduction

Gold [9] considered a learning scenario where the learner is fed with piecewise increasing amounts of finite data about a given target language L; at every stage where a new input datum is given, the learner makes a conjecture about L. If there is exactly one correct representation of L that the learner always outputs after some finite time (assuming that it never stops receiving data about L), then the learner is said to have "identified L in the limit." In this paper, it is assumed that all target languages are recursively enumerable (r.e.), and that the learner uses Gödel numbers as its hypotheses.

Since its introduction, Gold's learning paradigm has been used as a basis for a variety of theoretical models in subjects such as human language acquisition [15]

* F. Stephan was partially supported by NUS grants R146-000-181-112 and R146-000-184-112; S. Zilles was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

and the theory of scientific inquiry in the philosophy of science [4, 14]. This paper is mainly concerned with the *partial learning* model [16], which retains several features of Gold's original framework – the modelling of learners as recursive functions, the use of texts as the mode of data presentation, and the restriction of target classes to the family of all r.e. sets – while liberalising the learning criterion by only requiring the learner to output exactly one hypothesis of the target set infinitely often while it must output any other hypothesis only finitely often. A major question studied in this paper is the following: if a class $\mathcal{C}$ can be learnt according to criterion A and also to criterion B, can $\mathcal{C}$ then be learnt by a learner satisfying both criteria A and B at the same time? This question has been treated fairly extensively in the inductive inference literature: Jantke and Beick [12], for example, investigated the existence of learning strategies that identify classes of recursive functions according to two different criteria simultaneously. The present paper will first revisit, in the context of partial learning, Fulk and Jain's [5] notion of approximating recursive functions, and then propose three notions of approximating languages – approximate learning, weak approximate learning, and finite approximate learning. The question as to which of these approximate learning criteria are combinable with various partial learning notions will occupy much of this paper. A secondary question studied is whether or not the combinations of some approximate learning notions with other non-convergent learning criteria, such as behaviourally correct learning with finitely many anomalies (BC^* learning), imply any other classical learning models. Two considerations motivate our investigation of these questions: one, philosophical; the other, learning-theoretic. First, if partial learning is to be taken seriously as a model of language acquisition, then it is quite plausible that learners are capable of gradually improving the quality of their hypotheses over time. For instance, if the learner M sees a sentence S in the text at some point, then it is conceivable that after some finite time, M will only conjecture grammars that generate S. This leads one to consider a notion of the learner "approximating" the target language even though it may not output any correct hypothesis on the input text. It will be shown that while partial learners can always be constructed in a way so that for any given finite set D, their hypotheses will almost always agree with the target language on D, the same does not hold if D must be a finite variant of a fixed infinite set. Thus trade-offs between certain approximate learning constraints and partial learning are sometimes unavoidable – an observation that perhaps has a broader implication in the philosophy of language learning.

The second motivation of this work comes from the field of inductive inference itself – the possibility of uncovering connections between the traditional learning models, in which some sort of strict convergence criterion is imposed, and the "inexact" models such as partial learning and learning with finitely many anomalies. By establishing relations between (a) the combinations of approximate learning criteria with various partial learning models and (b) convergent learning models, one could hope to derive alternative characterisations of some well-studied learning models such as learning in the limit and behaviourally correct learning.

Fulk and Jain [5] defined a notion of approximately learning recursive functions; this was not studied in combination with partial learning, but in combination with BC^* learning. They proved that the class of all recursive functions is approximately learnable, and posed the question of whether or not approximate learning of recursive functions is combinable with BC^* learning; in other words, is there a recursive learner that, when learning an arbitrary recursive function, outputs a sequence of hypotheses such that (a) the hypotheses approximate the function to be learnt (in the sense they defined) and (b) almost all hypotheses differ from the target only on finitely many places? This paper will show that such a learner does exist; moreover, the learner constructed in the proof is partial, that is, it outputs exactly one correct hypothesis infinitely often and outputs every other hypothesis only finitely often. We will also study analogous problems for language learning, that is, whether or not various approximate learning criteria are combinable with BC^*, Vac^*, and BC learning. Section 5 introduces three notions of approximating languages and studies the combinability of these learning criteria with some partial learning notions. Table 1 summarises most of this paper's results on the combinability of various learning constraints for language learning. If criteria A and B are combinable, then the corresponding entry in the table is "✓"; if criteria A and B are not always combinable, then the corresponding entry in the table is "✗". If the question of whether or not A and B are combinable is open, then the corresponding entry in the table is "?".

Learning criterion	*Part*	*ConsvPart*	*ConsPart*
FinApprox	Thm 37	Thm 40	Cor 39
WeakApprox	Prop 31	?	Prop 31
Approx	Thm 34	?	Thm 34

Table 1. Combinability of learning criteria

2 Preliminaries

The notation and terminology from recursion theory adopted in this paper follows in general the book of Rogers [17]. Background on inductive inference can be found in [10]. The symbol $\mathbb{N}$ denotes the set of natural numbers, $\{0, 1, 2, \ldots\}$. Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ denote a fixed acceptable numbering [17] of all partial-recursive functions over $\mathbb{N}$. Given a set S, $\overline{S}$ denotes the complement of S and S^* denotes the set of all finite sequences in S. One defines the e-th r.e. set W_e as $dom(\varphi_e)$ and the e-th canonical finite set by choosing D_e such that $\sum_{x \in D_e} 2^x = e$. This paper fixes a one-one padding function pad with $W_{pad(e,d)} = W_e$ for all

e, d. Furthermore, $\langle x, y\rangle$ denotes Cantor's pairing function, given by $\langle x, y\rangle = \frac{1}{2}(x+y)(x+y+1)+y$. A triple $\langle x, y, z\rangle$ denotes $\langle\langle x, y\rangle, z\rangle$. The notation $\eta(x)\downarrow$ means that $\eta(x)$ is defined, and $\eta(x)\uparrow$ means that $\eta(x)$ is undefined. The notation $\varphi_{e,s}(x)\downarrow$ means that $\varphi_e(x)$ is defined within s steps, that is, the computation of $\varphi_e(x)$ halts within s steps.

Turing reducibility is denoted by $\leq_T$; $A \leq_T B$ holds if A can be computed via a machine that uses B as an oracle; that is, it can give information on whether or not x belongs to B. $A \equiv_T B$ means that $A \leq_T B$ and $B \leq_T A$ both hold, and $\{A : A \equiv_T B\}$ is called the Turing degree of B. g^A denotes a function relative to an oracle A. A' denotes the jump of A, that is, $A' = \{e : \varphi_e^A(e)\downarrow\}$. For any partial-recursive function g, $graph(g) = \{\langle x, y\rangle : g(x)\downarrow= y\}$. K denotes the diagonal halting problem $\{e : \varphi_e(e)\downarrow\}$. Furthermore, K_s is a recursive approximation to K; without loss of generality, $K_s \subseteq \{0, 1, \ldots, s\} \cap K$ and the set $\{\langle x, s\rangle : x \in K_s\}$ is primitive recursive.

For any two sets A and B, $A \oplus B = \{2x : x \in A\} \cup \{2y+1 : y \in B\}$. Analogously, $A \oplus B \oplus C = \{3x : x \in A\} \cup \{3y+1 : y \in B\} \cup \{3z+2 : z \in C\}$. $A \subseteq^* B$ means that all but finitely many elements of A are in B.

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 $n = 0$. The concatenation of two strings σ and τ shall be denoted by $\sigma \circ \tau$; for convenience, and whenever there is no possibility of confusion, this is occasionally denoted by $\sigma\tau$. Let $\sigma[n]$ denote the sequence $\sigma(0) \circ \sigma(1) \circ \ldots \circ \sigma(n-1)$. The length of σ is denoted by $|\sigma|$.

The basic learning paradigms investigated in the present paper are *explanatory learning* [9], *behaviourally correct learning* [2, 3], and *partial learning* [16]. These learning models assume that the learner is presented with just positive examples of the target language, and that the learner is fed with a finite amount of data at every stage. Explanatory learning (or "learning in the limit"), first introduced by Gold [9], is the requirement that the learner must output in the limit a single correct representation h of the target language L; if L is an r.e. set, then h is usually an r.e. index of L with respect to the standard numbering $W_0, W_1, W_2, \ldots$ of all r.e. sets. Bārzdiņš [2] and Case [3] considered a more powerful variant of "learning in the limit" known as behaviourally correct learning, whereby the learner must almost always output a correct hypothesis of the input set, but some of the correct hypotheses may be syntactically distinct. Still more general is the criterion of partial learning that Osherson, Stob and Weinstein [16] defined; in this model, the learner must output exactly one correct index of the input set infinitely often and output any other conjecture only finitely often.

In addition, one can specify various constraints on the learner; the next definition imposes a restriction on the hypothesis space of the learner. As was shown by Lange and Zeugmann [13], restricting the hypothesis space of the learner to the target class reduces the power of learning in the limit. Thus, it may be natural to ask if the same constraint on the hypothesis space reduces the strengths of various partial learning criteria.

Definition 1. [18] M is said to *class-preservingly* (*ClsPresv*) learn $\mathcal{C}$ if it learns $\mathcal{C}$ with respect to a hypothesis space $\{H_0, H_1, H_2, \ldots\}$, where $H_0, H_1, H_2, \ldots$ are r.e. sets, such that $\mathcal{C} = \{H_0, H_1, H_2, \ldots\}$.

Throughout this paper, successful learning with respect to a class $\mathcal{C}$ will generally refer to class-comprising learning unless specified otherwise.

Let $\mathcal{C}$ be a class of r.e. sets. Throughout this paper, the mode of data presentation is that of a *text*, by which is meant an infinite sequence of natural numbers and the # symbol. Formally, a *text* T_L for some L in $\mathcal{C}$ is a map $T_L : \mathbb{N} \to \mathbb{N} \cup \{\#\}$ such that $L = range(T_L)$; here, $T_L[n]$ denotes the sequence $T_L(0) \circ T_L(1) \circ \ldots \circ T_L(n-1)$ and the range of a text T, denoted $range(T)$, is the set of numbers occurring in T. Analogously, for a finite sequence σ, $range(\sigma)$ is the set of numbers occurring in σ. A text, in other words, is a presentation of positive data from the target set. In the following definitions, M is a recursive function mapping $(\mathbb{N} \cup \{\#\})^*$ into $\mathbb{N} \cup \{?\}$; the ? symbol permits M to abstain from conjecturing at any stage. M may also be equipped with an oracle. In this case, a learner that has access to oracle A is an A-recursive function mapping $(\mathbb{N} \cup \{\#\})^*$ into $\mathbb{N} \cup \{?\}$.

Definition 2. (i) [16] M *partially* (*Part*) learns $\mathcal{C}$ if, for every L in $\mathcal{C}$ and each text T_L for L, there is exactly one index e such that $M(T_L[k]) = e$ for infinitely many k; furthermore, if M outputs e infinitely often on T_L, then $L = W_e$.

(ii) [9] M *explanatorily* (*Ex*) learns $\mathcal{C}$ if, for every L in $\mathcal{C}$ and each text T_L for L, there is a number n for which $L = W_{M(T_L[n])}$ and, for any $j \geq n$, $M(T_L[j]) = M(T_L[n])$.

(iii) [3] M *behaviourally correctly* (*BC*) learns $\mathcal{C}$ if, for every L in $\mathcal{C}$ and each text T_L for L, there is a number n for which $L = W_{M(T_L[j])}$ whenever $j \geq n$.

(iv) [1] M is *consistent* (*Cons*) if for all $\sigma \in (\mathbb{N} \cup \{\#\})^*$, $range(\sigma) \subseteq W_{M(\sigma)}$.

(v) [1] For any text T, M is *consistent on* T if $range(T[n]) \subseteq W_{M(T[n])}$ for all $n > 0$.

