On Conservative Learning of R.e. Languages

Ziyuan Gao[1], Sanjay Jain*[2], and Frank Stephan†[3]

1 **Department of Mathematics,**
National University of Singapore, Singapore 119076, Republic of Singapore.
`ziyuan84@yahoo.com`

2 **Department of Computer Science,**
National University of Singapore, Singapore 117417, Republic of Singapore.
`sanjay@comp.nus.edu.sg`

3 **Department of Mathematics and Department of Computer Science,**
National University of Singapore, Singapore 119076, Republic of Singapore.
`fstephan@comp.nus.edu.sg`

Abstract

Partially conservative learning is a variant of partial learning whereby the learner, on a text for a target language L, outputs one index e with $L = W_e$ infinitely often and every further hypothesis d is output only finitely often and satisfies $L \not\subseteq W_d$. The present paper studies the learning strength of this notion, comparing it with other learnability criteria such as confident partial learning, explanatory learning, as well as behaviourally correct learning. It is further established that for classes comprising infinite sets, partially conservative learnability is in fact equivalent to explanatory learnability relative to the halting problem.

1998 ACM Subject Classification Dummy classification

Keywords and phrases Partially conservative learning, inductive inference, prudent learning, conservative learning

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

Gold [12] introduced a framework of learning in the limit to analyse linguistic structures and the learnability of these structures. The underlying goal of his original model, which has since been picked up and thoroughly expanded by various researchers [2, 5, 14, 18], was to formulate precisely the intuitive notion of human language ability. Learning, according to this model, is an inductive inference process in which the learner is presented piecewise with examples of a concept, every particular example eventually appearing in the presentation, while it outputs a sequence of guesses as each datum is revealed. If the learner's sequence of guesses converges to a single correct description of the underlying concept, then the inference is correct; on the other hand, if it does not converge or converges to an incorrect description, then the inference is incorrect. Gold [12, Theorem I.8] also showed that any class of formal languages over an alphabet Σ that contains all finite languages and at least one infinite language is not inferrable in the limit from a text presentation of positive data. This negative result was widely interpreted as evidence that Gold's original paradigm of learning in the limit from positive data is too weak to adequately model the process of human language acquisition.

* Work is supported in part by NUS grants C252-000-087-001 and R252-000-420-112.
† Work is supported in part by NUS grant R252-000-420-112.

Various alternative models of learning by inference from positive data have since been proposed [9, 18]. In particular, Osherson, Stob and Weinstein [18] generalised the notion of learning in the limit to *partial learning*. Similar to the former setting, a recursive learner receives piecewise information about the elements of an unknown recursively enumerable language. At each stage, the learner is required to output a conjecture based on a pre-assigned hypothesis space – usually taken to be a fixed acceptable numbering of all recursively enumerable sets – and is judged to have successfully inferred a target language if it outputs exactly one correct index of the language infinitely often and outputs any other conjecture only finitely often. As it stands, partial learning is even more general than behaviourally correct learning; in fact, a recursive learner can partially learn the class of all recursively enumerable sets [18, Exercise 7.5A]. Osherson, Stob and Weinstein [18] called an explanatory learner *confident* iff it converges on any text for any language to a hypothesis; this concept was brought over to partial learning [11] by defining that the learner must issue exactly one hypothesis infinitely often on any text for any language. This new hybrid of the two learnability notions turns out to be restrictive in the case of language learning [11]: the class of all cofinite sets is not confidently partially learnable. Other learning constraints considered in the inductive inference literature, but studied mainly in the context of learning in the limit, include *consistency* [7], *conservativeness* [2] and *prudence* [18].

The present paper continues the study of *conservativeness* and in particular transfers it from learning in the limit to more general criteria like behaviourally correct learning and partial learning. Angluin [2] noted that inference from positive rather than positive and negative data is often stymied by the problem of overgeneralisation, by which is meant that the learner conjectures a proper superset of the target language, so that it never witnesses a counterexample to its hypothesis from a presentation of positive data. She suggested a strategy termed *conservative learning* whereby a learner may avoid overgeneralisation, and gave sufficient conditions for an indexed family of nonempty recursive languages to be inferable by a conservative learner [2, Theorem 5]. Further research into the properties of conservatively learnable classes, particularly for indexed families of recursive languages as well as recursively enumerable languages, has been carried out in [8, 17]. This paper adapts a slight variation of the original conservative learner to fit the partial learning model: a recursive learner is said to *partially conservatively* learn a recursively enumerable language L from text if it outputs exactly one correct index for L on each text for L and does not conjecture any proper superset of L.

In the present work, it is shown that imposing conservativeness yields a close connection between the models of learning in the limit (in both the syntactic and semantic sense) and partial learning. To exemplify this point, Theorem 7 states that the set of all partially conservatively learnable classes of r.e. languages strictly subsumes that of all conservatively behaviourally correctly learnable classes of r.e. languages. We begin by proving that the class containing all finite sets and the set of natural numbers is not partially conservatively learnable, thereby showing that this new learning notion constitutes a greater restriction than ordinary partial learning. However, partial conservativeness may still be a fairly robust concept, as the class of graphs of all recursive functions is learnable according to this criterion (Example 9). Theorem 8 frames a learning criterion that appears to be more general than partially conservative learnability; nevertheless, one can show that this apparent generalisation is equivalent in terms of learning strength to the original definition of partial conservativeness. The paper also revisits vacillatory learning, introduced by Case [6], when it is combined with conservativeness. The principal result obtained here is that conservative vacillatory learning is as powerful as conservative explanatory learning. This result stands in contrast to that for the case when conservativeness is not stipulated: in [6], one finds an example of a vacillatorily learnable class which is not explanatorily learnable.

As further evidence of the tie-in between partially conservative learning and learning in the limit, several results draw comparisons between partially conservative learning and explanatory learning relative to oracles; Theorem 17, for instance, asserts that for any class of r.e. languages composed of infinite sets, the learning strengths of both notions coincide when the oracle used for the explanatory learner is Turing equivalent to the diagonal halting problem. In addition, this work proposes a characterisation of partially conservatively learnable classes of r.e. languages as an analogue of Angluin's Theorem 1 in [2]. The original theorem, formulated in the intuitively appealing notion of a family of finite "telltale" sets, gave necessary and sufficient conditions for an indexed family of recursive languages to be explanatorily learnable from positive data. De Jongh and Kanazawa [8] later generalised Angluin's telltale condition to characterise indexed families of r.e. languages. Similarly, Zeugmann, Lange, and Kapur [17] proved a characterisation of conservative learnability for indexed families of recursive languages; this was subsequently extended by de Jongh and Kanazawa [8] to the case of indexed families of r.e. languages. In Theorem 20, a characterisation similar to that of Angluin [2] and Zeugmann, Lange and Kapur [17] is given for partial conservative learning.

The final part of the paper is devoted to the study of *consistency* in partial learning. Although consistency is quite a stringent criterion, even when partial learning is permitted, Theorem 22 demonstrates that a weaker variant of consistent and conservative partial learnability does follow from partially conservative learnability. To complete the analysis of the relative learning strength of consistent partial learning in relation to other learning criteria studied in this paper, several separating class examples are given towards the end.

2 Notation

The notation and terminology from recursion theory adopted in this paper follows the book of Rogers [19]. Background on inductive inference can be found in [14]. $\mathbb{N}$ denotes the set of natural numbers. Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ denote a fixed acceptable numbering of all partial-recursive functions. Given a set S, $\overline{S}$ denotes the complement of S, and S^* denotes the set of all finite sequences in S. $D_0, D_1, D_2, \ldots$ is a canonical indexing of all finite sets. Let $W_0, W_1, W_2, \ldots$ be a universal numbering of all r.e. sets, where W_e is the domain of φ_e. $\langle x, y\rangle$ denotes Cantor's pairing function, given by $\langle x, y\rangle = \frac{1}{2}(x+y)(x+y+1)+y$. $\varphi_e(x)\uparrow$ means that $\varphi_e(x)$ remains undefined; $\varphi_{e,s}(x)\downarrow$ means that $\varphi_e(x)$ is defined, and that the computation of $\varphi_e(x)$ halts within s steps.

Turing reducibility is denoted by $\leq_T$; $A \leq_T B$ holds if A can be computed via a machine which 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. The class of all recursive functions is denoted by REC; the class of all $\{0,1\}$-valued recursive functions is denoted by $REC_{0,1}$. For any two partial-recursive functions f and g, $f =^* g$ denotes that for cofinitely many x, $f(x)\downarrow = g(x)\downarrow$. The symbol $\mathbb{K}$ denotes the diagonal halting problem $\{e : \varphi_e(e)\downarrow\}$. Furthermore, $\mathbb{K}_s$ is an approximation to $\mathbb{K}$; without loss of generality, $\mathbb{K}_s \subseteq \{0,1,\ldots,s\} \cap \mathbb{K}$ and the set $\{\langle x, s\rangle : x \in \mathbb{K}_s\}$ is primitive recursive.

The jump of a set A is denoted by A' and denotes the relativised halting problem $A' = \{e : \varphi_e^A(e)\downarrow\}$. An r.e. set S is *simple* if it is coinfinite and intersects every infinite r.e. set. For any two sets A and B, $A \oplus B = \{2x : x \in A\} \cup \{2y+1 : y \in B\}$. Analogously,

$A \oplus B \oplus C = \{3x : x \in A\} \cup \{3y+1 : y \in B\} \cup \{3z+2 : z \in C\}$.

For any $\sigma, \tau \in (\mathbb{N} \cup \{\#\})^*, \sigma \preceq \tau$ if and only if $\sigma = \tau$ or τ is an extension of σ, $\sigma \prec \tau$ if and only if σ is a proper prefix of τ, and $\sigma(n)$ denotes the element in the nth position of σ, starting from $n = 0$. $\sigma[n]$ denotes the sequence $\sigma(0) \circ \sigma(1) \circ \ldots \circ \sigma(n-1)$. Given a number a and some fixed $n \geq 1$, denote by a^n the finite sequence $a \ldots a$, where a occurs n times. a^0 denotes the empty string. $|\sigma|$ is the length of σ. 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 $\sigma\tau$.

3 Learnability

Let $\mathcal{C}$ be a class of r.e. languages. 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 $L = \text{content}(T_L)$; here $T_L[n]$ denotes the sequence $T_L(0) \circ T_L(1) \circ \ldots \circ T_L(n-1)$ and the content of a text T, denoted content(T), is the set of numbers in the range of T. Analogously, for a finite sequence σ, content(σ) is the set of numbers in the range of σ. The main learning criteria studied in this paper are *partial learning*, *explanatory learning* and *behaviourally correct learning*. 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.

▶ **Definition 1.** **i.** [18] 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. [12] 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. [5] 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 $j \geq n$.

In some cases, learners are equipped with oracles and then $Ex[A]$ denotes the criterion of explanatory learning with oracle A and so on. Furthermore, in some cases, the learner does not use the default hypothesis space $W_0, W_1, W_2, \ldots$ but instead uses a uniformly r.e. hypothesis space $H_0, H_1, H_2, \ldots$ where in the corresponding definitions W_e is replaced by H_e. The next series of definitions impose additional constraints on the learner.