(vi) [7] M is said to *consistently partially* (*ConsPart*) learn $\mathcal{C}$ if it partially learns $\mathcal{C}$ from text and is consistent.

(vii) [6] M is said to *conservatively partially* (*ConsvPart*) learn $\mathcal{C}$ if it partially learns $\mathcal{C}$ and outputs on each text for every L in $\mathcal{C}$ exactly one index e with $L \subseteq W_e$.

(viii) [4] M is said to *behaviourally correctly* learn $\mathcal{C}$ *with at most a anomalies* (BC^a) iff for every $L \in \mathcal{C}$ and each text T_L for L, there is a number n for which $|(W_{M(T_L[j])} - L) \cup (L - W_{M(T_L[j])})| \leq a$ whenever $j \geq n$.

(ix) [4] M is said to *behaviourally correctly* learn $\mathcal{C}$ *with finitely many anomalies* (BC^*) iff for every $L \in \mathcal{C}$ and each text T_L for L, there is a number n for which $|(W_{M(T_L[j])} - L) \cup (L - W_{M(T_L[j])})| < \infty$ whenever $j \geq n$.

3 Partial Learning

This section collects some useful lemmas and known results on partial learning. The following proposition will be occasionally used in this paper.

Proposition 3. *Let M be a recursive learner. There is a recursive learner N such that if $e_0, e_1, e_2, \ldots$ is the (possibly finite) sequence of all distinct conjectures of M on some text T, then for each e_i, the following hold.*

- *If e_i is not the least index that M outputs infinitely often on T, then N outputs indices for W_{e_i} finitely often.*
- *If e_i is the least index that M outputs infinitely often on T, then there is exactly one index d_{e_i} with $W_{e_i} = W_{d_{e_i}}$ that N outputs infinitely often on T.*

Proof. One can simulate M on the given text T. If $e_0, e_1, e_2, \ldots$ is the (possibly finite) sequence of all distinct conjectures of M on T, then for each e_i, N outputs $pad(e_i, m_i)$, where m_i is the number of indices $j < i$ such that $e_j < e_i$. It follows that if M outputs infinitely often a least index e, then the total number of $j < e$ such that $e_j < e$ is some finite number k, so that N outputs $pad(e, k)$ infinitely often and any other index only finitely often. ∎

Theorem 4 (Gao, Jain and Stephan [6]). *A class $\mathcal{C}$ is ConsPart learnable iff $\mathcal{C}$ is a subclass of a uniformly recursive class of sets.*

Proof. Given a recursive learner M, one can define a new learner N as follows. Let $c_0, c_1, c_2, \ldots$ be the sequence of conjectures of M on some text T. For each k, N outputs $pad(c_k, d_k)$, where $d_k = |\{m < k : c_m < c_k\}|$. Thus if c_i is not the least index that M outputs infinitely often on T, then N outputs any index of the form $pad(c', f)$, where $c' = c_i$, only finitely often. On the other hand, if c_i is the least index that M outputs infinitely often on T, then there is some $j \geq i$ such that $c_j = c_i$ and N outputs $pad(c_j, d_j)$ infinitely often. ∎

Theorem 5 (Gao, Stephan and Zilles [8]). *A class $\mathcal{C}$ of r.e. sets is ConsConsvPart learnable iff $\mathcal{C}$ is ConsConsvEx learnable.*

4 Approximate Learning of Functions

Fulk and Jain [5] proposed a mathematically rigorous definition of *approximate inference*, a notion originally motivated by studies in the philosophy of science.

Definition 6. [5] An approximate (*Approx*) learner outputs on the graph of a function f a sequence of hypotheses such that there is a sequence $S_0, S_1, \ldots$ of sets satisfying the following conditions:
(a) The S_n form an ascending sequence of sets such that their union is the set of all natural numbers;
(b) There are infinitely many n such that $S_{n+1} - S_n$ is infinite;
(c) The n-th hypothesis is correct on all $x \in S_n$ but nothing is said about the $x \notin S_n$.

The next proposition simplifies this set of conditions.

Proposition 7. *M Approx learns a recursive function f iff the conditions hold:*
(d) For all x and almost all n, M's n-th hypothesis is correct at x;
(e) There is an infinite set S such that for almost all n and all $x \in S$, M's n-th hypothesis is correct at x.

Proof. If one has (a), (b), (c), then the set S is just the first set S_n which is infinite and the other conditions follow.

If one has (d) and (e) and one distinguishes two cases: If n is so small that the n-th and all subsequent hypotheses are not yet correct on S then one lets $S_n = \emptyset$ else one defines that S_n contains all $x \leq n$ such that each m-th hypothesis with $m \geq n$ is correct on x plus half of those members of S which are not in any S_m with $m < n$. So the trick is just not to put all members of S at one step into some S_n but just to put at each step which is applicable an infinite new amount while still another infinite amount remains outside S_n to be put later. ∎

Fulk and Jain interpreted their notion of approximation as a process in scientific inference whereby physicists take the limit of the average result of a sequence of experiments. Their result that the class of recursive functions is approximately learnable seems to justify this view.

Theorem 8 (Fulk and Jain [5]). *There is a learner M that approximates every recursive function.*

The following theorem answers an open problem of Fulk and Jain [5] on whether the class of recursive functions has a learner which outputs a sequence of hypotheses which (a) approximates the function to be learnt and (b) almost always differs from the target only on finitely many places.

Theorem 9. *There is a learner M which learns the class of all recursive functions such that (a) M is a BC^* learner, (b) M is a partial learner and (c) M is an approximate learner.*

Proof. Let $\psi_0, \psi_1, \ldots$ be an enumeration of all recursive functions and some partial ones such that in every step s there is exactly one pair (e, x) for which $\psi_e(x)$ becomes defined at step s and this pair satisfies in addition that $\psi_e(y)$ is already defined by step s for all $y < x$. Furthermore, one says that a function ψ_e makes progress on σ at step s iff $\psi_e(x)$ becomes defined at step s and $x \in dom(\sigma)$ and $\psi_e(y) = \sigma(y)$ for all $y \leq x$.

Now one defines for every σ a partial-recursive function $\vartheta_{e,\sigma}$ with the following properties:

- $\vartheta_{e,\sigma}(x) = \sigma(x)$ for all $x \in dom(\sigma)$;
- Let $e_t = e$;
- Inductively for all $s \geq t$, if some index $d < e_s$ makes progress on σ at step $s+1$ then let $e_{s+1} = d$ else let $e_{s+1} = e_s$;
- For each value $x \notin dom(\sigma)$, if there is a step $s \geq t + x$ for which $\psi_{e_s,s}(x)$ is defined then $\vartheta_{e,\sigma}(x)$ takes this value for the least such step s else $\vartheta_{e,\sigma}(x)$ remains undefined.

The learner M, now to be constructed, uses these functions as hypothesis space; on input τ, M outputs the index of $\vartheta_{e,\sigma}$ for the unique e and shortest prefix σ of τ such that the following three conditions are satisfied at some time t:

- t is the first time such that $t \geq |\tau|$ and some function makes progress on τ;
- ψ_e is that function which makes progress at τ;
- for every $d < e$, ψ_d did not make progress on τ at any $s \in \{|\sigma|, \ldots, t\}$ and either $\psi_{d,|\sigma|}$ is inconsistent with σ or $\psi_{d,|\sigma|}(x)$ is undefined for at least one $x \in dom(\sigma)$.

For finitely many strings τ there might not be any such function $\vartheta_{e,\sigma}$, as the conditions need that τ is longer than the largest value up to which some function has made progress at time $|\tau|$; this is true only for almost all but not for all τ. For these finitely many exceptions, one defines that M takes some default value, say the everywhere undefined function. Now the three conditions (a), (b) and (c) of M are verified. For this, let ψ_d be the function to be learnt, note that ψ_d is total.

Condition (a): M is an BC^* learner. Let d be the least index of the function ψ_d to be learnt and let u be the last step where some ψ_e with $e < d$ makes progress on ψ_d. Then every $\tau \preceq \psi_d$ with $|\tau| \geq u+1$ satisfies that first $M(\tau)$ conjectures a function $\vartheta_{e,\sigma}$ with $e \geq d$ and $|\sigma| \geq u+1$ and $\sigma \preceq \psi_d$ and second that almost all e_s used in the definition of $\vartheta_{e,\sigma}$ are equal to d; thus the function computed is a finite variant of ψ_d and M is a BC^* learner.

Condition (b): M is a partial learner. Let $t_0, t_1, \ldots$ be the list of all times where ψ_d makes progress on itself with $u < t_0 < t_1 < \ldots$; now note that whenever $\tau \preceq \psi_d$ and $|\tau| = t_k$ for some k then the conjecture $\vartheta_{e,\sigma}$ given by $M(\tau)$ satisfies that $e = d$ and that $|\sigma| = u+1$ in order to meet the requirement with respect to the indices below d. As none of these make progress from step $u+1$ onwards on ψ_d, they also do not make progress on σ after step $|\sigma|$ and $\vartheta_{e,\sigma} = \psi_d$; hence the learner outputs infinitely often an index for ψ_d. Furthermore, all other indices $\vartheta_{e,\sigma}$ are output only finitely often: if $e < d$ then ψ_e makes after step u not progress towards ψ_d which is the function to be learnt; if $e > d$ then the length of σ depends on the so far made progress of ψ_d on itself and if $|\tau| > t_k$ then also $|\sigma| > t_k$, hence no other index is output infinitely often.

Condition (c): M is an approximate learner. Conditions (d) and (e) in Proposition 7 are used. To see that M is an approximate learner, it is shown that for all $\tau \preceq \psi_d$ with $t_k \leq |\tau| < t_{k+1}$, that the hypothesis $\vartheta_{e,\sigma}$ issued by $M(\tau)$ satisfies that it is correct on the set $\{t_0, t_1, \ldots\}$. If $|\tau| = t_k$ then the hypothesis is correct everywhere as shown under condition (b). So assume that $e > d$ for the e in the hypothesis. Then $|\tau| > t_k$ and $|\sigma| > t_k$, hence $\vartheta_{e,\sigma}(x) = \psi_d(x)$ for all $x \leq t_k$. Furthermore, as ψ_d makes progress on σ in step t_{k+1} and as no ψ_c with $c < d$ makes progress on σ beyond step $|\sigma|$, it follows that the e_s defined in the algorithm of $\vartheta_{e,\sigma}$ all satisfy $e_s = d$ for $s \geq t_{k+1}$; hence $\vartheta_{e,\sigma}(x) = \psi_d(x)$ for all $x \geq t_{k+1}$. ∎

5 Approximate Learning of Languages

This section proposes several notions of approximation in language learning. The first two notions, *approximate* learning and *weak approximate* learning, are adaptations of the set of conditions for approximately learning recursive functions given in Proposition 7. Recall that V is a finite variant of W iff there is an x such that for all $y > x$ it holds that $V(y) = W(y)$.

Definition 10. A class S of languages is *approximately (Approx) learnable* iff for every language $L \in S$ there is an infinite set W such that for all texts T and all finite variants V of W and almost all hypothesis H of the learner on T, $H \cap V = L \cap V$.

A class S of languages is *weakly approximately (WeakApprox) learnable* iff for every language $L \in S$ and for every text T for L there is an infinite set W such that for all finite variants V of W and almost all hypothesis H of the learner on T, $H \cap V = L \cap V$.

So the difference between the two definitions is that approximate learning has one W for all texts while weakly approximate learning has for each text T a W depending on T.

Definition 11. Let $\mathcal{C}$ be a class of r.e. sets. A recursive learner M is said to *WPart* learn $\mathcal{C}$ iff for every text T for any $L \in \mathcal{C}$, M on T outputs at least one index for L. A recursive learner M is said to BC^* *WPart* learn $\mathcal{C}$ iff for every text T for any $L \in \mathcal{C}$, M on T BC^* learns L and outputs at least one index for L.

Definition 12. Let $\mathcal{C}$ be a class of r.e. sets. A recursive learner M is said to Vac^* learn $\mathcal{C}$ iff M outputs on any text T for every $L \in \mathcal{C}$ only finitely many indices, and for almost all n, $W_{M(T[n+1])}$ is a finite variant of L.

Example 13. Case and Smith [4] showed that Vac^* and Ex^* learning of recursive functions are equivalent. However, this equivalence does not extend to all classes of r.e. sets. Take, for example, the class $\mathcal{C} = \{\{e\} \oplus \mathbb{N} : e \in \mathbb{N}\} \cup \{\{e\} \oplus \{x : x \leq |W_e|\} : e \in \mathbb{N}\}$. $\mathcal{C}$ is *Vac* learnable: on any input σ whose range is of the form $\{e\} \oplus D$, determine whether $\max(D) > |W_{e,|\sigma|}|$; if so, conjecture $\{e\} \oplus \mathbb{N}$; otherwise, conjecture $\{e\} \oplus \{x : x \leq |W_e|\}$. If $range(\sigma)$ does not contain any even number, conjecture $range(\sigma)$.

On the other hand, $\mathcal{C}$ is not Ex^* learnable. Assume by way of a contradiction that a recursive learner M Ex^* learns $\mathcal{C}$. Using K as an oracle, one can determine for any e whether W_e is finite. By the assumption that M is an Ex^* learner, one can enumerate a text T for $L_e = \{e\} \oplus \{x : x \leq |W_e|\}$ until at least one of the following holds.

1. There is some m such that for all $x > m$, $x \notin W_e$. This immediately implies that W_e is finite.
2. For some $\sigma \in (L_e \cup \{\#\})^*$ such that σ is a prefix of T, it holds that for all $\eta \in (L_e \cup \{\#\})^*$, $M(\sigma\eta) = M(\sigma)$; in other words, σ is a locking sequence for L_e.

Now one can use K again to determine whether or not there exists an $\eta \in (\{e\} \oplus \mathbb{N})^*$ such that $M(\sigma\eta) \neq M(\sigma)$. Suppose that $|W_e|$ is finite. Then $\{e\} \oplus \mathbb{N}$ is not a finite variant of L_e; furthermore, as M must Ex^* learn $\{e\} \oplus \mathbb{N}$, there must exist some $\eta \in (\{e\} \oplus \mathbb{N})^*$ for which $M(\sigma\eta) \neq M(\sigma)$. Suppose, on the other hand, that $|W_e|$ is infinite. Then $L_e = \{e\} \oplus \mathbb{N}$, so that by the locking sequence property of σ, $M(\sigma\eta) = M(\sigma)$ for all $\eta \in (\{e\} \oplus \mathbb{N})^*$. Hence the Ex^* learnability of $\mathcal{C}$ would imply that $\{e : |W_e| < \infty\}$ is Turing reducible to K, which is known to be false. ∎

Proposition 14. *If there is an infinite r.e. set W such that all members of the class contain W then the class is Approx learnable.*

Proof. The learner for this just conjectures $range(\sigma) \cup W$ on any input σ. ∎

Thus approximate learning does, for languages, not imply BC^* learning. Note that for infinite coinfinite r.e. sets W, the class of all r.e. supersets of W is not BC^* learnable. The next result is the mirror image of the previous result by just considering a learner which conjectures the range of the data seen so far; for each set L in the class the infinite set S in item (e) of Proposition 7 is just the complement of L.

Proposition 15. *If a class $\mathcal{C}$ consists only of coinfinite r.e. sets then $\mathcal{C}$ is Approx learnable.*

While the class of all coinfinite r.e. sets can be approximated, this is not true for the class of all cofinite sets.

Theorem 16. *The class of all cofinite sets is BC^* learnable but not Approx learnable.*

Proof. The BC^* learner just conjectures on all inputs the set of natural numbers.

Assume now that the set L to be learnt is approximated with parameter set W. Given an approximate learner M for this class, one can construct inductively a text T such that either the text is for some set $L - \{w\}$ and it conjectures almost always that w is in the set to be learnt or the text is for L while there are infinitely many conjectures which do not contain W as a subset.

The idea is to construct the text T step by step by starting in (a) below and by alternating between (a) and (b) as needed:

(a) Select a $w \in L \cap W$ not contained in the part of the text constructed so far and add to the part of the text the elements of $L - \{w\}$ in ascending order until the learner M on the so far constructed initial segment conjectures a set not containing w;

(b) Append to the so far constructed part of the text all elements of L up to the element w (inclusively) and go back to step (a).

This gives then a text T with the desired properties: if the learner eventually stays in (a) forever, it is wrong on w considered when it the last time goes into (a); if the learner goes to (b) infinitely often, the text T is for L while the learner

M conjectures infinitely often sets which are not supersets of W. Thus there is no approximate learner for the class of all cofinite sets. ∎

Call a learner recursively approximating iff for every set L in the class there is an infinite recursive set S such that on every text for L, almost all hypotheses of the learner are correct on S. Note that the previous result can be modified as follows: if a class contains all finite variants of a simple set then it does not have a recursively approximating learner.

Case and Smith [4] published Harrington's observation that the class of recursive functions is BC^* learnable. This result does not carry over to the class of r.e. sets; for example, Gold's class consisting of the set of natural numbers and all finite sets is not BC^* learnable. In the light of Theorem 9, which established that the class of recursive functions can be BC^* and *Part* learnt simultaneously, it may be interesting to know whether *any* BC^* learnable class of r.e. sets can be both BC^* and *Part* learnt at the same time. The next results deal with the question how various learning criteria can be combined with each other (with a special emphasis on BC^*, partial and approximate learning due to the corresponding result for function learning).

Theorem 17. *If a class $\mathcal{C}$ of r.e. sets is BC learnable, then there is a learner M for $\mathcal{C}$ such that M BC and Part learns $\mathcal{C}$.*

Proof. Let M be a BC learner of $\mathcal{C}$. On any given text T for some $L \in \mathcal{C}$, let $e_n = M(T[n+1])$. Define a new learner N_1 such that for each e_i, N_1 outputs e_i at least m times iff there is a stage $t > m$ such that for all $x < m$, $x \in W_{e_i,t}$ iff $x \in range(T[t{+}1])$. Then N_1 outputs any incorrect index of M only finitely often, while it outputs every correct index of M infinitely often. Using the construction in Proposition 3, with N_1 in place of M, one can define a further learner N that preserves the BC learning property of N_1, outputs exactly one index for the target set infinitely often, and outputs every other index only finitely often. ∎

Theorem 18. *Let $\mathcal{C}$ be a class of r.e. sets. For any n, if $\mathcal{C}$ is BC^n learnable, then there is a recursive learner M such that M BC^n and Part learns $\mathcal{C}$.*

Proof. Fix any n such that $\mathcal{C}$ is BC^n learnable. Given a recursive BC^n learner M of $\mathcal{C}$, one can construct a new learner N_1 as follows. First, let $F_0, F_1, F_2, \ldots$ be a one-one enumeration of all finite sets such that $|F_i| \leq n$ for all i. Fix a text T, and let $e_0, e_1, e_2, \ldots$ be the sequence of M's conjectures on T.

For each set of the form $W_{e_i} \cup F_j$ (respectively $W_{e_i} - F_j$), N_1 outputs a canonical index for $W_{e_i} \cup F_j$ (respectively $W_{e_i} - F_j$) at least m times iff the following two conditions hold.

1. There is a stage $s > j$ for which the number of distinct $x < j$ such that either $x \in W_{e_i,s} \wedge x \notin range(T[s+1])$ or $x \in range(T[s+1]) \wedge x \notin W_{e_i,s}$ does not exceed n.
2. There is a stage $t > m$ such that for all $x < m$, $x \in W_{e_i,t} \cup F_j$ iff $x \in range(T[t+1])$ (respectively $x \in W_{e_i,t} - F_j$ iff $x \in range(T[t+1])$).

At any stage $T[s+1]$ where no set of the form $W_{e_i} \cup F_j$ or $W_{e_i} - F_j$ satisfies the conditions above, or each such set has already been output the required number of times (up to the present stage), N_1 outputs $M(T[s+1])$. Suppose T is a text for some $L \in \mathcal{C}$. Since M is a BC^n learner of $\mathcal{C}$, it holds that for almost all i, there are at most n x's such that $W_{e_i}(x) \neq L(x)$. Furthermore, for all j such that $W_{e_j}(x) \neq L(x)$ for at least $n+1$ distinct x's, there is an l such that for all $l' > l$, neither $W_{e_j} \cup F_{l'}$ nor $W_{e_j} - F_{l'}$ will satisfy Condition 1.; thus, for any set S such that $S(x) \neq L(x)$ for more than n distinct values of x, N_1 will conjecture S only finitely often. On the other hand, if there are at most n distinct x's such that $W_{e_i}(x) \neq L(x)$, then there is some l such that either $L = W_{e_i} \cup F_l$ or $L = W_{e_i} - F_l$; consequently, either $W_{e_i} \cup F_l$ or $W_{e_i} - F_l$ will satisfy Conditions 1. and 2. for infinitely many m. Hence N_1 is a BC^n learner of L and it outputs at least one correct index for L infinitely often on any text for L. Using the construction in Proposition 3, with N_1 in place of M, one can define a further learner N that $BC^n Part$ learns $\mathcal{C}$. ▮

Theorem 19. *Let $\mathcal{C}$ be a class of r.e. sets. Suppose there is a recursive learner M for $\mathcal{C}$ such that M BC^* learns $\mathcal{C}$ and outputs on every text for any $L \in \mathcal{C}$ at least one index infinitely often. Then there is a recursive learner N for $\mathcal{C}$ such that N BC^* and Part learns $\mathcal{C}$.*

Proof. Let M be a recursive BC^* learner of $\mathcal{C}$ such that M outputs on every text for any $L \in \mathcal{C}$ at least one index infinitely often. Define a learner N_1 as follows.

On any given text T for some $L \in \mathcal{C}$, let $e_n = M(T[n+1])$. Let $F_0, F_1, F_2, \ldots$ be a one-one enumeration of all finite sets. On input $T[k+1]$, N_1 outputs a canonical index $d_{e_k,l}$ for $W_{e_k} \cup F_l$ (respectively $g_{e_k,l}$ for $W_{e_k} - F_l$) at least m times iff the following conditions hold:

1. M outputs e_k at least $l+1$ times;
2. there is a stage $s > m$ such that for all $x < m$, $x \in range(T[s+1])$ iff $x \in W_{e_k,s} \cup F_l$ (respectively $x \in range(T[s+1])$ iff $x \in W_{e_k,s} - F_l$).

It will be shown that N_1 has the following two learning properties: first, it BC^* learns $\mathcal{C}$; second, it outputs at least one correct index infinitely often; third, it outputs an incorrect index only finitely often. Consider any e_k.

First, suppose that W_{e_k} is not a finite variant of L. Then M outputs e_k only finitely often. Further, N_1 will consider sets of the form $W_{e_k} \cup F_l$ or $W_{e_k} - F_l$ for only finitely many F_l. Since, for each such $W_{e_k} \cup F_l$ (or $W_{e_k} - F_l$), item 2. will be satisfied for only finitely many m, it follows that N_1 will conjecture a set of the form $W_{e_k} \cup F_l$ or $W_{e_k} - F_l$ only finitely often.

Second, suppose that W_{e_k} is a finite variant of L. Then for any F_l, $W_{e_k} \cup F_l$ and $W_{e_k} - F_l$ are both finite variants of L. Hence N_1 preserves its BC^* learning property by outputting any indices for $W_{e_k} \cup F_l$ or $W_{e_k} - F_l$. Moreover, M outputs infinitely often at least one index e_h such that W_{e_h} is a finite variant of L. If $L = W_{e_h} \cup F_c$ (respectively $L = W_{e_h} - F_c$) for some F_c, then N will consider $W_{e_h} \cup F_c$ (respectively $W_{e_h} - F_c$) after M has output e_h at least $c+1$

times. As $W_{e_h} \cup F_c$ (respectively $W_{e_h} - F_c$) satisfies item 2. for almost all m, N_1 will output at least one index for L infinitely often.