▶ **Definition 2.** **i.** [11] M is said to *confidently partially* ($ConfPart$) learn $\mathcal{C}$ if it partially learns $\mathcal{C}$ from text and, on every infinite sequence, outputs exactly one index infinitely often.

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

iii. [2] M *conservatively explanatorily* ($ConsvEx$) learns $\mathcal{C}$ if it Ex learns $\mathcal{C}$ and, for any $\sigma, \tau \in (\mathbb{N} \cup \{\#\})^*$ such that $M(\sigma) \neq M(\sigma \circ \tau)$, there is a number x with $x \in \text{range}(\sigma \circ \tau) - W_\sigma$.

iv. M is said to *conservatively behaviourally correctly* ($ConsvBC$) learn $\mathcal{C}$ if it BC learns $\mathcal{C}$ and is semantically conservative. Here, a learner M is semantically conservative iff, for any $\sigma, \tau \in (\mathbb{N} \cup \{\#\})^*$ with $W_{M(\sigma)} \neq W_{M(\sigma \circ \tau)}$, there is a number x with $x \in \text{range}(\sigma \circ \tau) - W_{M(\sigma)}$.

v. [18] M is *prudent* if it learns the class $\{W_{M(\sigma)} : \sigma \in (\mathbb{N} \cup \{\#\})^*, M(\sigma) \neq ?\}$. In other words, M learns every set it conjectures.

vi. [6] M *vacillatorily* (Vac) learns $\mathcal{C}$ if it BC learns $\mathcal{C}$ from text, and outputs on each text for every L in $\mathcal{C}$ only finitely many different indices.

vii. M is said to *conservatively vacillatorily* ($ConsvVac$) learn $\mathcal{C}$ if it $ConsvBC$ learns $\mathcal{C}$ from text, and outputs on each text for every L in $\mathcal{C}$ only finitely many different indices.

viii. [7] M is *consistent* ($Cons$) if for all $\sigma \in (\mathbb{N} \cup \{\#\})^*$, $\mathrm{content}(\sigma) \subseteq W_{M(\sigma)}$.

ix. M is said to *class-comprisingly* ($ClsCom$) learn $\mathcal{C}$ if it learns $\mathcal{C}$ from text with respect to a hypothesis space $\{H_0, H_1, H_2, \ldots\}$. As M learns $\mathcal{C}$, it is immediate that $\mathcal{C} \subseteq \{H_0, H_1, H_2, \ldots\}$.

x. M is said to *class-preservingly* ($ClsPresv$) learn $\mathcal{C}$ if it learns $\mathcal{C}$ from text with respect to a hypothesis space $\{H_0, H_1, H_2, \ldots\}$ such that $\mathcal{C} = \{H_0, H_1, H_2, \ldots\}$.

Blum and Blum [7] introduced the notion of a *locking sequence* for explanatory learning, whose existence is a necessary criterion for a learner to successfully identify the r.e. language generating the text seen.

▶ **Lemma 3.** *[7] Let M be a recursive learner and L be an r.e. language explanatorily learnt by M. Then there is a finite sequence σ such that*
- $W_{M(\sigma)} = L$*;*
- *For all $\tau \in (L \cup \{\#\})^*$, $M(\sigma) = M(\sigma \circ \tau)$.*

This σ is called a locking sequence *for L.*

One may consider a weakened notion of a locking sequence, whereby the learner need not output an index of the target language L on this sequence; such a sequence is called a *stabilising sequence* for L.

▶ **Definition 4.** [14] Let M be a recursive learner and L be an r.e. language. Then σ is a *stabilising sequence* for M on L if
- $\mathrm{content}(\sigma) \subseteq L$;
- For all $\tau \in (L \cup \{\#\})^*$, $M(\sigma) = M(\sigma \circ \tau)$.

With a slight modification, one can adapt Lemma 3 to the partial learning model.

▶ **Lemma 5.** *Let M be a recursive learner and L be an r.e. language partially learnt by M. Then there is a finite sequence σ such that*
- $W_{M(\sigma)} = L$*;*
- *For all $\tau \in (L \cup \{\#\})^*$, there is an $\eta \in (L \cup \{\#\})^*$ such that $M(\sigma \circ \tau \circ \eta) = M(\sigma)$.*

4 Partially conservative learning

Conservativeness is a learnability constraint that has been studied fairly extensively in the inductive inference literature, especially in the setting of indexed families [2, 17]. In this paper, we consider the notion of *partial conservativeness* in language learning; in brief, this is partial learning combined with the constraint that if a learner outputs e infinitely often on a text for some target language L, then none of its other conjectures on this text can contain L as a subset. Gold [12] observed that the class $\{\mathbb{N}\} \cup \{F : F \text{ is finite}\}$ is not behaviourally correctly learnable, even when granting access to any oracle. Nonetheless, Gold's class is partially learnable from text: it is only necessary to output a canonical index for $\mathbb{N}$ as many times as a new datum appears, and to conjecture for every input σ a canonical index for the finite set $\mathrm{content}(\sigma)$. By contrast, one can show that Gold's class cannot be partially conservatively learnt.

▶ **Example 6.** The class $\mathcal{C} = \{\mathbb{N}\} \cup \{F : F \text{ is finite}\}$ is not partially conservatively learnable.

Proof. Assume by way of contradiction that M were a recursive partially conservative learner of $\mathcal{C}$. Since M learns $\mathbb{N}$, it follows from Lemma 5 there is a sequence $a_0 \circ a_1 \circ \ldots \circ a_n \in (\mathbb{N} \cup \{\#\})^*$ such that $M(a_0 \circ a_1 \circ \ldots \circ a_n) = e$ for some e with $\mathbb{N} = W_e$. Then $a_0 \circ a_1 \circ \ldots \circ a_n$ is the initial segment of a text for the finite set $\{a_0, a_1, \ldots, a_n\} - \{\#\}$, but since M outputs an index e with $\mathbb{N} = W_e \supset \{a_0, a_1, \ldots, a_n\} - \{\#\}$, M cannot be a partially conservative learner of $\mathcal{C}$. ◀

The main theorem of the present paper establishes a hierarchy of the major learnability notions treated herein. It yields a connection between the criteria of partial learning and learning in the limit - both syntactic and semantic - when one imposes the learning constraint of conservativeness.

▶ **Theorem 7.** *$ConsvEx = ConsvVac \subset ConsvBC \subset ConsvPart \subset ConsvEx[\mathbb{K}]$.*

Before proceeding with the proof of Theorem 7, we shall lay out a series of somewhat more general statements, from which Theorem 7 may be deduced. The first of these results shows that the criterion for partial conservativeness may in fact be slightly relaxed. It asserts that for any class $\mathcal{C}$ of r.e. languages to be *ConsvPart* learnable, it is sufficient that there is a recursive learner which, on every text for a target language L in $\mathcal{C}$, outputs at least one correct index and does not overgeneralise at any stage.

▶ **Theorem 8.** *Let $\mathcal{C}$ be a class of r.e. languages such that there is a recursive learner which, on any text for some $L \in \mathcal{C}$, outputs at least one correct index for L, and does not output an index for a proper superset of L. Then $\mathcal{C}$ is ConsvPart learnable.*

Proof. Given a recursive learner M satisfying the hypothesis of the theorem, one can make the following learner N. N outputs a canonical index for $\emptyset$ until N sees the first datum. Then N analyses the sequence $e_0, e_1, \ldots$ of M given as response to some intput text. N conjectures for each input σ a canonical index for $range(\sigma)$. Furthermore, for each σ with $range(\sigma) \neq \emptyset$ and each e_n, N outputs an index $f(e_n, \sigma)$ where $W_{f(e_n,\sigma)}$ is the union of $\emptyset$ and all $W_{e_n,s}$ where $W_{e_n,s} \subset W_{e_n,s+1}$ and $range(\sigma) \not\subseteq W_{e_m,s}$ for all $m < n$ and $W_{e_n,s} \supseteq range(\sigma)$ and it can be deduced that $W_{e_n,s} \not\subseteq W_{f(e_n,\tau)}$ for all proper prefixes τ of σ. Here "it can be deduced that $W_{e_n,s} \not\subseteq W_{f(e_n,\tau)}$" means there is some $m < n$ and some $t < s$ with $range(\tau) \subseteq W_{e_m,t}$ and $W_{e_n,t} \subset W_{e_n,s}$; it then follows that $W_{f(e_n,\tau)} \subseteq W_{e_n,t} \subset W_{e_n,s}$.

Clearly N learns the empty set. If M learns a finite non-empty set L then N outputs a canonical index for $range(\sigma)$. Furthermore, no index e_n output by M while learning L is a proper superset of L; hence the $W_{e_n,s}$ used in the union of a conjecture of the form $f(e_n, \sigma)$ satisfy that either $L \not\subseteq W_{e_n}$ or $W_{e_n,s} \subset W_{e_n,s+1} \subseteq L$; hence $L \not\subseteq W_{f(e_n,\sigma)}$. So the hypothesis of N for L is unique. If M learns an infinite set L then there is a first n such that $W_{e_n} = L$. For all sufficiently long prefixes σ of L, $range(\sigma) \not\subseteq W_{e_m}$ for all $m < n$. Let σ be the first prefix of the text with this property. Then there are infinitely many stages s such that $W_{e_n,s} \subset W_{e_n,s+1}$ $range(\sigma) \subseteq W_{e_n,s}$ and for $\tau \prec \sigma$ it can be deduced that $W_{e_n,s} \not\subseteq W_{f(e_n,\tau)}$. $W_{f(e_n,\sigma)}$ is the union of these $W_{e_n,s}$ and hence equal to W_{e_n} and equal to L. For prefixes η of the text properly extending σ, there is no s such that it can be deduced that $W_{e_n,s} \not\subseteq W_{f(e_n,\sigma)}$; thus $W_{e_n,\eta}$ is empty. Furthermore, for $k > n$, each $W_{f(e_k,\eta)}$ satisfies that $range(\eta) \subseteq W_{e_n}$ and hence $W_{f(e_k,\eta)}$ is a finite set different from L. Hence $f(e_n, \sigma)$ is the unique index of L output by N and all other indices are not supersets of L.

To finish the construction, one can build a new learner N' that first observes the conjectures $c_0, c_1, c_2, c_3, \ldots$ output by N on the given text T. For each conjecture c_i that N outputs, N' outputs c_i at least n times iff there is a stage $s > n$ such that $\forall x < n[x \in W_{c_i,s}(x) \Leftrightarrow x \in \text{content}(T[s])]$ holds. Consequently, N' outputs on T the unique index of L

issued by N infinitely often, and it outputs all other conjectures of N only finitely often. N' is therefore a *ConsvPart* learner of $\mathcal{C}$. ◀

Whilst partially conservative learnability may appear at first sight to be quite a stringent learning criterion, one can furnish a relatively natural example of a partially conservatively learnable class of r.e. languages which is neither behaviourally correctly nor confidently partially learnable.