Third, suppose that for some F_l, neither $W_{e_k} \cup F_l$ nor $W_{e_k} - F_l$ is equal to L. Then $W_{e_k} \cup F_l$ and $W_{e_k} - F_l$ will satisfy Condition 2. for all but finitely many m, and so N_1 will output a canonical index for $W_{e_k} \cup F_l$ or $W_{e_k} - F_l$ only finitely often. This establishes the three learning properties of N_1.

Now define a further learner N based on the construction in Proposition 3, with N_1 in place of M. N preserves the BC^* learning property of N_1; further, if e'_h is the minimum index that N_1 outputs infinitely often on T, then there is a h' with $e'_h = e_{h'}$ such that N will output $pad(e'_{h'}, d_{h'})$ infinitely often, and every other index is output only finitely often. Therefore N is both a BC^* and a $Part$ learner of $\mathcal{C}$. ∎

Corollary 20. *If a class $\mathcal{C}$ of r.e. sets is Vac^* learnable, then there is a recursive learner M for $\mathcal{C}$ such that M BC^* and $Part$ learns $\mathcal{C}$.*

Example 21. Case and Smith [4] showed that the class of recursive functions $\mathcal{F} = \{f : f \text{ is recursive} \wedge \forall^\infty x[f = \varphi_{f(x)}]\}$ is BC learnable but not Ex^* learnable. By the equivalence of Ex^* and Vac^* in the setting of learning recursive functions, $\mathcal{F}$ is also not Vac^* learnable. Furthermore, by Theorem 17, the class $\mathcal{F}$ witnesses the separation of Vac^* and BC^*Part learnability.

The following proposition shows that two relatively strong learning criteria can be synthesized to produce quite a strict learning criterion.

Proposition 22. *If a class $\mathcal{C}$ of r.e. sets is Vac^*WPart learnable, then $\mathcal{C}$ is Vac learnable.*

Proof. Assume that M is a Vac^*WPart learner of $\mathcal{C}$. Define a new learner N as follows. On input σ, let $e_0, e_1, \ldots, e_k$ be all the distinct conjectures of M on prefixes of σ. For each e_i, let p_i be the maximum number such that for all $x < p_i$, $x \in W_{e_i,|\sigma|}$ holds iff x is contained in $range(\sigma)$. Furthermore, let $q = \max(\{p_i : 0 \le i \le k\})$ and m be the least index such that $p_m = q$; N then outputs e_m.

Let $d_0, \ldots, d_l$ be all the distinct conjectures of M on some text T for an $L \in \mathcal{C}$. Since M is a $WPart$ learner, it must output at least one index for L. on T. Consider any d_i, d_j such that $W_{d_i} \neq L$ and $W_{d_j} = L$. Let z_i be the maximum number such that for all $x < z_i$, $x \in W_{d_i}$ holds iff $x \in L$. Then on almost all text prefixes $T[s]$, there must exist some $y_j > z_i$ such that for all $x < y_j$, $x \in W_{d_j,s+1}$ iff x is contained in $range(T[s])$. As there are only finitely many incorrect indices that M outputs, it follows that N will almost always output some index d_c for which $W_{d_c} = L$. Therefore N is a Vac learner of $\mathcal{C}$. ∎

The following proposition implies that vacillatory learning cannot in general be combined with partial learning; in other words, a vacillatorily learnable class may not necessarily be vacillatorily as well as partially learnable at the same time.

Proposition 23. *If a class $\mathcal{C}$ of r.e. sets is Vac*Part learnable, then $\mathcal{C}$ is Ex learnable.*

Proof. If M is a recursive learner of $\mathcal{C}$ such that on any text T for some $L \in \mathcal{C}$, M outputs only finitely many indices and outputs exactly one index d for L infinitely often, then M almost always outputs d on T. ∎

Remark 24. As a consequence of Proposition 23, *Vac** learning is more powerful than *Vac*Part* learning. To see this, it suffices to construct an *Ex** learnable class that is not *BC* learnable; one could take, for example, the class of all cofinite sets.

The following result shows that weak approximate learning is quite powerful.

Theorem 25. *The class of all infinite sets is WeakApprox learnable.*

Proof. Consider the learner M which conjectures on input σ the set

$$W_{M(\sigma)} = range(\sigma) \cup \{x : \forall y \in range(\sigma)\,[x > y]\}$$

and consider any text T for an infinite set. Let $S = \{x \in range(T)$: when x appears first in T, no larger datum of T has been seen so far$\}$. Note that the set S is infinite. Now all conjectures $M(T[n])$ are a superset of S: if an $x \in S$ has not yet appeared in $T[n]$ then all members of $range(T[n])$ are smaller than x and $x \in W_{M(T[n])}$ else x has already appeared in $T[n]$ and is therefore also in $range(T[n])$. Furthermore, if $x \notin range(T)$ then almost all n satisfy $\max(range(T[n])) > x$ and therefore $x \notin W_{M(T[n])}$, thus for every x almost all hypotheses $W_{M(T[n])}$ are correct at x. ∎

Unfortunately, the weakly approximate learning property of any class of infinite sets may be lost if finite sets are added to the target class.

Proposition 26. *Gold's class consisting of the set of natural numbers and all sets $\{0, 1, \ldots, m\}$ is not WeakApprox learnable.*

Proof. Make a text T where $T(0) = 0$ and iff the n-th hypothesis of the learner contains $T(n) + 1$ then $T(n + 1) = T(n)$ else $T(n + 1) = T(n) + 1$.

In the case that the text T is for a finite set with maximum m then $T(n) = m$ for almost all n and the n-th hypothesis contains $m + 1$ for almost all n; thus the approximations are in the limit false at $m + 1$.

In the case that the text T is for the set of all natural numbers then consider any $m > 0$ and consider the first n such that $T(n + 1) = m$. Then the n-th hypothesis does not contain m. Therefore, one can conclude that for every m there is an $n \geq m$ such that the n-th hypothesis is conjecturing m not to be in the set to be learnt although the set to be learnt is the set of all natural numbers. In particular there is no infinite set on which from some time on all approximations are correct.

Thus the class considered is not weakly approximately learnable. ∎

It may be observed that in the proof of Theorem 25, the parameter sets S with respect to which the learner M approximates the class of all infinite sets may not necessarily be r.e. (or be of any fixed Turing degree). This motivates the question of whether or not the class of all infinite sets is still weakly approximately learnable if one restricts the class of parameter sets in Definition 10 to some countable family.

Definition 27. For any sets L and W, where W is infinite, and any text T for L, say that a recursive learner M *weakly approximately (WeakApprox) learns L via W on T* iff for all finite variants V of W, it holds that for almost all hypotheses H of M on T, $H \cap V = L \cap V$. For any class $\mathcal{W}$ of infinite sets, a class S of sets is *weakly approximately (WeakApprox) learnable via $\mathcal{W}$* iff there is a recursive learner M such that for every $L \in S$ and every text T for L, M *WeakApprox* learns L via some $W \in \mathcal{W}$ on T.

Proposition 28. *For any countable class $\mathcal{W}$ of infinite sets, the class of all cofinite sets is not WeakApprox learnable via $\mathcal{W}$.*

Proof. Suppose M is a recursive learner that weakly approximately learns all cofinite sets via some countable class $\mathcal{W}$ of infinite sets. First, note that there exist $\sigma \in \mathbb{N}^*$ and $V \in \mathcal{W}$ such that for all $\tau \in \mathbb{N}^*$, $V \subseteq W_{M(\sigma\tau)}$. For, assuming otherwise, one can build a text T for $\mathbb{N}$ as follows. Let $V_0, V_1, V_2, \ldots$ be a one-one enumeration of $\mathcal{W}$, and set $T_0 = \epsilon$, where T_s denotes the text prefix built until stage s. Let m_s be the minimum number not contained in $range(T_s)$, and find strings $\eta_0, \eta_1, \ldots, \eta_s$ such that for all $i \in \{0, \ldots, s\}$, $V_i \not\subseteq W_{M(T_s \circ m_s \eta_0 \ldots \eta_i)}$; by assumption, such strings $\eta_0, \eta_1, \ldots, \eta_s$ must exist. Let $T = \lim_s T_s$. T is a text for $\mathbb{N}$; furthermore, for any $V_l \in \mathcal{W}$, $V_l \not\subseteq W_{M(T[s+1])}$ for infinitely many s, so that M does not weakly approximately learn $\mathbb{N}$ via V_l on T.

Now fix $\sigma \in \mathbb{N}^*$ and $V \in \mathcal{W}$ such that for all $\tau \in \mathbb{N}^*$, $V \subseteq W_{M(\sigma\tau)}$. As V is infinite, one can choose some $w \in V - range(\sigma)$. Let T' be a text for $\mathbb{N} - \{w\}$ that extends σ. Then M conjectures a set containing w on almost all text prefixes of T', which shows that it cannot weakly approximately learn $\mathbb{N} - \{w\}$. In conclusion, the class of all cofinite sets is not weakly approximately learnable via $\mathcal{W}$. ∎

Theorem 29. *Assume that M is a BC^* learner for $\mathcal{C}$ such that M outputs for each $L \in \mathcal{C}$ at least once a fully correct hypothesis, that is, M is a BC^* WPart learner for $\mathcal{C}$. Then there is a partial learner N for $\mathcal{C}$ which is at the same time a BC^* learner.*

Proof. Let M be given and on a text T, let $e_0, e_1, \ldots$ be the sequence of hypotheses by M. Now one can make a learner O which on input $T(0)T(1)\ldots T(n)$, first computes $e_0, e_1, \ldots, e_n$ and then computes for every e_m the quality $q_{m,n}$ which is the maximal number $y \leq n$ such that for all $x \leq y$ the number x has been enumerated into $W_{e,n}$ iff $x \in \{T(0), T(1), \ldots, T(n)\}$. In each step the learner O outputs either the hypothesis for the least m such that either (a) e_m has been output so far less than $q_{m,n}$ times or (b) all $k \leq n$ satisfy that e_k has been output

$q_{k,n}$ times and $q_{k,n} \leq q_{m,n}$. One can see that false hypotheses e_m get output only finitely often output while at least one correct hypotheses gets output infinitely often; as all but finitely many hypotheses of M are finite variants of L, the same is true for the modified learner O. By combining this learner O with the padding given in Proposition 3, one can obtain a resulting learner N which is at the same time a BC^* learner and a partial learner. ∎

Proposition 30. *If $\mathcal{C} = \{L_0, L_1, L_2, \ldots\}$ is a uniformly recursive family of sets that is BC^* learnable with respect to a uniformly recursive hypothesis space, then $\mathcal{C}$ is $Ex^*[K']$ learnable with respect to the hypothesis space $\{L_0, L_1, L_2, \ldots\}$.*

Proof. Let $\{H_0, H_1, H_2, \ldots\}$ be the uniformly recursive hypothesis space that the learner M uses. Now construct an $Ex^*[K']$ learner N as follows. On input τ, N queries K' to find the least $i \leq |\tau|$ (if such an i exists) such that L_i is a finite variant of $H_{M(\tau)}$. If no such i exists, then N conjectures $range(\tau)$. It is immediate from the construction of N that for any text T of some L_j, since M outputs indices for finite variants of L_j on almost all prefixes T, N will converge to the least index l such that L_l is a finite variant of L_j. ∎

The subsequent proposition gives a negative answer to the question as to whether either partial or consistent partial learning can be combined with weakly approximate learning.

Proposition 31. *The class consisting of all sets which either contain all even and some odd numbers or contain all odd numbers and some even numbers or contain all natural numbers does not have a learner which is (a) weakly approximating and (b) partial at the same time, though it has learners which satisfy (a), learners which satisfy (b), and is also ConsPart learnable.*

Proof. That (a) and (b) can be satisfied by different learners follows from known results (for partial learning) and Theorem 25 (for weak approximation). The class is also *ConsPart* learnable by Theorem 4. Furthermore, one can easily make a text T which makes sure that a given partial learner M for the class does not also weakly approximate it. The idea is to define the text T inductively as follows by going through the following loop:

1. Let $n = 0$;
2. As long as $M(T[n])$ does not conjecture a set which contains all even numbers and only finitely many odd numbers let $T(n)$ be the least even number not yet in the text and update $n = n + 1$;
3. As long as $M(T[n])$ does not conjecture a set which contains all odd numbers and only finitely many even numbers let $T(n)$ be the least odd number not yet in the text and update $n = n + 1$;
4. Go to Step 2.

It is easy to see that as the learner is partial it cannot get stuck in Step 2 or Step 3 forever, as it would not output an index for $range(T)$ infinitely often in that case. Hence it alternates between Steps 2 and 3 infinitely often and will therefore alternating between sets containing all even and only finitely many odd numbers and all odd and only finitely many even numbers. Hence there is no infinite set which is contained in almost all hypotheses; however, the range of T is the set of natural numbers and thus the learner is not weakly approximating it. ∎

The following theorem considerably improves an earlier result of Gao, Stephan and Zilles [7] that every conservatively partially learnable class is also BC^* learnable.

Theorem 32. *If a class $\mathcal{C}$ has a conservative partial learner M then $\mathcal{C}$ has also a learner N which is (a) a BC^* learner, (b) a partial learner and (c) an approximate learner.*

Proof. Let M be given and for a text T for a language $L \in \mathcal{C}$, one considers the sequence $e_0, e_1, \ldots$ of distinct hypotheses issued by M; note that one correct hypothesis is there and all others are not containing L. For each hypothesis e_n one has two numbers which track its quality: $b_{n,t}$ is the maximal $s \leq n+t$ such that all $T(u)$ with $u < s$ are in $W_{e_n,n+t} \cup \{\#\}$ and $a_{n,t} = 1 + \max\{b_{m,t} : m < n\}$.

Now one defines the hypothesis set $H_{e_n,\sigma}$ where σ is any sequence by defining $e_{n,0}, e_{n,1}, \ldots$ be a sequence with $e_{n,0} = e_n$ and $e_{n,u}$ be the e_m for the minimum m such that $m = n$ or W_{e_m} has enumerated all members of $range(\sigma)$ within $u + t$ time steps. The set $H_{e_n,\sigma}$ contains all x for which there is an $u \geq x$ with $x \in W_{e_{n,u}}$.

An intermediate learner O now conjectures some canonical index of a set $H_{e_n,\sigma}$ at least k times iff there is a t with $\sigma = T(0)T(1)\ldots T(a_{n,t})$ and $b_{n,t} > k$. Thus O conjectures $H_{e_n,\sigma}$ infinitely often iff W_{e_n} contains $range(T)$ and $a_{n,t} = |\sigma|$ for almost all t.

If e_n is the correct index for the set to be learnt then, by conservativeness, the sets W_{e_m} with $m < n$ do not contain the set to be learnt as a subset and therefore the values $b_{m,t}$ converge which implies that $a_{n,t}$ converges to some fixed value s. It follows that for the prefix σ of T of length s, the canonical index of $H_{e_n,\sigma}$ is conjectured infinitely often while no other index is conjectured infinitely often. Thus O is a partial learner. Furthermore, for all sets $H_{e_m,\tau}$ conjectured after $a_{n,t}$ has reached its final value s, it holds that the $e_{m,u}$ in the construction of $H_{e_m,\tau}$ converge to e_n and therefore the set $H_{e_m,\tau}$ is the union of W_{e_n} and a finite set. Thus the learner O is a BC^* learner. However, the third condition on approximate learning is not yet guaranteed and therefore O will be translated into another learner N.

Let $d_0, d_1, \ldots$ be the sequence of O output on the text T. Now N will copy this sequence but with some delay. Assume that $N(\sigma_k) = d_k$ and σ_k is a prefix of T. Now N will keep the hypothesis d_k until the current prefix σ_{k+1} considered satisfies either that $range(\sigma_{k+1}) \not\subseteq range(\sigma_k)$ or $W_{d_k,|\sigma_{k+1}|} \neq range(\sigma_{k+1})$.

In the case that the range of T is infinite, the sequence of hypotheses of N will be the same as the sequence of hypotheses of O, only some additional delay and repeating of hypotheses might be there. Furthermore, almost all W_{d_n} contain $range(T)$, thus the resulting learner N learns $range(T)$ and is almost always correct on the infinite set $range(T)$, thus N learns $range(T)$ partially, in the sense BC^* and approximately. If $range(T)$ is finite, there will be some correct index which equals to infinitely many d_n. There is a step t by which all elements of $range(T)$ have been seen in the text and also be enumerated into this set. Therefore, when the learner conjectures this correct index again, it will never withdraw it; furthermore, it will replace eventually every incorrect conjecture due to the comparison of the two sets. Thus the learner converges explanatorily to $range(T)$ and is also in this case learning $range(T)$ in a BC^* way, partially and approximately. It follows that N has all the three requirements (a), (b) and (c). ∎

Theorem 33. *If a class $\mathcal{C}$ has a BC^* learner then it is WeakApprox learnable.*

Proof. Let M be the learner which weakly approximates the class of all infinite sets. Let O be a given BC^* learner for $\mathcal{C}$. Now the new learner N is given as follows:

On input σ, $N(\sigma)$ outputs an index of the following set which first enumerates $range(\sigma)$ and then searches for τ such that the following conditions holds:

- $range(\tau) = range(\sigma)$;
- $|\tau| = 2 * |range(\sigma)|$;
- $W_{O(\tau\#^s)}$ enumerates at least $|\sigma|$ many elements for all $s \leq |\sigma|$.

If all three conditions are satisfied then the set enumerates also all elements of $W_{M(\sigma)}$.

If $L \in C$ is finite then for every τ of length $2 * |L|$ with range L, the learner outputs on some input $\tau\#^{s_\tau}$ a finite set with c_τ many elements. As there are only finitely many such τ, there is an upper bound t of all c_τ and s_τ. Then it follows from the construction that the learner N on any input σ with $range(\sigma) = L$ and $|\sigma| \geq t$ outputs a hypothesis for the set L, as the corresponding τ cannot be found. Thus N weakly approximately learns L.

If $L \in \mathcal{C}$ is infinite then there is a locking sequence $\gamma \in L^*$ for L such that $O(\gamma\eta)$ conjectures an infinite set whenever $\eta \in L^*$. It follows for all σ with $range(\gamma) \subseteq range(\sigma)$ and $|range(\sigma)| > |\gamma|$ that $N(\sigma)$ considers also a τ which is an extension of γ in its algorithm and which therefore meets all three conditions, thus $N(\sigma)$ will conjecture a set consisting of the union of $range(\sigma)$ and $W_{M(\sigma)}$. As adding $range(\sigma)$ to the hypothesis $W_{M(\sigma)}$ cannot make $W_{N(\sigma)}$ to be incorrect at any x where $W_{M(\sigma)}$ is correct, it follows that also N is weak approximately learning L. Thus, by case distinction, N is a weak approximate learner for $\mathcal{C}$. ∎

The next theorem shows that neither partial learning nor consistent partial learning can be combined with approximate learning. In fact, it establishes a stronger result: consistent partial learnability and approximate learnability are insufficient to guarantee both partial and weakly approximate learnability simultaneously.

Theorem 34. *There is a class with the following properties:*
(a) The class is not BC^* learnable;
(b) The class does not have a learner which is both, partial and weakly approximate;

(c) The class is approximately learnable;
(d) The class is $Ex[K']$ learnable.
(e) The class is ConsPart learnable.

Proof. Since the three constraints, BC^* learning, partial learning and weak approximate learning, are all compatible with delaying, one can use a list M_0, M_1, ... of learners which are all total and which contains for every learner to be considered a delayed version. This permits to ignore the case that some learner is undefined on some input.

The class consists now of all sets L_d such that for each d, either L_d is $\{d, d+1, \ldots\}$ or L_d is a subset built by a diagonalisation-procedure. This diagonalisation-procedure assigns to each number $x \geq d$ the level $\ell(x)$. The key idea of the construction is the following:

– If some set $L_{d,e} = \{x \geq d : \ell(x) \leq e\}$ is infinite
– then let $L_d = L_{d,e}$ for the least such e and M_d does not partially learn L_d
– else let $L_d = \{d, d+1, \ldots\}$ and M_d does not weakly approximate L_d.

The construction of the sets is inductive over stages. At stage s one decides on $\ell(d+s)$ which then implicitly decides on the other parameters as well. One does the following algorithm for each $s = 0, 1, 2, \ldots$:

– Let τ_e be a sequence of all $x \in \{d, d+1, \ldots, d+s-1\}$ with $\ell(x) = e$ in ascending order;
– If there is an $e < s$ such that e has not been cancelled in any previous step and for each $\eta \preceq \tau_e$ the intersection

$$W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta),s} \cap \{y : d \leq y < d+s \wedge \ell(y) > e\}$$

contains at least $|\tau_e|$ elements
– Then choose the least such e and let $\ell(d+s) = e$ and cancel all e' with $e < e' \leq s$
– Else let $\ell(d+s) = s$.

One can define inductively a text $T = \lim_e \sigma_e$ as follows (where σ_0 is the empty sequence):

– Let τ_e be the sequence of all x with $\ell(x) = e$ in ascending order;
– If σ_e is finite then let $\sigma_{e+1} = \sigma_e\tau_e$ else let $\sigma_{e+1} = \sigma_e$.

In the case that some σ_e are infinite, let e be the least index where this happens. Then $T = \sigma_e$ and $L_d = L_{d,e}$ and T is a text for L_d. As $L_{d,e}$ is infinite, one can conclude that

$$\forall \eta \preceq \sigma_e \, \forall c \, [|W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta)} \cap \{y : \ell(y) > e\}| \geq c]$$

and therefore M_d outputs on the text T for L_d almost always a set which contains infinitely many elements outside L_d; so M_d does neither partially learn L_d nor BC^* learn L_d.

In the case that all σ_e are finite and therefore all $L_{d,e}$ are finite there must be infinitely many e which never get cancelled. Each such e satisfies

$$\exists \eta \preceq \tau_e \, [W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta)} \cap \{y : \ell(y) > e\} \text{ is finite}]$$

and therefore e also satisfies

$$\exists \eta \preceq \tau_e \, [W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta)} \text{ is finite}].$$

Thus M_d outputs on the text T for the cofinite set $L_d = \{d, d+1, \ldots\}$ infinitely often a finite set and M_d is neither weakly approximately learning L_d (as there is no infinite set on which almost all conjectures are correct) nor BC^*-learning L_d. Thus conditions (a) and (b) are true.

Now it is shown that the class of all L_d is approximately learnable by some learner N. This learner N will on a text for L_d eventually find the minimum d which is needed to compute the function ℓ correctly. Once N has found this d, N will on each input σ conjecture the set

$$W_{N(\sigma)} = \{x : x \geq \max(range(\sigma)) \vee \exists y \in range(\sigma) \, [\ell(x) \leq \ell(y)]\}$$

In the case that $L_d = L_{d,e}$ for some e then $L_{d,e}$ is infinite and for each text and almost all prefixes σ of this text satisfy $\max\{\ell(y) : y \in range(\sigma)\} = e$ and $L_{d,e} \subseteq W_{N(\sigma)}$. So almost all conjectures are correct on the infinite set L_d itself. Furthermore, $W_{N(\sigma)}$ does not contain any $x < \max(range(\sigma))$ with $\ell(x) > e$, hence N gets in the limit correct also on any $x \notin L_{d,e}$ and therefore N approximates $L_{d,e} = L_d$.