▶ **Example 9.** Let $\mathcal{G} = \{graph(f) : f \text{ is recursive}\}$. Then $\mathcal{G}$ is *ConsvPart* learnable but neither confidently partially learnable nor behaviourally correctly learnable.

Proof. By Theorem 17 and the fact that $\mathcal{G}$ composes of infinite r.e. sets, it suffices to show that $\mathcal{G}$ is $Ex[\mathbb{K}]$ learnable. Adleman and Blum [1, Theorem 3] have shown that *REC* is $Ex[\mathbb{K}]$ learnable, and so $\mathcal{G}$ is indeed $Ex[\mathbb{K}]$ learnable as well.

Assume by way of contradiction that $\mathcal{G}$ were confidently partially learnable via a recursive learner M. By the confidence of M and the corresponding variant of Lemma 5, one may find a finite sequence $\alpha = \langle 0, y_0\rangle \circ \langle 1, y_1\rangle \circ \ldots \circ \langle n, y_n\rangle$ such that, for some unique index e, $M(\alpha) = e$, and for each $\sigma \in (\mathbb{N} \cup \{\#\})^*$ of the form $\sigma = \langle n+1, z_{n+1}\rangle \circ \ldots \circ \langle n+k, z_{n+k}\rangle$, there is a sequence $\tau \in (\mathbb{N} \cup \{\#\})^*$ of the form
$\tau = \langle n+k+1, z_{n+k+1}\rangle \circ \ldots \circ \langle n+k+i, z_{n+k+i}\rangle$ with $M(\alpha \circ \sigma \circ \tau) = e$. A new recursive function g may now be defined inductively as follows.

- Set $g(i) = y_i$ for all $i \leq n$.
- Assume that $k \geq n$ and $g(x)$ has been defined for all $x \leq k$. Run a search for a sequence of the form $\langle k+1, z_{k+1}\rangle \circ \ldots \circ \langle k+l, z_{k+l}\rangle$ such that $M(\langle 0, g(0)\rangle \circ \langle 1, g(1)\rangle \circ \ldots \circ g(k) \circ \langle k+1, z_{k+1}\rangle \circ \ldots \circ \langle k+l, z_{k+l}\rangle) = e$; this search must terminate due to the definition of e as mentioned above. Set $g(k+j) = z_{k+j}$ for $j = 1, \ldots, l$, and $g(k+l+1) = \varphi_{e'}(k+l+1)+1$ if W_e is the graph of a recursive function $\varphi_{e'}$; otherwise, $g(k+l+1)$ remains undefined until the next stage.

If W_e is not the graph of a recursive function, then $W_e \neq \{\langle x, y\rangle : x \in \mathbb{N} \wedge g(x) \downarrow = y\}$; M, however, outputs e infinitely often on the text $\langle 0, g(0)\rangle \circ \langle 1, g(1)\rangle \circ \langle 2, g(2)\rangle \circ \ldots$, and so it cannot confidently partially learn the graph of g. In the case that W_e were the graph of some recursive function $\varphi_{e'}$, then, since g is defined to be such that $\langle k, g(k)\rangle \neq \langle k, \varphi_{e'}(k)\rangle$ for infinitely many k, $W_e \neq \{\langle x, y\rangle : x \in \mathbb{N} \wedge g(x) \downarrow = y\}$ still holds, and thus M fails to confidently partially learn the graph of g. This contradiction establishes that $\mathcal{G}$ is not confidently partially learnable. Furthermore, it has already been established [21, Corollary 20] that the class of all graphs of recursive functions is not *BC* learnable. ◀

As an immediate consequence of Theorem 8 and Example 9, one has that the notion of partially conservative learning strictly subsumes that of conservative behaviourally correct learning, as claimed in Theorem 7. The subsequent theorem places the last inclusion relation of Theorem 7 in a broader setting, characterising the oracles relative to which a partially conservatively learnable class is also explanatorily learnable.

▶ **Theorem 10.** *Every ConsvPart learnable class is ConsvEx[A] learnable if and only if* $\mathbb{K} \leq_T A$.

Proof. Assume that $\mathbb{K} \leq_T A$, and that M is a *ConsvPart* learner of $\mathcal{C}$. On input σ, a new A-recursive learner N may effectively search via oracle A for the shortest $\tau \preceq \sigma$ such that $\text{content}(\sigma) \subseteq W_{M(\tau)}$, and output $M(\tau)$; if no such τ exists, it outputs a canonical index for $\emptyset$. As N processes a text T for some $L \in \mathcal{C}$, $N^A(T[s+1]) \neq N^A(T[s])$ only if $T(s) \notin W_{N^A(T[s])}$, and so it is conservative. Furthermore, if $T[s]$ is the shortest text segment such that $W_{M(T[s])} = L$, it follows from the partial conservativeness of M that for every $k < s$, there is a number $x_k \in L - W_{M(T[k])}$. Hence there is a sufficiently long text segment $T[l]$ with $l > s$ for which $\text{content}(T[u]) \not\subseteq W_{M(T[k])}$ and $\text{content}(T[u]) \subseteq W_{M(T[s])}$ whenever $u \geq l$ and $k < s$, which establishes that N^A *ConsvEx*$[A]$ learns $\mathcal{C}$.

Conversely, one may show that the class $\{\mathbb{K} \cup D : D \text{ is finite}\}$, while *ConsvPart* learnable, is *Ex*$[A]$ learnable iff $\mathbb{K} \leq_T A$. To verify this, one may first construct an *Ex*$[\mathbb{K}]$ learner $M^{\mathbb{K}}$ of the given class: on input σ, $M^{\mathbb{K}}$ outputs a canonical index for $\mathbb{K} \cup \{x \in \text{content}(\sigma) : x \notin \mathbb{K}\}$. An application of Theorem 17 then gives that this class, consisting of infinite sets, is *ConsvPart* learnable. Next, consider a locking sequence (Lemma 3) $\sigma \in (\mathbb{K} \cup \{\#\})^*$ for $\mathbb{K}$ corresponding to an *Ex*$[A]$ learner M^A of $\{\mathbb{K} \cup D : D \text{ is finite}\}$. Then it holds that $x \notin \mathbb{K} \Leftrightarrow \exists \tau \in (\mathbb{K} \cup \{x\} \cup \{\#\})^*[M^A(\sigma \circ \tau) \neq M^A(\sigma)]$, and therefore $\mathbb{K} \leq_T A$. Furthermore, suppose that $\mathbb{K} \leq_T A$; then, since it was shown above that the given class is *Ex*$[\mathbb{K}]$ learnable, this class must also be *Ex*$[A]$ learnable. ◀

To complement the preceding theorem, one can show that for every nonrecursive set A, *ConsvEx*$[A]$ learning is more powerful than *ConsvPart* learning.

▶ **Example 11.** If A is not recursive, then there is a class which is *ConsvEx*$[A]$ learnable but not partially conservatively learnable.

Proof. Let $\mathcal{C}$ contain the following sets:

1. All sets of the form $D \oplus E$ where D, E is finite and $|D| \neq 1$;
2. All sets $\{e\} \oplus E$ where E is finite and $e \notin A \oplus \overline{A}$;
3. All sets $\{e\} \oplus \mathbb{N}$ where $e \in A \oplus \overline{A}$.

The following *ConsvEx*$[A]$-learner infers $\mathcal{C}$: If $e \in A \oplus \overline{A}$ and $range(\sigma) = \{e\} \oplus E$ for a finite set E then M conjectures $\{e\} \oplus \mathbb{N}$ else M conjectures $range(\sigma)$. In the case that $e \in A \oplus \overline{A}$ and the text is for $\{e\} \oplus \mathbb{N}$ then the learner will eventually converge to an index for this set; if the text is for a finite set different from all $\{e\} \oplus E$ with $e \in A \oplus \overline{A}$ then the learner will converge to an index for this finite set. Hence the class is *ConsvEx*$[A]$-learnable.

Assume now by way of contradiction that a partial conservative learner infers this class. Then, for each e, this learner overgeneralises on some input with range $\{e\} \oplus E$ iff $e \in A \oplus \overline{A}$: in the case that e is in this set, the learner must overgeneralise eventually as it has to output the set $\{e\} \oplus \mathbb{N}$, in the case that e is not in this set, the learner has to learn $\{e\} \oplus E$ and therefore cannot conjecture any proper superset of this language. Hence $A \oplus \overline{A}$ is recursively enumerable and thus recursive, a contradiction to the assumption on A. ◀

Returning to the first equality of Theorem 7, the concept of *vacillatory* learning, introduced by Case in [6], will be considered conjointly with conservativeness. Vacillatory learning, when imposed together with conservativeness, is a fairly stringent criterion; the next theorem asserts that it implies conservative explanatory learning in general.

▶ **Theorem 12.** *If a class of r.e. languages is ConsvVac learnable, then it is ConsvEx learnable.*

Proof. Assume that the class $\mathcal{C}$ of r.e. languages is *ConsvVac* learnable via a recursive learner M. We first build an *Ex* learner N_1 of $\mathcal{C}$ which is semantically conservative, in the sense that for any two finite sequences σ, τ with $W_{N_1(\sigma)} \neq W_{N_1(\sigma \circ \tau)}, \text{content}(\sigma \circ \tau) \not\subseteq W_{N_1(\sigma)}$. On input σ, simulate M and observe the conjectures $e_0, e_1, \ldots, e_{|\sigma|-1}$ output by M on piecewise increasing segments of σ. N_1 then outputs e_l iff l is the maximum number such that $e_l \neq e_i$ for all $i < l$, that is, e_l is the latest conjecture of M which differs from all its prior ones.

For the verification, suppose that N_1 processes a text for some $L \in \mathcal{C}$. Owing to the fact that M Vac learns L, M outputs only finitely many different hypotheses $e_0, e_1, \ldots, e_l$ on this text, where it is assumed that $i < j$ iff M conjectures e_i prior to e_j; moreover, at least one of these conjectures is correct. Then e_l is the last hypothesis of M which is different from all its previous ones, and so N_1 outputs e_l in the limit. If $W_{e_l} \neq L$, then there is some $k < l$ such that $L = W_{e_k}$. But if σ_{e_l} is the text segment on which M outputs e_l, the conservativeness of M implies that content$(\sigma_{e_l}) \not\subseteq W_{e_k}$, contradicting the fact that $L = W_{e_k}$. The learner N_1 is also semantically conservative: if it outputs e_a and e_b on the text segments σ_{e_a} and σ_{e_b} respectively with $\sigma_{e_a} \preceq \sigma_{e_b}$, and $W_{e_a} \neq W_{e_b}$, then M conjectures e_b on some text segment η such that $\sigma_{e_a} \preceq \eta \preceq \sigma_{e_b}$. By the conservativeness of M, content$(\eta) \not\subseteq W_{e_a}$. This verifies the required properties of N_1.

Based on the learner N_1, one may construct a $ConsvEx$ learner N of $\mathcal{C}$ as follows. First, for any $\sigma \in (\mathbb{N} \cup \{\#\})^*$, let f be a recursive function for which $W_{f(\sigma)} = \bigcup_{s \in A_\sigma} W_{N_1(\sigma),s}$, where $A_\sigma = \{t : \forall \tau \in (W_{N_1(\sigma),t} \cup \{\#\})^* [|\tau| > t \vee M(\sigma) = M(\sigma \circ \tau)]\}$. N is defined inductively on a text $T(0) \circ T(1) \circ T(2) \circ \ldots$ according to the following algorithm:

- Stage 0. Set $\sigma_0 = T(0)$ and $N(T(0)) = f(\sigma_0)$.
- Stage $s+1$; N reads the current input $T[s+2]$. Assume that $f(\sigma_s)$ is the conjecture of N at stage s. For any finite sequence σ, denote by $W_{f(\sigma),s}$ the set $\bigcup_{t \in (A_\sigma \cap \{0,1,\ldots,s\})} W_{N_1(\sigma),t}$, and let x_{s+1} be the least number in the range of $T[s+2]$ which is not contained in content(σ_s); if no such number exists, let $x_{s+1} = \#$. N searches for the shortest sequence γ with $|\gamma| \leq s+1$ and $\gamma \in (\text{range}(T[s+2]) \cup \{\#\})^*$ such that $N_1(\sigma_s) \neq N_1(\sigma_s \circ x_{s+1} \circ \gamma)$ and content$(\sigma_s \circ x_{s+1} \circ \gamma) \not\subseteq W_{f(\sigma_s),|\gamma|+1}$; if γ is found, N outputs $f(\sigma_s \circ x_{s+1} \circ \gamma)$, and one defines $\sigma_{s+1} = \sigma_s \circ x_{s+1} \circ \gamma$. If no such γ exists, N outputs $f(\sigma_s)$, and one defines $\sigma_{s+1} = \sigma_s$.

The syntactic conservativeness of N follows directly from construction: retaining the notation in the algorithm described above, it suffices to note that whenever N makes a mind change, then it has found a sequence γ with content(γ) contained in the current range of the text such that content$(\gamma) \not\subseteq W_{f(\sigma_s),|\gamma|+1}$, which implies, by the definition of f and the property that $x_{s+1} \circ \gamma$ enforces a mind change of N_1, that content$(\gamma) \not\subseteq W_{f(\sigma_s)}$. It remains to show that N is an Ex learner of $\mathcal{C}$. For a contradiction, assume first that N makes infinitely many mind changes on an input text T_L for some $L \in \mathcal{C}$. It follows that there are infinitely many stages s such that $\sigma_s \prec \sigma_{s+1}$ and $N_1(\sigma_s) \neq N_1(\sigma_{s+1})$; furthermore, $\bigcup_s \text{content}(\sigma_s) = L$. Hence $lim_s \sigma_s$ is a text for L on which N_1 makes infinitely many mind changes, contradicting the fact that it Ex learns L. Secondly, suppose that on T_L, N outputs $f(\sigma_s)$ in the limit. Fix a text extension T'_L of σ_s for L. If σ_s is a locking sequence for N_1 on L, then $W_{f(\sigma_s)} = W_{N_1(\sigma_s)} = L$, so that N explanatorily learns L. If σ_s is not a locking sequence for N_1 on L, then there is a text prefix $T'_L[l]$ extending σ_s such that $W_{N_1(T'_L[l])} = L$ and $N_1(T'_L[l]) = N_1(T'_L[k])$ for all $k \geq l$. Since N converges to the index $f(\sigma_s)$, it follows from construction that whenever $\sigma_s \preceq T'_L[p] \preceq T'_L[l]$ and $N_1(T'_L[p]) \neq N_1(\sigma_s)$, then content$(T'_L[p]) \subseteq W_{f(\sigma_s)} \subseteq W_{N_1(\sigma_s)}$; thus, by the semantic conservativeness of N_1, $W_{N_1(\sigma_s)} = W_{N_1(T'_L[p])}$, whence, $W_{N_1(\sigma_s)} = W_{N_1(T'_L[l])} = L$. Therefore, whenever $N_1(\sigma_s) \neq N_1(\sigma_s \circ \tau)$ for some $\tau \in (W_{N_1(\sigma_s)} \cup \{\#\})^*$, one has content$(\tau) \subseteq W_{f(\sigma_s)}$, giving that $W_{f(\sigma_s)} = W_{N_1(\sigma_s)} = L$, as required. ◀

The proof of the preceding theorem hinges on the construction of a specific text for some language in the class to be learnt, on which an explanatory learner may output indices of languages outside this class. In the next example, it is shown that if one imposes the condition that the learner must use a class-preserving hypothesis space, then semantic conservativeness does not in general imply its syntactic analogue for explanatory learning.

▶ **Example 13.** If A is a nonrecursive r.e. set, then the class $\mathcal{C} = \{A \cup \{x\} : x \notin A\}$ is semantically conservatively and explanatorily learnable using a class-preserving hypothesis space, as well as syntactically conservatively and prudently explanatorily learnable using a class-comprising hypothesis space, but not *ConsvEx* learnable using a class-preserving hypothesis space.

Proof. Given a non-recursive r.e. set A, let $a_0, a_1, a_2, \ldots$ be a recursive one-one enumeration of A, and let A_s denote the sth approximation $\{a_0, a_1, \ldots, a_s\}$. One may construct a learner M which executes the following instructions. First, it fixes any $x \notin A$. On input σ, if $x \in \text{range}(\sigma)$ or if $\text{content}(\sigma) \subseteq A_{|\sigma|}$, then M conjectures a canonical index for $A \cup \{x\}$; if $x \notin \text{range}(\sigma)$ and $\text{content}(\sigma) \not\subseteq A_{|\sigma|}$, then it searches for the least $y \in \text{range}(\sigma)$ with $y \notin A_{|\sigma|}$, and conjectures the index e such that

$$W_e = \begin{cases} A \cup \{x\} & \text{if } y \in A; \\ A \cup \{y\} & \text{if } y \notin A. \end{cases}$$

The class-preserving property of M is immediate from construction. Furthermore, if M makes a semantic mind change between two stages s and s' with $s < s'$ on some text T, then it must hold that $W_{M(T[s+1])} = A \cup \{x\}$ and $W_{M(T[s'+1])} = A \cup \{y\}$ for some $y \notin A$ with $y \neq x$ and $y \in \text{content}(T[s'+1])$; hence M is also semantically conservative. In addition, on any text for some $L \in \mathcal{C}$, M eventually identifies the unique $z \in L - A$, and therefore explanatorily learns L.

To produce a syntactically conservative *PrudEx* learner P of $\mathcal{C}$ that uses a class-comprising uniformly r.e. hypothesis space, one may set P to work as follows. For any $\sigma \in (\mathbb{N} \cup \{\#\})^*$, denote by σ' the sequence $\sigma[|\sigma|-1]$ if $|\sigma| > 1$, and the empty sequence λ otherwise. On input σ, if $\text{content}(\sigma) \subseteq A_{|\sigma|}$ and $\sigma(|\sigma|-1) \notin \text{content}(\sigma') \cup \{\#\}$, P conjectures a canonical index for $\{a_n : \exists x \in \text{content}(\sigma)[x \geq a_n]\}$. If $\text{content}(\sigma) \subseteq A_{|\sigma|}$ and $\sigma(|\sigma|-1) \in \text{content}(\sigma') \cup \{\#\}$, P repeats its previous hypothesis. If y is the least number such that $y \in \text{content}(\sigma) - A_{|\sigma|}$, then P outputs the index e satisfying $W_e = \{y\} \cup \{a_n : y \notin \{a_m : m < n\}\}$. P uses the uniformly r.e. hypothesis space $\mathcal{L} = \{L_0, L_1, L_2, \ldots\}$, where $L_x = \{x\} \cup \{a_n : x \notin \{a_m : a_m < a_n\}\}$, and $\mathcal{C} \subset \mathcal{L}$; it only remains to show that it syntactically conservatively *Ex* learns $\mathcal{L}$. On any text T, if there are stages s, s' with $s < s'$ such that $P(T[s+1]) \neq P(T[s'+1])$, then exactly one of the following cases must hold: (i) $W_{P(T[s+1])} = \{a_0, a_1, \ldots, a_m\}$ and $W_{P(T[s'+1])} = \{a_0, a_1, \ldots, a_{m'}\}$ for some m, m' such that $m < m'$, and $a_{m'} \in \text{range}(T[s'+1]) - W_{P(T[s+1])}$; or (ii) $W_{P(T[s+1])} = \{a_0, a_1, \ldots, a_m\}$ and $W_{P(T[s'+1])} = \{y\} \cup A$ for some m and $y \in \text{range}(T[s'+1]) - A$. In either case, P is always syntactically conservative; moreover, if T is a text for some $L_x \in \mathcal{L}$, then it must eventually converge to an index for L_x.

Assume by way of contradiction that $\mathcal{C}$ were class-preservingly *ConsvEx* learnable via some recursive learner N. Fix any $\tau \in A^*$ such that $N(\tau) \in \mathbb{N}$. Such a τ must exist; otherwise, since N learns $\mathcal{C}$, it must hold that for given any x, $x \notin A$ iff there is some $\eta \in A^*$ with $N(x \circ \eta) \in \mathbb{N}$, contradicting the nonrecursiveness of A. By the class-preservingness of N, $W_{N(\tau)} = A \cup \{b\}$ for some $b \notin A$. It follows from the syntactic conservativeness of N and the fact that it explanatorily learns any language $A \cup \{z\}$ with $z \notin A$ that for all $z \neq b$, $z \notin A \Leftrightarrow \exists \sigma \in (A \cup \{\#\})^*[N(\tau) \neq N(\tau \circ z \circ \sigma)]$ holds, that is, $z \notin A \cup \{b\} \Leftrightarrow z \neq b \wedge \exists \sigma \in (A \cup \{\#\})^*[N(\tau) \neq N(\tau \circ z \circ \sigma)]$, giving that A is recursive, a contradiction. ◀

Fulk [10] proved that every explanatorily learnable class is also prudently explanatorily learnable; Jain, Stephan, and Ye [15] established the corresponding result for behaviourally correct learning. In connection to these results, one may ask whether the relation in Theorem

12 still holds when the learner is required to be prudent. The following theorem answers this question affirmatively, showing that every conservatively explanatorily learnable class of r.e. languages is also conservatively explanatorily learnable by a prudent learner.

▶ **Theorem 14.** *If a class of r.e. languages is ConsvEx learnable, then it is PrudConsvEx learnable.*

Proof. Assume some recursive 1–1 listing of all the finite sequences, and consider ordering among the sequences based on this listing.

Let $I_s = \{x : x \leq s\}$.

Let $r(\sigma) = \max \bigcup_{\tau \leq \sigma} \text{content}(\tau)$.