In the case that $L_d = \{d, d+1, \ldots\}$ then all $L_{d,e}$ are finite and one can look at the set $S = \{x : \forall y > x \, [\ell(y) > \ell(x)]\}$. Note that S is infinite. Let $x \in S$ and consider any σ with $\min(range(\sigma)) = d$. If $x \geq \max(range(\sigma))$ then $x \in W_{N(\sigma)}$. If $x < \max(range(\sigma))$ then $\ell(\max(range(\sigma))) \geq \ell(x)$ and again $x \in W_{N(\sigma)}$. Thus $W_{N(\sigma)}$ contains S. Furthermore, for all $x \geq d$ and sufficiently long prefixes σ of the text, $\ell(\max(range(\sigma))) \geq \ell(x)$ and therefore all $x \in W_{N(\sigma)}$ for almost all prefixes σ of the text. So again N approximates L_d. Thus condition (c) is true.

Furthermore, there is a K'-recursive learner O which explanatorily learns the class. On input σ with at least one element in $range(\sigma)$, the learner determines $d = \min(range(\sigma))$. If there is now some $e \leq |\sigma|$ such that $L_{d,e}$ is infinite then O conjectures $L_{d,e}$ for the least such e else O conjectures $\{d, d+1, \ldots\}$. It is easy to see that these hypotheses converge to the set L_d to be learnt: eventually the minimum of the range of each input is d. In the case that $L_d = L_{d,e}$ for some e this e is detected whenever the input is longer than e and therefore the learner converges to $L_{d,e}$. In the case that all $L_{d,e}$ are finite, the learner almost always outputs the same hypothesis for $\{d, d+1, \ldots\}$. Thus O is a $Ex[K']$ learner and condition (d) is true.

It remains to show that the class is *ConsPart* learnable. This follows from the fact that the class is a subclass of the uniformly recursive family $\mathcal{U} = \{L_{e,d}\}_{e,d \in \mathbb{N}} \cup \{\{d + x : x \in \mathbb{N}\} : d \in \mathbb{N}\}$. To see that $\mathcal{U}$ is uniformly recursive, it may be observed from the construction of $L_{e,d}$ that for each d, $\ell(x)$ is defined for all $x \geq d$; each of these values, moreover, can be calculated effectively. Thus one can uniformly decide for all d, e and y whether or not $y \geq d$ and $\ell(y) \leq e$, that is, whether or not $y \in L_{e,d}$. Consequently, by Theorem 4, the given class is consistently partially learnable, as required. ∎

Theorem 35. *If $\mathcal{C}$ is ConsvEx$[K']$ learnable, then $\mathcal{C}$ is WeakApproxBC* learnable.*

Proof. Let M be the learner which conservatively explanatorily learns the class with oracle $A = \{a : W_a \text{ is infinite}\}$ which is Turing equivalent to K'. Now first one makes a new learner O which is a BC^* learner and almost always outputs a superset of the set to be learnt. Assume an input σ is given.

Now for each finite string α of length up to σ and each $\tau \preceq \sigma$ one computes the following hypothesis $W_{O(\tau,\sigma,\alpha)}$: If $M^\alpha(\tau)$ is defined and in particular all oracle queries can be answered by using α and $W_{M^\alpha(\tau)}$ contains $range(\sigma)$ then $W_{O(\sigma,\tau,\alpha)}$ contains all x for which there is a stage $s > |\sigma|$ with $x \in W_{O(\tau,\sigma,\alpha),s}$ and $|W_{a,s}| > |\sigma| + x$ for all a with $\alpha(a) = 1$ and $|W_{a,s}| < |\sigma|$ for all a with $\alpha(a) = 0$ else $W_{O(\sigma,\tau,\alpha)}$ is empty. Note that $W_{O(\sigma,\tau,\alpha)}$ is infinite iff $M^\alpha(\tau)$ is defined and $range(\sigma) \subseteq W_{M^\alpha(\tau)}$ and $W_{M^\alpha(\tau)}$ is infinite and $\alpha \preceq A$ and all a with $\alpha(a) = 0$ satisfy that W_a has less than $|\sigma|$ elements. Now $W_{O(\sigma)}$ is the union of all $W_{O(\sigma,\tau,\alpha)}$ with $\tau \preceq \sigma$ and $|\alpha| \leq |\sigma|$.

It follows that when M^A learns $L \in \mathcal{C}$ and σ is a sufficiently long prefix of a given text of L that $range(\sigma)$ is not contained in any of the finitely many wrong indices output by M^A on this text and that there is $\tau \preceq \sigma$ such that $M^A(\tau)$ is correct and can be computed without querying A beyond $|\sigma|$. Then it follows that $W_{O(\sigma)}$ is the union of L and a finite set, hence O is a BC^* learner.

Now one forms a learner N from O in a way which is similar to Theorem 33. This new learner N is given as follows.

On input σ, $N(\sigma)$ outputs an index of the following set which first enumerates $range(\sigma)$ and then searches for τ such that the following conditions holds:

- $range(\tau) = range(\sigma)$;
- $|\tau| = 2 * |range(\sigma)|$;
- $W_{O(\tau\#^s)}$ enumerates at least $|\sigma|$ many elements for all $s \leq |\sigma|$.

If all three conditions are satisfied then the set enumerates also all elements $x \in \{W_{O(\sigma)}$ with $\forall y \in range(\sigma) : y < x\}$.

If $L \in C$ is finite then for every τ of length $2 * |L|$ with range L, the learner outputs on some input $\tau\#^{s_\tau}$ a finite set with c_τ many elements. As there are only finitely many such τ, there is an upper bound t of all c_τ and s_τ. Then it follows from the construction that the learner N on any input σ with $range(\sigma) = L$ and $|\sigma| \geq t$ outputs a hypothesis for the set L, as the corresponding τ cannot be found. Thus N weak approximately BC^* learns L.

If $L \in \mathcal{C}$ is infinite then there is a locking sequence $\gamma \in L^*$ for L such that $O(\gamma\eta)$ conjectures a superset of L whenever $\eta \in L^*$. It follows for all σ with $range(\gamma) \subseteq range(\sigma)$ and $|range(\sigma)| > |\gamma|$ that $N(\sigma)$ considers also a τ which is an extension of γ in its algorithm and which therefore meets all three conditions, thus $N(\sigma)$ will conjecture a set consisting of the union of $range(\sigma)$ and the elements of $W_{O(\sigma)}$ which are greater than every element of $W_{O(\sigma)}$ which includes all elements of L which are greater than $\max(range(\sigma))$. One can now verify as in Theorem 25 the that learner N is approximating; furthermore, all of its hypotheses are finite variants of those of O and therefore finite variants of L. Thus again N is an approximating BC^* learner for L. Thus, by case distinction, N is a weak approximate learner for $\mathcal{C}$. ∎

Now consider a watered-down version of approximate learning, whereby the learner M on any text T for a target language L is only required to be almost always correct on any finite set; that is, for any given finite set D, it holds that for almost all hypotheses H of M, $H \cap D = L \cap D$. It will be shown that this learning criterion is so powerful that the whole class of r.e. sets can be finitely approximated and partially learnt at the same time.

Definition 36. A class S of languages is *finitely approximately (FinApprox) learnable* iff for every language $L \in S$, all texts T for L, and any finite set D, it holds that for almost all hypotheses H of the learner on T, $H \cap D = L \cap D$.

Theorem 37. *The class of all r.e. sets is FinApproxPart learnable.*

Proof. Let M_1 be a partial learner of all r.e. sets. Define a learner M_2 as follows. On text T, let $e_0, e_1, e_2, \ldots$ be all the distinct conjectures of M_1. For each s such that $T(s) \in range(T[s+1]) - range(T[s]) - \{\#\}$ and for each $t < T(s)$, M_2 outputs a fixed index d for $range(T[s+1]) \cup (W_{e_t} \cap \{y : y > T(s)\})$ at least n times iff there is a stage $u > n$ such that for all $x < n$, it holds that $x \in range(T[u+1])$ iff $x \in W_{d,u}$. At every other stage where no set of the form $range(T[s+1]) \cup (W_{e_t} \cap \{y : y > T(s)\})$ must be conjectured or where no new datum is seen, M_2 outputs a canonical index for the range of the current input.

Let T be any given text for an r.e. set L. If L is finite, then M_2 will almost always output a canonical index for L. Suppose that L is infinite. Note that M_1 on T will output exactly one index e for L infinitely often. Furthermore, there are infinitely many stages s at which $range(T[s+1]) \cap \{0, \ldots, T(s)\} = L \cap \{0, \ldots, T(s)\}$. Thus $range(T[s+1]) \cup (W_e \cap \{y : y > T(s)\}) = L$, so that M_2 will output a fixed index for $range(T[s+1]) \cup (W_e \cap \{y : y > T(s)\})$ infinitely often. Furthermore, for every z there is a least k such that $range(T[l+1]) \cap \{0, \ldots, z\} = L \cap \{0, \ldots, z\}$ for all $l > k$; for each $l' \leq k$ such that $T(l') \in range(T[l'+1]) - range(T[l']) - \{\#\}$, since M_2 will conjecture incorrect sets of the form $range(T[l'+1]) \cup (W_{e_h} \cap \{y : y > T(l')\})$ for some $h < T(l')$ only finitely often, M_2 will almost always conjecture sets H for which $H \cap \{0, \ldots, z\} = L \cap \{0, \ldots, z\}$. Therefore M_2 finitely approximately learns $\mathcal{C}$, implying that it never outputs any incorrect index infinitely often. One may then apply Proposition 3 to construct a further learner M_3 that outputs on every text for any r.e. set L exactly one index e with $W_e = L$ infinitely often, and outputs every other index only finitely often; in addition, M_3 finitely approximately learns L. ∎

Theorem 38. *If $\mathcal{C}$ is a uniformly recursive family of sets, then $\mathcal{C}$ is FinApprox-ConsPart learnable.*

Proof. Let $\mathcal{C} = \{L_0, L_1, L_2, \ldots,\}$ be a uniformly recursive family of sets. Define a learner M_1 as follows. M_1 maintains a list I of pairs $\{(e_0, m_0), (e_1, m_1), \ldots, (e_n, m_n)\}$, which it updates whenever a new datum appears in the input. At stage s, I is set as I_s. At the initial stage, M_1 sets $I_0 = \emptyset$. For any string τ, let τ' denote the string obtained from τ by deleting its last symbol. On input σ with $s = |\sigma|$, if $range(\sigma) - \{\#\} = range(\sigma') - \{\#\}$, then M_1 outputs a canonical index for $range(\sigma)$. Otherwise, let $range(\sigma) - range(\sigma') - \{\#\} = \{x\}$; M_1 then determines all $i \leq s$ such that $L_i \cap \{0, \ldots, x\} = range(\sigma) \cap \{0, \ldots, x\}$. For each such i, if there is a maximum $t \leq s$ such that $(i, j) \in I_t$ for some j, then M_1 adds $(i, j+1)$ to I_{s+1}; otherwise, M_1 adds $(i, 1)$ to I_{s+1}. I is then set as I_{s+1}. Furthermore, M_1 at this stage outputs a fixed index e' for L_e such that $(e, l) \in I_h$ for some $l, h \leq s+1$ and $range(\sigma) - \{\#\} \subseteq L_e$. M_1 chooses e so that on the given text, for every pair (i', j') such that $(i', j') \in I_g$ for some g, M_1 outputs a fixed index for $L_{i'}$ at least j' times; this can be achieved if the range of the given text is infinite because each I_k is finite. If no such e exists, then M_1 outputs a canonical index for $range(\sigma) - \{\#\}$.