Let M be a conservative learner for a class $\mathcal{L}$. Assume without loss of generality that if $\text{content}(\sigma) = \emptyset$, then σ is not a stabilizing sequence for M on any language except maybe $\emptyset$.

By s-m-n theorem, there exists a recursive function $P(\sigma, S)$ such that $W_{P(\sigma,S)}$ is defined as follows:
$W_{P(\sigma,S)} = W_{M(\sigma)} - [I_{r(\sigma)} - S]$, if

(1) $\text{content}(\sigma) \subseteq S \subseteq W_{M(\sigma)} \cap I_{r(\sigma)}$, and
(2) for all $\tau < \sigma$, at least one of following holds:

  (a) $\text{content}(\tau) \not\subseteq S$
  (b) For some τ' with $\text{content}(\tau') \subseteq W_{M(\sigma)} - [I_{r(\sigma)} - S]$, $M(\tau\tau') \neq M(\tau)$.

If any of the above conditions fails, then let $W_{P(\sigma,S)} = \emptyset$.

▶ **Claim 15.** *For all σ and S such that content$(\sigma) \subseteq S \subseteq I_{r(\sigma)}$, if $W_{P(\sigma,S)} \neq \emptyset$, then the following properties hold.*
(A) $W_{P(\sigma,S)} = W_{M(\sigma)} - [I_{r(\sigma)} - S]$,
(B) $W_{P(\sigma,S)} \cap I_{r(\sigma)} = S$.
(C) σ is the least stabilizing sequence for M on $W_{P(\sigma,S)}$.

(A) is trivial by definition of $W_{P(\sigma,S)}$. (B) holds by condition (1), and the fact that if $W_{P(\sigma,S)} \neq \emptyset$, then $W_{P(\sigma,S)} = W_{M(\sigma)} - [I_{r(\sigma)} - S]$. (C) holds as by conservativeness of M, and content(σ) being contained in $W_{M(\sigma)}$, σ is a stabilizing sequence for M on $W_{P(\sigma,S)}$. Furthermore, by (2), no smaller τ is a stabilizing sequence for M on $W_{P(\sigma,S)}$.

Note that for every $L \in \mathcal{L} - \{\emptyset\}$, for the least stabilizing sequence σ and $S = L \cap I_{r(\sigma)}$, it holds that $P(\sigma, L \cap I_{r(\sigma)})$ is an index for L. Thus, every language in $\mathcal{L} - \{\emptyset\}$ has an index of the form $P(\sigma, S)$, with content$(\sigma) \subseteq S \subseteq I_{r(\sigma)}$.

Now define M' as follows. On input $T[n]$, if content$(T[n]) = \emptyset$, then M' outputs a standard grammar for $\emptyset$. Otherwise, it searches for a σ such that, for $S = \text{content}(T[n]) \cap I_{r(\sigma)}$

(a) content$(\sigma) \subseteq S$, $|\sigma| \leq n$
(b) $W_{P(\sigma,S)} \neq \emptyset$,
(c) for all $\tau < \sigma$, either content$(\tau) \not\subseteq S$ or for some τ' with $|\tau'| \leq n$, content$(\tau') \subseteq$ content$(T[n])$, $M(\tau\tau') \neq M(\tau)$,
(d) for all τ' with $|\tau'| \leq n$, content$(\tau') \subseteq$ content$(T[n])$, $M(\sigma\tau') = M(\sigma)$

If such a least σ is found, then M' outputs $P(\sigma, S)$. Otherwise, M' repeats its previous hypothesis.

Note that every non-empty $W_{P(\sigma,S)}$ is learnt by M', as once the input sequence contains all elements in $I_{r(\sigma)}$ along with witnesses content(τ') satisfying $M(\tau\tau') \neq M(\tau)$, for each $\tau < \sigma$ such that content$(\tau) \subseteq S$ (such a witness τ' exists by (2) in the definition of $W_{P(\sigma,S)}$) and is large enough, we have that M' chooses $P(\sigma, S)$ as its hypothesis. Thus, M' is prudent. Thus, it suffices to show that M' is conservative. For this suppose M' outputs $P(\sigma, S)$ on

$T[n]$ and then later outputs a different $P(\sigma', S')$ on input $T[n']$ where $n' > n$. Note that content$(\sigma') \subseteq S' \subseteq$ content$(T[n'])$.

Case 1: $\sigma' < \sigma$: Then, clearly, content$(\sigma') \not\subseteq S$, as otherwise, by definition of M', there exists a τ' such that $M(\sigma'\tau') \neq M(\sigma')$, with $|\tau'| \leq n$ and content$(\tau') \subseteq$ content$(T[n])$ (by condition (c) on input $T[n]$). But then condition (d) on input $T[n']$ will prevent M' from outputting $P(\sigma', S')$. Thus, content$(\sigma') \not\subseteq W_{P(\sigma,S)}$ by Claim 15(B), and thus the mind change is conservative.

Case 2: $\sigma' > \sigma$: Then, by (c) in the definition of M', we have that for some τ', $M(\sigma\tau') \neq M(\sigma)$, where content$(\tau') \subseteq$ content$(T[n'])$. But then content$(\tau') \not\subseteq W_{M(\sigma)} \supseteq W_{P(\sigma,S)}$, as M is conservative. Thus, the mind change is conservative again. ◀

The following example emphasises the distinction between prudence and the use of a class-preserving hypothesis space; even when combined with conservativeness, prudence is not sufficient to guarantee that a uniformly r.e. class of languages is partially conservatively learnable with respect to a class-preserving hypothesis space.

► **Example 16.** Let $\mathcal{L} = \{L_{\langle d,2s\rangle} : d, s \in \mathbb{N}\} \cup \{L_{\langle d,2s+1\rangle} : d, s \in \mathbb{N}\}$, where

$$L_{\langle d,2s\rangle} = \begin{cases} \{d\} & \text{if } W_{d,s} = W_d; \\ \{d, t+1\} & \text{if } t \text{ is the first stage with } W_{d,s} \subset W_{d,t}, \end{cases}$$

$$L_{\langle d,2s+1\rangle} = \{d, d+s+1\}.$$

$\mathcal{L}$ is *PrudConsvEx* learnable but not *ClsPresvConsvPart* learnable.

Proof. If W_d is infinite then all $L_{\langle d,e\rangle}$ are different from $\{d\}$; if W_d is finite and $W_d = W_{d,s}$ then $L_{\langle d,e\rangle} = \{d\}$ for even $e \geq 2s$.

A prudent conservative explanatory learner is given by a learner for all finite sets which always conjectures the set of data items observed so far.

Furthermore, assume now that there is a learner which is conservative and uses a class preserving hypothesis space $H_0, H_1, H_2, \ldots$ of uniformly r.e. sets. One can check with oracle $\mathbb{K}$ whether the learner outputs on text $d \circ d \circ d \circ d \circ \ldots$ any hypothesis e with $d \in H_e$. Furthermore, with oracle $\mathbb{K}$ one can check whether $H_e = \{d\}$. If e exists and $H_e = \{d\}$ for the first such e found then W_d is finite as the learner uses a class-preserving hypothesis space else the learner does not conservatively learn $\{d\}$ by either overgeneralising or never outputting a hypothesis containing d on the text $d \circ d \circ d \circ \ldots$, hence W_d must be infinite. This would result in a decision procedure for $\{d : W_d \text{ is finite}\}$ relative to $\mathbb{K}$ which does not exist. Hence the given class is neither *ClsPresvConsvBC* learnable nor *ClsPresvConsvPart* learnable. ◀

► Remark. One also has that the conservatively vacillatorily learnable classes of r.e. languages constitute a strict subset of the prudently conservatively behaviourally correctly learnable classes when a class-preserving hypothesis space is enforced. For example, the class $\{\mathbb{K} \cup D : D \text{ is finite}\}$ is prudently conservatively behaviourally correctly learnable using a class-preserving hypothesis space, but it is not vacillatorily learnable.

Proof of Theorem 7. As was shown in Theorem 12, every *ConsvVac* learnable class of r.e. languages is *ConsvEx* learnable. A *ConsvEx* learnable class of r.e. languages is, a fortiori, also *ConsvBC* learnable. On the other hand, the class $\{\mathbb{K} \cup D : D \text{ is finite}\}$ is *ConsvBC* learnable but not *Ex* learnable, and this establishes the first strict inclusion.

A *ConsvBC* learner satisfies the learning criterion in the hypothesis of Theorem 8, and therefore every *ConsvBC* learnable class is *ConsvPart* learnable. Example 9 shows that *ConsvPart* learning is in fact less restrictive than *BC* learning.

The last inclusion $ConsvPart \subseteq ConsvEx[\mathbb{K}]$ follows as a particular case of Theorem 10. Moreover, Example 11 shows that the oracle $\mathbb{K}$ permits some classes which are not $ConsvPart$ learnable to be $ConsvEx$ learnt. ◀

Coming to a more particular setting, the following result demonstrates the equivalence of partially conservative learnability and $Ex[\mathbb{K}]$ learnability for classes comprising infinite sets. The hypothesis that all the languages in the class be infinite cannot, however, be dropped, as may be seen from Example 11.

▶ **Theorem 17.** *Let $\mathcal{C}$ be a class of infinite r.e. sets. Then $\mathcal{C}$ is $ConsvPart$ learnable if and only if it is $Ex[\mathbb{K}]$ learnable.*

Proof. Suppose that $\mathcal{C}$ is $Ex[\mathbb{K}]$ learnable via a $\mathbb{K}$-recursive learner M that outputs r.e. indices. On a text $T(0) \circ T(1) \circ T(2) \circ T(3) \circ \ldots$, let N be a learner that works by outputting, for each $\sigma \in (\text{range}(T) \cup \{\#\})^*$ and all s, the index $f(\sigma, s)$, where f is a recursive function such that $W_{f(\sigma,s)} = \{x : \exists t > x \forall \tau \in (W_{M^{\mathbb{K}_s}(\sigma),t} \cup \{\#\})^* \exists u > t[[|\tau| \leq x \Rightarrow M^{\mathbb{K}_u}(\sigma \circ \tau) = M^{\mathbb{K}_s}(\sigma)] \wedge [\text{content}(\sigma \circ x) \subseteq W_{M^{\mathbb{K}_s}(\sigma),t}]]\}$.

It shall be argued that N, on any text for a given $L \in \mathcal{C}$, outputs at least one correct index for L, and does not conjecture any proper superset of L; it will then follow as a consequence of Theorem 8 that $\mathcal{C}$ is $ConsvPart$ learnable.