It follows directly by construction that M_1 is consistent. Now suppose that M_1 is fed with a text T for some $L \in \mathcal{C}$. If L is finite, then there is a least m such that $range(T[m]) - \{\#\} = L$, so that N will output a canonical index for L on almost all prefixes of T. If L is infinite, then I will be updated at infinitely many stages; furthermore, there are infinitely many stages s at which $range(T[s+2]) - range(T[s+1]) - \{\#\}$ contains some $x \in L$ such that $range(T[s+2]) \cap \{0, \ldots, x\} = L \cap \{0, \ldots, x\}$. Thus for every i such that $L_i = L$ and every j, there is a t such that $(i, j) \in I_t$. This implies that M_1 will output a fixed index for L_i infinitely often. Further, for every x, there is a least k such that $range(T[k+1]) \cap \{0, \ldots, x\} = L \cap \{0, \ldots, x\}$; consequently, for all $t > k$ and all $(i', j') \in I_t$, it must hold that $L_{i'} \cap \{0, \ldots, x\} = L \cap \{0, \ldots, x\}$. As there are only finitely many pairs (i'', j'') that are contained in I_t for some $t \leq k$, this means that almost all hypotheses d of M_1 satisfy $W_d \cap \{0, \ldots, x\} = L \cap \{0, \ldots, x\}$. Therefore M_1 is a finitely approximate learner. Note that since M_1 is a finitely approximate learner, M_1 never outputs any incorrect index infinitely often. Hence M_1 is a consistent and finitely approximate learner that outputs at least one index for L infinitely often and outputs any incorrect index only finitely often. One may then apply the padding construction in Proposition 3 to build a new learner M that is consistent, finitely approximate, and outputs exactly one index for L infinitely often. In conclusion, $\mathcal{C}$ is both finitely approximately learnable as well as consistently partially learnable at the same time. ∎

Proposition 38 and Theorem 4 together give the following corollary.

Corollary 39. *If $\mathcal{C}$ is ConsPart learnable, then $\mathcal{C}$ is FinApproxConsPart learnable.*

The next result shows conservative partial learning may always be combined with finitely approximate learning.

Theorem 40. *If $\mathcal{C}$ is ConsvPart learnable, then $\mathcal{C}$ is FinApproxConsvPart learnable.*

Proof. Let M be a conservative partial learner of $\mathcal{C}$. Given a text T, let e_0, e_1, e_2, $\ldots$ be all the distinct conjectures of M on prefixes of T. For each pair (n, m), fix an index $d_{n,m}$ so that $W_{d_{n,m}} = (W_{e_n,m} \cap \{0, \ldots, n\}) \cup (W_{e_n} \cap \{x : x > n\})$. Furthermore, fix $d'_{n,m}$ so that $W_{d'_{n,m}} = \bigcup_{s \in A_{n,m}} W_{d_{n,m}}$, where $s \in A_{n,m}$ iff $W_{d_{n,m},s} \subset W_{d_{n,m},s+1}$. Define a learner N on text T as follows. At any stage s where no new datum appears, that is, $range(T[s+1]) - \{\#\} = range(T[s]) - \{\#\}$, N outputs a canonical index for $range(T[s+1]) - \{\#\}$. At every stage where a new datum appears in the input, N outputs, if necessary, some index of the form $d'_{n,m}$; N chooses this index so that on T, $d'_{n,m}$ is output at least k times iff there is an $s > k$ for which the following conditions hold:

1. $range(T[k + n + m]) - \{\#\} \subseteq W_{d_{n,m}}$;
2. $W_{d_{n,m},n+k+s} - \{\#\} \subseteq range(T[s + n + m])$;
3. $m < s$ and m is the least number with $W_{d_{n,m},n+k+s} = W_{e_n,n+k+s}$;
4. M outputs e_n at least k times.

If, at the present stage, there is no index of the form $d_{n,m}$ that must be output, N outputs a canonical index for the range of the current input. By construction, if T is a text for a finite set D, then N on T will converge to a canonical index for $range(D)$; furthermore, every other index that N outputs on T is not a superset (proper or otherwise) of D. Now suppose that T is a text for an infinite set in $\mathcal{C}$. Note that N is a finitely approximate learner: for each n, N outputs $d'_{n,m}$ at least once only if $W_{d_{n,m}}$ is a subset of W_{e_n} such that $W_{d_{n,m}} \cap \{0, \ldots, n\} = range(T[n + m]) \cap \{0, \ldots, n\}$; at any stage where N does not output an index of the form $d'_{n,m}$, it conjectures a canonical index for the range of the current input. Furthermore, N is conservative as it either conjectures proper subsets of each W_{e_n} or outputs exactly one index $d'_{n,m}$ such that $W_{d'_{n,m}} = W_{d_{n,m}} = W_{e_n}$; moreover, no finite set that N conjectures is a superset of $range(T)$. In addition, suppose that M outputs e_n infinitely often on T. If m is the least number such that $W_{e_n,m} \cap \{0, \ldots, n\} = W_{e_n} \cap \{0, \ldots, n\}$, then N outputs $d'_{n,m}$ infinitely often and outputs every other index of the form $d'_{n',m'}$ only finitely often; moreover, as $range(T)$ is infinite, N conjectures finite sets only finitely often. Thus N is a finitely approximate conservative partial learner of $\mathcal{C}$. ∎

Jain, Stephan and Ye [11] proved that for uniformly r.e. classes, class-comprising explanatory learning is equivalent to *uniform* explanatory learning; the latter means that one can construct a numbering of partial-recursive learners M_0, M_1, $M_2, \ldots$ such that for any given r.e. numbering $H_0, H_1, H_2, \ldots$ of the target class $\mathcal{C}$ with $W_e = \{\langle d, x \rangle : x \in H_d\}$, the e-th learner explanatorily learns $\mathcal{C}$ with respect to $\{H_0, H_1, H_2, \ldots\}$. In particular, uniformly r.e. explanatorily learnable classes are always explanatorily learnable with respect to a class-preserving hypothesis space. The next theorem shows, however, that none of the approximate learning criteria considered so far can be combined with class-preservingness. This may be interpreted intuitively as follows: in general, any successful approximation of languages must involve sets not contained in the target hypothesis space.

Theorem 41. *There is a uniformly r.e. class that is Approx learnable but not ClsPresvFinApprox learnable.*

Proof. Let $M_0, M_1, M_2, \ldots$ be an enumeration of all partial-recursive learners. For each e, define a strictly increasing r.e. sequence $x_{e,1}, x_{e,2}, \ldots$ as follows. First, for any given finite set D and number $y \notin D$, let $\alpha_{D,y}$ denote the string $1 \circ 4 \circ \ldots \circ 3y+1$, which is a concatenation (in increasing order) of all numbers of the form $3z+1$ with $0 \leq z \leq y$ and $z \notin D$. $x_{e,1}$ is defined to be the first number found (if such a number exists) such that for some $m_{e,1}$ with $x_{e,1} > m_{e,1}$, it holds that $\{3e, 3x_{e,1}+1\} \subseteq W_{M_e(3e \circ \alpha_{\emptyset, m_{e,1}})}$. Suppose that $x_{e,1}, \ldots, x_{e,k}$ have been defined. $x_{e,k+1}$ is then defined to be the first number found (if such a number exists) such that for some $m_{e,k+1}$ with $x_{e,k+1} > m_{e,k+1} > x_{e,k}$, it holds that $\{3e, 3x_{e,k+1}+1\} \subseteq W_{M_e(3e \circ \alpha_{\{x_{e,i}: 1 \leq i \leq k\}, m_{e,k+1}})}$.

For each pair $\langle e, i \rangle$, define $L_{\langle e,i \rangle}$ according to the following case distinction.

Case (1): $x_{e,i}$ is defined for all i. Set $L_{\langle e,0 \rangle} = \{e\} \oplus (\mathbb{N} - \{x_{e,i} : i \in \mathbb{N}\}) \oplus \emptyset$. For each $j > 0$, set $L_{\langle e,j \rangle} = \{e\} \oplus (\mathbb{N} - \{x_{e,i} : i < j\}) \oplus \{0\}$.

Case (2): There is a minimum l such that $x_{e,l}$ is undefined. Set $L_{\langle e,0 \rangle} = \{e\} \oplus (\{y : (l = 1 \Rightarrow y < 0) \wedge (l > 1 \Rightarrow y < x_{e,l-1})\} - \{x_{e,i} : i < l\}) \oplus \emptyset$. For each j with $1 \leq j \leq l-1$, set $L_{\langle e,j \rangle} = \{e\} \oplus (\mathbb{N} - \{x_{e,i} : i < j\}) \oplus \{0\}$. Set $L_{\langle e,l \rangle} = \{e\} \oplus (\mathbb{N} - \{x_{e,i} : i < l\}) \oplus \emptyset$. For each $j \geq l+1$, set $L_{\langle e,j \rangle} = \emptyset$.

Set $\mathcal{C} = \{L_{\langle e,i \rangle} : e, i \in \mathbb{N}\}$.

Now it is shown that $\mathcal{C}$ is approximately learnable with respect to a class-comprising hypothesis space. On input σ, the learner M outputs a canonical index for $\emptyset$ if $range(\sigma)$ does not contain any multiple of 3. Otherwise, let e be the minimum number such that $3e \in range(\sigma)$; M then checks whether or not $2 \in range(\sigma)$. If $2 \in range(\sigma)$, M searches (with computational time bounded by $|\sigma|$) for the least l (if such an l exists) such that $3x_{e,l}+1 \in range(\sigma)$; it then conjectures $L_{\langle e,l \rangle}$. If no such l exists, M conjectures $L_{\langle e,1 \rangle}$. If $2 \notin range(\sigma)$, M searches for the minimum l' such that $x_{e,l'}$ has not yet been defined at stage $|\sigma|$. If $3x_{e,l'}+1 \notin range(\sigma)$, then M conjectures $L_{\langle e,0 \rangle}$. If $3x_{e,l'}+1 \in range(\sigma)$, then M outputs an index d such that

$$W_d = \begin{cases} range(\sigma) \cup \{3z+1 : (l' = 1 \Rightarrow 0 \leq z \leq s) & \text{if } s > x_{e,l'} \text{ is the first step at} \\ \wedge (l' > 1 \Rightarrow x_{e,l'-1}+1 \leq z \leq s)\} \cup L_{\langle e,0 \rangle} & \text{which } x_{e,l'} \text{ is defined;} \\ range(\sigma) \cup \{3z+1 : (l' = 1 \Rightarrow z \geq 0) & \text{if } x_{e,l'} \text{ is undefined.} \\ \wedge (l' > 1 \Rightarrow z \geq x_{e,l'-1}+1)\} & \end{cases}$$

For the verification that M approximately learns $\mathcal{C}$, suppose that M outputs the sequence of conjectures $e_0, e_1, e_2, \ldots$ on text T. Assume first that $x_{e,i}$ is defined for all i. If T is a text for $L_{\langle e,0 \rangle}$, then for almost all n, W_{e_n} is a finite variant of $L_{\langle e,0 \rangle}$; furthermore, if $e_{j_0}, e_{j_1}, \ldots$ is the subsequence of conjectures for which $W_{e_{j_i}} \neq L_{\langle e,0 \rangle}$, then the sequence $y_0, y_1, y_2, \ldots$ of minimum numbers such that $W_{e_{j_i}}(y_i) \neq L_{\langle e,0 \rangle}(y_i)$ is almost always monotone increasing and contains a strictly increasing subsequence. In addition, for almost all i, $W_{e_i}(y) = L_{\langle e,0 \rangle}(y)$ for all y contained in $L_{\langle e,0 \rangle}$, which is an infinite set. Hence M approximately learns $L_{\langle e,0\rangle}$. If T is a text for $L_{\langle e,j\rangle}$ for some $j > 0$, then $2 \in range(T)$ and so M will eventually identify j as the minimum l such that $3x_{e,l}+1 \in range(T)$. Thus M will converge to an index for $L_{\langle e,j\rangle}$. Next, assume that there is a minimum l such that $x_{e,l}$ is undefined. If T is a text for $L_{\langle e,0\rangle}$, then M will in the limit identify l as the minimum l' such that $x_{e,l'}$ is undefined; thus, as $3x_{e,l'} + 1 \notin range(T)$, M on T will converge to an index for $L_{\langle e,0\rangle}$. If T is a text for some nonempty $L_{\langle e,j\rangle}$ with $j > 1$, M on T will again converge to an index for $L_{\langle e,j\rangle}$: if $2 \in L_{\langle e,j\rangle}$, then M will eventually identify j as the minimum number l such that $3x_{e,l} + 1 \in range(T)$ and converge to indices for $L_{\langle e,j\rangle}$; if $2 \notin L_{\langle e,j\rangle}$, then $3x_{e,j} + 1 \in range(T)$ and the fact that j is the minimum number for which $x_{e,j}$ is undefined together imply that M on T will converge to indices for $L_{\langle e,j\rangle}$. By construction, M converges to a canonical index for $\emptyset$ on any text with an empty range. This completes the verification that M approximately learns $\mathcal{C}$.

It remains to show that $\mathcal{C}$ is not *FinApprox* learnable using a class-preserving hypothesis space. Assume that M_e *ClsPresvFinApprox* learns $\mathcal{C}$. If there is a minimum l such that $x_{e,l}$ is undefined, then there is a text U for $L_{\langle e,l\rangle}$ on which M_e almost always outputs a conjecture that is different from $L_{\langle e,l\rangle}$. Since M_e finitely approximates $L_{\langle e,l\rangle}$, almost all of M_e's hypotheses on U must contain $3e$. But for all $j > 0$ such that $j \neq l$, either $L_{\langle e,j\rangle} = \emptyset$ or $2 \in L_{\langle e,j\rangle}$. As $2 \notin L_{\langle e,l\rangle}$ and $L_{\langle e,l\rangle}$ is infinite, while $L_{\langle e,0\rangle}$ is finite, it follows that M_e, being a finitely approximate learner, must almost always conjecture a set different from any $L_{\langle e,j\rangle}$ with $j \neq l$. Hence M_e is not a finitely approximate learner of $L_{\langle e,l\rangle}$. Suppose, on the other hand, that $x_{e,i}$ is defined for all i. Then one can build a text U' for $L_{\langle e,0\rangle}$ on which M_e infinitely often conjectures a set containing 2; but since $2 \notin L_{\langle e,0\rangle}$, it follows that M_e does not finitely approximately learn $L_{\langle e,0\rangle}$. This establishes that $\mathcal{C}$ is not *ClsPresvFinApprox* learnable. ∎

The main content of the following proposition may be summed up as follows: the quality of the hypotheses issued by a BC^* learner may be improved so that for any given finite set D, the learner's hypotheses will eventually agree with the target language on D.

Proposition 42. *If $\mathcal{C}$ is BC^* learnable, then $\mathcal{C}$ is FinApproxBC^* learnable.*

Proof. Given a BC^* learner M of $\mathcal{C}$, one can make a new learner N as follows. On input σ, N conjectures $range(\sigma) \cup (W_{M(\sigma)} \cap \{z : z > |\sigma|\})$. Suppose that N is fed with a text T for some $L \in \mathcal{C}$. N is a BC^* learner because it always conjectures finite variants of M's conjectures. Furthermore, for every finite set D there is some s_D such that $s_D > \max(D)$ and $range(T[s]) \cap D = L \cap D$ for all $s > s_D$. It follows by construction that for all $s > s_D$, $W_{N(T[s])} \cap D = range(T[s]) \cap D = L \cap D$, and so N finitely approximately BC^* learns L. ∎

The next two results consider combinations of finite approximation and some learning models that permit finitely many anomalies. It is readily seen that the additional constraint of finite approximation implies that any anomaly in the learner's hypotheses will eventually be corrected.

Proposition 43. *If $\mathcal{C}$ is Vac*FinApprox learnable, then $\mathcal{C}$ is Vac learnable.*

Proof. Given a *Vac*FinApprox* learner M of $\mathcal{C}$, suppose that on a text T for some $L \in \mathcal{C}$, M outputs an index i infinitely often iff $i \in \{e_0, \ldots, e_n\}$. Let $D = \bigcup_{0 \leq i \leq n}(W_{e_i} \triangle L)$. Note that by the finitely approximate learning property of M, there is an s such that for all $s' > s$, $W_{M(T[s'])} \cap D = L \cap D$. Thus $D = \emptyset$, and so $W_{e_i} = L$ for all $i \in \{0, \ldots, n\}$. Therefore M is a *Vac* learner of $\mathcal{C}$. ∎

Proposition 44. *If $\mathcal{C}$ is Ex*FinApprox learnable, then $\mathcal{C}$ is Ex learnable.*

Proof. Let M be an *Ex*FinApprox* learner of $\mathcal{C}$. Suppose M is fed with a text T for some $L \in \mathcal{C}$ and M converges on T to e. Then W_e is a finite variant of L, and by the finitely approximate learning property of M, $W_e \cap D = L \cap D$ for all finite sets D. Therefore $W_e = L$, and so M *Ex* learns L. ∎

As a side remark, *ConsPart* learning cannot in general be combined with *ConsvPart* learning. The final proposition establishes this fact.

Proposition 45. *There is a uniformly r.e. class of recursive sets that is both ConsPart learnable and ClsPresvConsvEx learnable but neither ConsEx learnable nor ConsConsvPart learnable.*

Proof. Let $\mathcal{C} = \{e + x : x \leq |W_e|\}$. First, note that $\mathcal{C}$ is contained in $\{\{e + x : x \in \mathbb{N}\} : e \in \mathbb{N}\} \cup \{\{e + x : x \leq d\} : e, d \in \mathbb{N}\}$, which has a uniformly recursive numbering. Thus by Theorem 4, $\mathcal{C}$ is *ConsPart* learnable. Second, let M_1 be a learner that outputs a canonical index for $\emptyset$ on any input with an empty range, and outputs a fixed index for $\{e + x : x \leq |W_e|\}$ on any input σ with $e = \min(\mathit{range}(\sigma))$. Then M_1 is a *ClsPresvConsvEx* learner of $\mathcal{C}$ as the minimum number e in any target set $L = \{e + x : x \leq |W_e|\}$ must occur at least once in any text for L; moreover, for any s_1, s_2 with $s_1 < s_2$, $M_1(T[s_1]) \neq M_1(T[s_2])$ only if $\mathit{min}(\mathit{range}(T[s_2])) \notin W_{M_1(T[s_1])}$, so that M_1 is conservative.

To prove that $\mathcal{C}$ is not *ConsConsvPart* learnable, note that by Theorem 5, it suffices to show that $\mathcal{C}$ is not *ConsEx* learnable. Assume by way of a contradiction that M *ConsEx* learns $\mathcal{C}$. It will be shown that one can use K as an oracle to decide for any given e whether or not W_e is finite. Let $L_e = \{e + x : x \leq |W_e|\}$. Assume a one-one ordering of all finite strings. One can use oracle K to find the first $\sigma \in L_e^*$ such that if $M(\sigma) = c$, then for all $\tau \in L_e^*$, $M(\sigma\tau) = c$. Note that such a σ must exist because M *Ex* learns $\mathcal{C}$; furthermore, $W_c = L_e$. Next, one can use oracle K again to determine whether or not there exists some $\eta \in (\{e + x : x \in \mathbb{N}\})^*$ such that $M(\sigma\eta) \neq c$. Suppose that W_e is finite. Then there exists a least l such that $e + l \notin W_c$. By the consistency of M, it must hold that $e + l \in M(\sigma \circ (e + l))$, so that $M(\sigma \circ (e + l)) \neq c$. Suppose that W_e is infinite. Then $L_e = \{e + x : x \in \mathbb{N}\}$, so that by the locking sequence property of σ, it holds that $M(\sigma\tau) = c$ for all $\tau \in (\{e + x : x \in \mathbb{N}\})^*$. Hence W_e is finite iff there exists some $\tau \in (\{e + x : x \in \mathbb{N}\})^*$ such that $M(\sigma \circ \tau) \neq c$, and therefore $\{e : |W_e| < \infty\} \leq_T K$, a contradiction. ∎

6 Conclusion

This paper introduced three notions of approximate learning of languages and studied the combinability of various learning criteria. It also positively resolved

Mark Fulk and Sanjay Jain's open question on whether the class of all recursive functions can be approximately learnt and BC^* learnt at the same time. However, questions on the combinability of some pairs of learning constraints remain open. In particular, it is unknown whether or not every BC^* learnable class has a learner that is both BC^* and $Part$.

References

1. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control* 45(2) (1980): 117–135.
2. Janis Bārzdiņš. Two theorems on the limiting synthesis of functions. In *Theory of Algorithms and Programs, vol. 1*, pages 82–88. Latvian State University, 1974. In Russian.
3. John Case and Chris Lynes. Machine inductive inference and language identification. *Proceedings of the Ninth International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science 140 (1982): 107–115.
4. John Case and Carl Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science* 25 (1983): 193–220.
5. Mark Fulk and Sanjay Jain. Approximate inference and scientific method. *Information and Computation* 114(2) (1994): 179–191.
6. Ziyuan Gao, Sanjay Jain and Frank Stephan. On conservative learning of recursively enumerable languages. *Ninth Conference on Computability in Europe*, CiE 2013, Milan, Italy, July 1–5, 2013. Springer LNCS 7921 (2013): 181–190.
7. Ziyuan Gao, Frank Stephan and Sandra Zilles. Partial learning of recursively enumerable languages. *Algorithmic Learning Theory, Twentyfourth International Conference*, ALT 2013, Singapore, October 6–9, 2013. Proceedings. Springer LNAI 8139: 113–127, 2013.
8. Ziyuan Gao, Frank Stephan and Sandra Zilles. Partial learning of recursively enumerable languages. Manuscript, 2014.
9. E. Mark Gold. Language identification in the limit. *Information and Control* 10 (1967): 447–474.
10. Sanjay Jain, Daniel Osherson, James S. Royer and Arun Sharma. 1999. *Systems that learn: an introduction to learning theory.* Cambridge, Massachusetts.: MIT Press.
11. Sanjay Jain, Frank Stephan and Nan Ye. Prescribed learning of r.e. classes. *Theoretical Computer Science* 410(19) (2009): 1796–1806.
12. Klaus P. Jantke and Hans-Rainer Beick. Combining postulates of naturalness in inductive inference. *Elektronische Informationsverarbeitung und Kybernetik* 17(8/9) (1981): 465–484.
13. Steffen Lange and Thomas Zeugmann. Language learning in dependence on the space of hypotheses. *Proceedings of the Sixth Annual ACM Conference on Computational Learning Theory*, pages 127–136, ACM Press, 1993.
14. Eric Martin and Daniel N. Osherson. 1998. *Elements of scientific inquiry.* Cambridge, Massachusetts.: MIT Press.
15. Daniel N. Osherson, Michael Stob and Scott Weinstein. Learning strategies. *Information and Control* 53 (1982): 32-51.
16. Daniel N. Osherson, Michael Stob and Scott Weinstein. 1986. *Systems that learn: an introduction to learning theory for cognitive and computer scientists.* Cambridge, Massachusetts.: MIT Press.

17. Hartley Rogers, Jr. 1987. *Theory of recursive functions and effective computability.* Cambridge, Massachusetts: MIT Press.
18. Thomas Zeugmann, Steffen Lange and Shyam Kapur. Characterizations of monotonic and dual monotonic language learning. *Information and Computation* 120(2) (1995): 155–173.