First, if σ is not a locking sequence of M for L, then for all s, $W_{f(\sigma,s)}$ cannot be a superset of L: for, if one assumes that $L \subseteq W_{M^{\mathbb{K}_s}(\sigma)}$, then whenever t is sufficiently large, there is some $\tau \in (W_{M^{\mathbb{K}_s}(\sigma),t} \cup \{\#\})^*$ and u' such that whenever $u > u'$, $M^{\mathbb{K}_s}(\sigma) \neq M^{\mathbb{K}_u}(\sigma \circ \tau)$. Thus if $W_{M^{\mathbb{K}_s}(\sigma)}$ contains L, $W_{f(\sigma,s)}$ does not enumerate any element greater than $|\tau| + u'$, thus $W_{f(\sigma,s)}$ is finite. On the other hand, by Lemma 3, there must be at least one locking sequence $\sigma \in (\text{range}(T) \cup \{\#\})^*$; if σ is a locking sequence of M for L, then whenever s is sufficiently large, $M^{\mathbb{K}_s}(\sigma)$ is the same index for L. So, for every $x \in L$, there is a $t > x$ such that whenever $\tau \in (W_{M^{\mathbb{K}_s}(\sigma),t} \cup \{\#\})^*$ and $|\tau| \leq x$, one can find a $u > t$ satisfying $M^{\mathbb{K}_u}(\sigma \circ \tau) = M^{\mathbb{K}_s}(\sigma)$ and $\text{content}(\sigma \circ x) \subseteq W_{M^{\mathbb{K}_s}(\sigma),t}$. Hence for such a σ and all large enough s, $W_{f(\sigma,s)} = W_{M^{\mathbb{K}_s}(\sigma)} = L$. Therefore N has the required properties stated in the preceding paragraph.

The converse direction of the proof has been demonstrated, in a more general form, in Theorem 10. ◀

The subsequent example gives an instance of a set A such that $A \not\leq_T \mathbb{K}$ and the relation $Ex[A] \subset ConsvPart$ no longer holds, even when confined to classes of infinite sets.

▶ **Example 18.** The class of infinite sets

$$\mathcal{C} = \{\{e\} \oplus (W_e \cup D) : D \text{ is finite and } W_e \text{ is cofinite}\} \cup \{\{e\} \oplus \mathbb{N} : e \in \mathbb{N}\}$$

is $Ex[\mathbb{K}']$ learnable but not partially conservatively learnable.

Proof. An $Ex[\mathbb{K}']$ learner M may be programmed as follows: on input σ, where $\text{content}(\sigma) = \{e\} \oplus D$ for some finite set D, M checks whether or not $\forall x > |\sigma| \exists s[x \in W_{e,s}]$ holds; if so, then M conjectures $\{e\} \oplus (W_e \cup D)$ else M conjectures $\{e\} \oplus \mathbb{N}$. Note that formally this learner is behaviourally correct, but it can easily be converted into an equivalent explanatory learner, as the oracle is above $\mathbb{K}$.

Suppose that M is presented with a text T for the set $\{e\} \oplus \mathbb{N}$. First, assume that W_e is cofinite. Then there is a least number x such that for all $y > x$, y is contained in W_e. Further, for a sufficiently long segment σ of the text, $\{z \leq x : z \notin W_e\} \subseteq \text{range}(\sigma)$ and $|\sigma| > x$ both hold. Hence M will converge on T to a fixed index for the set $\{e\} \oplus \mathbb{N}$. Secondly, assume that W_e is coinfinite. In this case, the condition $\forall y > x \exists s[y \in W_{e,s}]$ fails to hold for all x, and so M will conjecture the set $\{e\} \oplus \mathbb{N}$ on all segments of T. Next, suppose that M is fed with a text T' for the set $\{e\} \oplus (W_e \cup D)$, where W_e is cofinite and D is finite. Let x be the minimum number such that for all $y \geq x$, $y \in W_e$ holds. Then, upon witnessing a segment σ of T' with $|\sigma| \geq x$ which contains all the elements of D, M will thenceforth always conjecture a fixed index for $\{e\} \oplus (W_e \cup D)$. Therefore M is an $Ex[\mathbb{K}']$ learner of $\mathcal{C}$, as required.

On the other hand, assume for the sake of a contradiction that N were a partially conservative learner of $\mathcal{C}$. Fix any number e and give the text $2e \circ 1 \circ 3 \circ 5 \circ \ldots \circ (2n+1) \circ \ldots$ to N. Since N partially learns the set $\{e\} \oplus \mathbb{N}$, there is a least number k such that N outputs an index for $\{e\} \oplus \mathbb{N}$ on the segment $2e \circ 1 \circ \ldots \circ 2k+1$; moreover, one can search for k by means of the oracle $\mathbb{K}'$. One may subsequently check relative to $\mathbb{K}'$ whether or not $\forall z > k \exists s[z \in W_{e,s}]$ holds. If it does hold, then W_e is cofinite; otherwise, W_e must be coinfinite, for if W_e were cofinite and $z > k$ were a number such that $z \notin W_e$, then the segment $2e \circ 1 \circ \ldots \circ 2k+1$ may be extended to a text for $\{e\} \oplus (W_e \cup \{0,1,\ldots,k\})$, and since N outputs an index for some set of which $\{e\} \oplus (W_e \cup \{0,1,\ldots,k\})$ is a proper subset, this implies that N cannot partially conservatively learn $\{e\} \oplus (W_e \cup \{0,1,\ldots,k\})$, contrary to hypothesis. Thus the initial assumption would lead to a decision procedure relative to $\mathbb{K}'$ for the Π^0_3-complete set $\{e : W_e \text{ is coinfinite}\}$, a contradiction. In conclusion, $\mathcal{C}$ is not partially conservatively learnable, as required. ◀

The following theorem sharpens Theorem 7, demonstrating that by imposing the further learning constraint of *prudence*, *ConsvBC* learnability does not in general guarantee *ConsvPart* learnability.

► **Theorem 19.** *There is a class of r.e. languages which is ConsvPart learnable and prudently ConsvBC learnable, but not prudently ConsvPart learnable.*

Proof. Let S be a simple set. We shall show that while the class $\mathcal{L} = \{S\} \cup \{D : D \text{ is finite and } D \not\subseteq S\}$ is *PrudConsvBC* learnable, it cannot be *PrudConsvPart* learnt. By Theorem 7, the class is also *ConsvPart* learnable.

A prudent *ConsvBC* learner N of $\mathcal{L}$ may work by conjecturing, on input σ, the r.e. set

$$W_{N(\sigma)} = \begin{cases} \text{content}(\sigma) & \text{if content}(\sigma) \not\subseteq S; \\ S & \text{if content}(\sigma) \subseteq S. \end{cases}$$

If N is fed with a text for S, then it will always conjecture S. If N is fed with a text for some $D \not\subseteq S$, then it will conjecture indices for S until the first member of $D - S$ appears in the text; from then onwards it will conjecture the finite set comprising the range of the current input. Since N only outputs r.e. indices of sets in $\mathcal{L}$, it also prudently learns $\mathcal{L}$.

Now proceeding by contradiction, assume that $\mathcal{L}$ is *PrudConsvPart* learnable via a recursive learner M. By Lemma 5, there is a sequence $\tau \in S^*$ such that $W_{M(\tau)} = S$. Fixing τ, let y be any given number. It shall be shown that the existence of M provides an effective procedure for deciding whether or not $y \in S$, contradicting the fact that S is nonrecursive.

First, let A be the r.e. set $\{x : \exists \sigma \succeq \tau[M(\sigma) \neq M(\tau) \wedge \text{range}(\tau) \cup \{x\} \subseteq W_{M(\sigma)} \wedge x \notin \text{range}(\sigma)]\}$. For any t, denote by S_t the t-th approximation to S. Let x be any number contained in A, so that for some $\sigma \succeq \tau$ with $x \notin \text{range}(\sigma)$, $\text{range}(\sigma) \cup \{x\} \subseteq W_{M(\sigma)}$. If $W_{M(\sigma)} \subseteq S$, then, by the prudence of M, M must *ConsvPart* learn $W_{M(\sigma)}$, but since $M(\sigma) \neq M(\tau)$, $W_{M(\sigma)} \neq S$, and so $W_{M(\sigma)}$ is a proper subset of S; this contradicts the partial conservativeness of M, as it conjectures S on τ, a prefix of σ. Hence there is some $z \in W_{M(\sigma)} \cap \overline{S}$. If $x \in S$, then $x \neq z$, and so M does not *ConsvPart* learn the set $\text{content}(\sigma) \cup \{z\} \in \mathcal{L}$ since its conjecture $W_{M(\sigma)}$ is a proper superset of $\text{content}(\sigma) \cup \{z\}$.

Therefore $A \cap S = \emptyset$, and as S is simple, A must be finite. If $y \in A$, then $y \notin S$ immediately follows; thus it may be assumed in the subsequent argument that $y \notin A$.

Now let $f(y)$ be the first number found such that $M(\tau \circ y^{f(y)}) \neq M(\tau)$ and $\text{content}(\tau) \cup \{y\} \subseteq W_{M(\tau \circ y^{f(y)})}$, or $y \in S_{f(y)}$; such a number must exist as M learns every finite set that intersects $\overline{S}$. For the same reason, one can find, for each $x \in A$, a number $g(x, y)$ such that $\text{content}(\tau) \cup \{x, y\} \subseteq W_{M(\tau \circ y^{f(y)} \circ x^{g(x,y)})}$.

Suppose that $y \in S$: as was argued previously, $W_{M(\tau \circ y^{f(y)})} \not\subseteq S$, and so by the definition of A there is some $x \in A \cap W_{M(\tau \circ y^{f(y)})}$. As $\text{content}(\tau) \cup \{x, y\} \in \mathcal{L}$, and for any $L \in \mathcal{L}$, M cannot conjecture a proper superset of L on any text segment of L, it must hold that $W_{M(\tau \circ y^{f(y)})} = W_{M(\tau \circ y^{f(y)} \circ x^{g(x,y)})} = \text{content}(\tau) \cup \{x, y\}$. Moreover, M must output exactly one correct index on any text segment of L, and thus $M(\tau \circ y^{f(y)}) = M(\tau \circ y^{f(y)} \circ x^{g(x,y)})$. On the other hand, if $y \notin S$, then $\text{content}(\tau) \cup \{y\} = W_{M(\tau \circ y^{f(y)})}$, giving that $M(\tau \circ y^{f(y)}) \neq M(\tau \circ y^{f(y)} \circ x^{g(x,y)})$ for all $x \in A$. Hence if $y \notin A$, one has the recursive condition $y \in S \Leftrightarrow \exists x \in A[M(\tau \circ y^{f(y)} \circ x^{g(x,y)}) = M(\tau \circ y^{f(y)})]$, contradicting the fact that S is nonrecursive. ◀

Angluin [2] introduced the concept of a *telltale set*, which is a finite subset E_L of some language L to be learnt such that no language different from L and containing E_L can be a proper superset of L, and she characterised indexed families of recursive languages which are explanatorily learnable from positive data using this notion. The next two theorems propose criteria for partially conservative learnability as well as conservative $Ex[\mathbb{K}]$ learnability in a similar fashion to Angluin's original formulation.

▶ **Theorem 20.** *A class $\mathcal{C}$ is ConsvPart learnable iff there is a recursive sequence of pairs (i_e, j_e) such that*

1. *$D_{i_e} \subseteq W_{j_e}$ for all e;*
2. *For all $L \in \mathcal{C}$ there is an e with $L = W_{j_e}$;*
3. *For all d, e, if $D_{i_e} \subseteq W_{j_d} \subset W_{j_e}$ then $W_{j_d} \notin \mathcal{C}$.*

Proof. If M *ConsvPart*-learns a class $\mathcal{C}$ then one can make the first sequence as an enumeration of the r.e. set of all (i, j) for which there is a σ with $D_i = range(\sigma)$ and $j = M(\sigma)$ and $range(\sigma) \subseteq W_j$. Clearly for every $L \in \mathcal{C}$ there is a pair (i, j) enumerated for which $M(\sigma) = j$, $W_j = L$ and $D_i = range(\sigma) \subseteq L$. Furthermore, if $L \in \mathcal{C}$ then there is no σ with $range(\sigma) \subseteq L \subset W_{M(\sigma)}$. Hence all three conditions are satisfied.

For the converse direction, assume that a sequence of such (i_e, j_e) is given. Then a learner N conjectures all sets W_{j_e} for which all elements of D_{i_e} have shown up in the text of the set L to be learnt. This set L equals some W_{j_d} by the second condition. Clearly j_d will be among the indices conjectured. Furthermore, for each index j_e conjectured by N, it holds that the condition $D_{j_e} \subseteq W_{j_d} \subset W_{j_e}$ is not satisfied and therefore W_{j_e} is not a proper superset of L. Hence the learner can be transformed into a conservative partial learner by means of Theorem 8. ◀

▶ **Theorem 21.** *A class $\mathcal{C}$ is $ConsvEx[\mathbb{K}]$ learnable iff there is a recursive sequence of pairs (i_e, j_e) such that*

1. *$W_{i_e} \subseteq W_{j_e}$ for all e;*
2. *For all $L \in \mathcal{C}$ there is an e with W_{i_e} being finite and $L = W_{j_e}$;*
3. *For all d, e, if W_{i_e} is finite and $W_{i_e} \subseteq W_{j_d} \subset W_{j_e}$ then $W_{j_d} \notin \mathcal{C}$.*

Proof. If M $ConsvEx[\mathbb{K}]$-learns a class $\mathcal{C}$ then one can enumerate for each $\sigma \preceq T$ and each s with $range(\sigma) \subseteq W_{M^{\mathbb{K}_s}(\sigma)}$ a pair (i,j) satisfying the following conditions:

$$W_i = \begin{cases} range(\sigma) & \text{if } \forall t \geq s\,[M^{\mathbb{K}_t}(\sigma) = M^{\mathbb{K}_s}(\sigma)]; \\ \mathbb{N} & \text{if } \exists t \geq s\,[M^{\mathbb{K}_t}(\sigma) \neq M^{\mathbb{K}_s}(\sigma)]. \end{cases}$$

$$W_j = \begin{cases} W_{M^{\mathbb{K}_s}(\sigma)} & \text{if } \forall t \geq s\,[M^{\mathbb{K}_t}(\sigma) = M^{\mathbb{K}_s}(\sigma)]; \\ \mathbb{N} & \text{if } \exists t \geq s\,[M^{\mathbb{K}_t}(\sigma) \neq M^{\mathbb{K}_s}(\sigma)]. \end{cases}$$

The first condition holds clearly. The second condition is satisfied as for each $L \in \mathcal{C}$ there is a $\sigma \in (L \cup \{\#\})^*$ such that $M(\sigma)$ is an index for L. So, for sufficiently large s, σ and s cause a pair (i,j) to be enumerated where $W_i = range(\sigma)$ and $W_j = L$. The third condition follows from the fact that M is conservative. So if W_{j_d} is in $\mathcal{C} - \{\mathbb{N}\}$ and $W_{i_e} \subseteq W_{j_d}$ then W_{i_e} is finite and W_{j_e} is equal to a set conjectured by M on a string of elements from W_{j_d} and therefore W_{j_e} is not a proper superset of W_{j_d}.

For the converse direction, assume that a recursive sequence of (i_e, j_e) as above is given. One can now make a $ConsvEx[\mathbb{K}]$-learner N which does the following: If there is no current conjecture or the current conjecture is inconsistent (that is, $range(\sigma)$ is not in the set what can be checked with $\mathbb{K}$), then N checks using $\mathbb{K}$ whether there is a $e \leq |\sigma|$ with $W_{i_e} \subseteq range(\sigma) \subseteq W_{j_e}$. This check can be done, as the learner checks whether W_{i_e} enumerates some element outside $range(\sigma)$ and whether W_{j_e} enumerates all elements inside $range(\sigma)$. If so, then the learner conjectures j_e for the least such e else the learner abstains from conjecturing a hypothesis. Assume now that a set W_{j_d} has to be learnt and that W_{i_d} is finite. Then each conjecture W_{j_e} of N on a text for W_{j_d} satisfies $W_{i_e} \subseteq W_{j_d}$ and hence $W_{j_d} \not\subset W_{i_e}$ by the third condition. So the learner does not overgeneralise and is conservative. Furthermore, eventually all hypotheses W_{j_e} which do not contain all elements of W_{j_d} get cancelled eventually and so the learner N ends up with either conjecturing d or conjecturing a set $W_{j_e} = W_{j_d}$. Thus the learner is indeed learning $\mathcal{C}$. ◀

The learning constraint of *consistency* is often studied closely with conservativeness; indeed, consistency may be viewed as a restrain on conservativeness, for a conservative learner performs a mind change only if its latest conjecture is inconsistent with the current input data. One can show that every $ConsvPart$ learnable class is learnable in the following sense: the learner is both consistent and partially conservative, and it outputs exactly one correct index at least once, although this index need not be output infinitely often.

▶ **Theorem 22.** *For every $ConsvPart$ learnable class $\mathcal{C}$, there is a recursive learner N such that, on any text for some $L \in \mathcal{C}$,*
- *N is consistent and does not output an index for a proper superset of L;*
- *N outputs exactly one index e with $L = W_e$ and every other index only finitely often.*

Proof. Assume that the class $\mathcal{C}$ of r.e. languages is $ConsvPart$ learnable via a recursive learner M. Let f be a recursive function such that for all numbers e and finite sets $D \subset \mathbb{N}$, $W_{f(e,D)} = \bigcup_{s \in A_{e,D}} W_{e,s} \cup D$, where $s \in A_{e,D}$ iff $D \subseteq W_{e,s} \subset W_{e,s+1}$. Denote by σ' the sequence $\sigma[|\sigma| - 1]$ if $|\sigma| > 1$, and the empty sequence λ otherwise. Let $B = \bigcup_{s \in \mathbb{N}} B_s$ be an auxiliary r.e. set with $B_0 = \emptyset$. Now define a new learner N that performs the following on input σ. First, let $e_0, e_1, \ldots, e_j$ be all the distinct indices output by M on piecewise increasing prefixes of σ.

1. If content(σ) = content(σ'), Then $N(\sigma) = N(\sigma')$ and $B_{|\sigma|} = B_{|\sigma|-1}$,
2. Else If $i = \min\{k \leq j : e_k \notin B_{|\sigma|-1}\}$ exists, Then N outputs $f(e_i, \text{content}(\sigma))$ and updates $B_{|\sigma|} = B_{|\sigma|-1} \cup \{e_i\}$,

3. Else (i does not exist) N outputs a canonical index for the set content(σ) and updates $B_{|\sigma|} = B_{|\sigma|-1}$.

For the verification, suppose that N is fed with a text T for some language $L \in \mathcal{C}$. First, suppose that L is finite. There is a shortest text prefix $T[l]$ with content$(T[l]) = L$, so that whenever $k > l$, case 1 applies, and N converges to the index $N(T[l])$. On input $T[l]$, either case 2 or case 3 applies. If case 2 applies, then, since the new conjecture e_i cannot be a proper superset of L, either it does not contain content$(T[l])$, or it is precisely equal to content$(T[l])$, and so $W_{N(T[l])} = W_{f(e_i,\text{content}(T[l]))} = L$. If 3. applies, then $W_{N(T[l])} = L$ also holds. Furthermore, for all $k < l$, as content$(T[k]) \subset L$ and every incorrect conjecture of M does not contain some element of L, while if $W_{e_i} = L$ then $W_{f(e_i,\text{content}(T[k]))} \subset L$, $W_{N(T[k])} \neq L$ holds. Secondly, suppose that L is infinite. Then case 1 does not hold at infinitely many stages, and therefore N outputs each index only finitely often. If case 2 applies on some text prefix $T[k]$ to an incorrect conjecture e_i with W_{e_i} infinite, then either content$(T[k]) \not\subset W_{e_i}$ and $W_{f(e_i,\text{content}(T[k]))} = \text{content}(T[k])$, or content$(T[k]) \subset W_{e_i}$ and there is some $x \in L - W_{e_i}$, so that $W_{N(T[k])} \neq L$ holds whenever $W_{e_i} \neq L$. If case 2 applies and e_i is the single correct index that M outputs on T, then $W_{N(T[k])} = W_{f(e_i,\text{content}(T[k]))} = W_{e_i}$, that is, N outputs a correct conjecture. Finally, every index that N outputs in case 3 is incorrect, and this establishes that N satisfies the learning requirements of the theorem. ◀

The following theorem gives a necessary condition on all consistently partially learnable classes of r.e. languages; it suggests that partial consistency may be quite a strong learning requirement, as only recursive languages can be partially consistently learnt.

▶ **Theorem 23.** *If a class $\mathcal{L}$ of r.e. languages is ConsPart learnable, then there is a uniformly recursive family $\mathcal{A}$ such that $\mathcal{L} \subseteq \mathcal{A}$.*

Proof. Given a recursive *ConsPart* learner M of $\mathcal{L}$, let $\mathcal{A}$ be the uniformly recursive family of all languages $L_{D,\sigma,\tau} = \{x : \exists \eta \preceq \tau[M(\sigma \circ x \circ \eta) \in D]\}$. For the verification that $\mathcal{L} \subseteq \mathcal{A}$, choose any $L \in \mathcal{L}$; it shall be shown that there are sequences $\sigma, \tau \in L^*$ such that for all $x \in L$, $M(\sigma \circ x \circ \eta) = M(\gamma)$ for some $\eta \preceq \tau$ and $\gamma \preceq \sigma$ with $W_{M(\gamma)} = L$. Proceeding by way of a contradiction, assume that for all $\sigma, \tau \in L^*$, there is an $x \in L$ such that $M(\sigma \circ x \circ \eta) \neq M(\gamma)$ for all $\gamma \preceq \sigma$ with $W_{M(\gamma)} = L$ and $\eta \preceq \tau$. Consequently, one can build a text T for L on which M does not output any correct index infinitely often: first, let $a_0, a_1, a_2, \ldots$ be a recursive one-one enumeration of L. Set $T(0) = a_0$. At stage $s+1$, suppose that T has been defined up to and including $2s$. By assumption, there is a number $x_{s+1} \in L$ such that at least one of the following conditions holds: $W_{M(T[2s+1]\circ x_{s+1})} \neq L$ and $W_{M(T[2s+1]\circ x_{s+1}\circ a_{s+1})} \neq L$; or, for all $k \leq 2s+1$, $M(T[2s+1] \circ x_{s+1}) \neq M(T[k])$ and $M(T[2s+1] \circ x_{s+1} \circ a_{s+1}) \neq M(T[k])$. Set $T(2s+1) = x_{s+1}$ and $T(2s+2) = a_{s+1}$. It follows from this construction that T is a text on which M never outputs a single correct index for L infinitely often, as claimed. Hence there are sequences $\sigma, \tau \in L^*$ such that for all $x \in L$, $M(\sigma \circ x \circ \eta) = M(\gamma)$ for some $\eta \preceq \tau$ and $\gamma \preceq \sigma$ with $W_{M(\gamma)} = L$, that is, $L \subseteq \{x : \exists \eta \preceq \tau[M(\sigma \circ x \circ \eta) \in \{M(\gamma) : \gamma \preceq \sigma \wedge W_{M(\gamma)} = L\}]\}$. Furthermore, by the partial consistency of M, the reverse inclusion $\{x : \exists \eta \preceq \tau[M(\sigma \circ x \circ \eta) \in \{M(\gamma) : \gamma \preceq \sigma \wedge W_{M(\gamma)} = L\}]\} \subseteq L$ also holds. By setting $D = \{M(\gamma) : \gamma \preceq \sigma \wedge W_{M(\gamma)} = L\}$, one has that $L = L_{D,\sigma,\tau} \in \mathcal{A}$, establishing that $\mathcal{L} \subseteq \mathcal{A}$, as required. ◀

▶ Remark. The class $\mathcal{L} = \{\{e\} \oplus \{x : x \leq |W_e|\} : W_e \text{ is finite}\}$ is prudently as well as consistently explanatorily learnable, but it has no uniformly recursive indexing. To *PrudConsEx* learn this class, the learner may output a canonical index for content(σ) on every input σ. However, if one assumes that $\mathcal{L}$ were uniformly recursively indexed by $\{L_0, L_1, L_2, \ldots\}$, then for every number e one may check via oracle $\mathbb{K}$ whether or not the

Σ^0_1 condition $\exists n[2e \in L_n]$ holds. If this condition holds, then W_e is finite; otherwise, W_e is infinite. Hence the assumption would imply that the Π^0_2-complete set $\{e : W_e \text{ is infinite}\}$ is Turing reducible to $\mathbb{K}$, a contradiction.

Furthermore, partial consistency does not follow directly from partial conservativeness; this is witnessed by the class of graphs of all recursive functions. Moreover, the class $\{\mathbb{K}, \mathbb{N}\}$ establishes one direction of separation between $PrudConsvEx$ and consistent partial learning, which also gives by Theorem 7 that consistent partial learning is not less restrictive than $ConsvPart$ learning. Blum and Blum [7] showed that this particular class is prudently $ConsvEx$ learnable but not consistently explanatorily learnable. For the converse separation of $PrudConsvEx$ learning and consistent partial learning, one may consider the following example.

▶ **Example 24.** The indexed family $\mathcal{C} = \{L_0, L_1, L_2, \ldots\}$ of recursive sets, where

$$L_{\langle e,0,0\rangle} = \{e + x : x \in \mathbb{N}\}, L_{\langle e,i,j\rangle} = \begin{cases} \{e + x : x \leq j\} & \text{if } e \in \mathbb{K}_i - \mathbb{K}_j; \\ \{e + x : x \in \mathbb{N}\} & \text{if } e \notin \mathbb{K}_i \vee e \in \mathbb{K}_j, \end{cases}$$

is prudently consistently Ex learnable but not $ConsvPart$ learnable.

Proof. A prudent consistent Ex learner M outputs a canonical index for $\emptyset$ until the range of the input σ is nonempty with $e = \min(\{x : x \in \text{content}(\sigma)\})$ and $e + d = \max(\{x : x \in \text{content}(\sigma)\})$. M then conjectures an index for the set $\{e+x : x \in \mathbb{N}\}$ if $e \notin \mathbb{K}_{|\sigma|}$ or if $e \in \mathbb{K}_d$, and an index for the set $\{e + x : x \leq d\}$ if $e \in \mathbb{K}_{|\sigma|} - \mathbb{K}_d$. Suppose that M is fed with a text for the set $\{e + x : x \in \mathbb{N}\}$. If $e \notin \mathbb{K}$ then M will always output an index for the correct set. If $e \in \mathbb{K}_{s+1} - \mathbb{K}_s$, then M will converge to a correct index once the element $e + s + 1$ occurs in a segment of the text of length at least s. On the other hand, if M processes a text of the set $\{e + x : x \leq d\}$ with $e \in \mathbb{K}_s - \mathbb{K}_d$ for some $s > d$, then it will also converge to a correct index on inputs containing $\{e + x : x \leq d\}$ with length at least s.

For the sake of a contradiction, suppose that N were a partially conservative learner of $\mathcal{C}$. Define a recursive function f by setting $f(e)$ to be the first number d found such that $\{e, e + 1, \ldots, e + d + 1\} \subseteq W_{N(e \circ e+1 \circ \ldots \circ e+d)}$. Since N learns the set $\{e + x : x \in \mathbb{N}\}$, such a number d must exist, and so f is a recursive function. Furthermore, owing to the partial conservativeness of N, it follows that $e \in \mathbb{K}$ holds if and only if $e \in \mathbb{K}_{f(e)}$. This provides a recursive procedure for the halting problem, which is a contradiction. Thus N cannot be a partially conservative learner of $\mathcal{C}$, as required. ◀

▶ **Example 25.** If A is a nonrecursive r.e. set, then the class $\mathcal{C} = \{A \cup \{x\} : x \in \mathbb{N}\}$ is $ConsvBC$ learnable but neither Ex learnable nor class consistently partially learnable.

Proof. A $ConsvBC$ learner of $\mathcal{C}$ may work by always outputting, on input σ, an index for the r.e. set $A \cup \text{content}(\sigma)$. Such a learner M is always conservative, for if $W_{M(\sigma)} \neq W_{M(\tau)}$ for some $\sigma \prec \tau$, then, since $W_{M(\sigma)} \subset W_{M(\tau)}$ and $W_{M(\tau)} - W_{M(\sigma)} \subseteq \text{content}(\tau)$, there is some number $x \in \text{content}(\tau) - W_{M(\sigma)}$.

Now assume by way of a contradiction that N were a class consistent partial learner of $\mathcal{C}$. The proof proceeds by building a text T for A on which N does not output any correct index infinitely often. Let a_0 be the first enumerated member of A, and set $T(0) = a_0$. At stage $s+1$, suppose that $T(i) = a_i$ and $a_i \in A$ for all $0 \leq i \leq 2s$. Let $b_{s+1} = \min(\{A - \text{range}(T[2s+1])\})$, and search for some $a \in A$ such that the condition $\forall k < 2s + 1[(N(T[2s + 1] \circ a) \neq N(T[k]) \wedge N(T[2s+1] \circ a \circ b_{s+1}) \neq N(T[k])) \vee (W_{N(T[2s+1] \circ a)} \neq A \wedge W_{N(T[2s+1] \circ a \circ b_{s+1})} \neq A)]$ holds. Set $T(2s + 1) = a$ and $T(2s + 2) = b_{s+1}$. This search must eventually terminate successfully, for if, whenever $a \in A$, then $e_i \in \{N(T[2s+1] \circ a), N(T[2s+1] \circ a \circ b_{s+1})\}$ for some $e_i \in \{e_0, e_1, \ldots, e_l\}$, where $\{e_0, e_1, \ldots, e_l\} \subseteq \{N(T[k]) : k < 2s+1\}$ and $W_{e_i} = A$ whenever $0 \leq i \leq l$, then the class consistency of N would imply that for all x, $x \in A$ iff $N(T[2s+1] \circ x) \in \{e_0, e_1, \ldots, e_l\} \vee N(T[2s+1] \circ x \circ b_{s+1}) \in \{e_0, e_1, \ldots, e_l\}$ holds, and this would furnish an effective decision procedure for the membership problem of A, a contradiction. Thus T is a text for A on which N does not output any correct index infinitely often, which is the desired contradiction. ◀

5 Conclusion

The main results of the foregoing discussion may be summarised in the following diagram. The set of all classes of r.e. languages that are learnable according to each criterion is represented by the corresponding notation given in Definitions 1 and 2. The learning criterion in Theorem 22 is denoted by $WConsConsvPart$.

$$
\begin{tikzcd}
 & & & \text{WConsConsvPart} \arrow[d, leftrightarrow] & \\
\text{ConsvEx} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \text{ConsvVac} \arrow[r] \arrow[d, leftrightarrow] & \text{ConsvBC} \arrow[r] & \text{ConsvPart} \arrow[r] & \text{ConsvEx}[\mathbb{K}] \\
\text{PrudConsvEx} \arrow[r, leftrightarrow] & \text{PrudConsvVac} \arrow[r] & \text{PrudConsvBC} & \text{PrudConsvPart} \arrow[u] & \\
\text{ClsPresvConsvEx} \arrow[r] \arrow[u] & \text{ClsPresvConsvVac} \arrow[r] \arrow[u] & \text{ClsPresvConsvBC} \arrow[u] & &
\end{tikzcd}
$$

References

1 Leonard Adleman and Manuel Blum. Inductive inference and unsolvability. *Journal of Symbolic Logic* 56 (1991): 891–900.

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

3 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.

4 Janis Bārzdiņš. Inductive inference of automata, functions and programs. In *Proceedings of the 20th International Congress of Mathematicians, Vancouver*, pages 455-560, 1974. In Russian. English translation in American Mathematical Society Translations: Series 2, 109: 107–112, 1977.

5 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.

6 John Case. The power of vacillation in language learning. *SIAM Journal on Computing* 28(6) (1999): 1941–1969.

7 Lenore Blum and Manuel Blum. Toward a mathematical theory of inductive inference. *Information and Control* 28 (1975): 125–155.

8 Dick de Jongh and Makoto Kanazawa. Angluin's theorem for indexed families of r.e. sets and applications. COLT 1996: 193–204.

9 Jerome Feldman. Some decidability results on grammatical inference and complexity. *Information and Control* 20 (1972): 244–262.

10 Mark Fulk. Prudence and other conditions on formal language learning. *Information and Computation* 85 (1990): 1–11.

11 Ziyuan Gao, Frank Stephan, Guohua Wu and Akihiro Yamamoto. Learning families of closed sets in matroids. *Computation, Physics and Beyond; International Workshop on Theoretical Computer Science*, WTCS 2012, Springer LNCS 7160 (2012): 120–139.

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

13 Gunter Grieser. Reflective inductive inference of recursive functions. *Theoretical Computer Science A* 397(1–3) (2008): 57–69.

14 Sanjay Jain, Daniel Osherson, James S. Royer and Arun Sharma. 1999. *Systems that learn: an introduction to learning theory.* Cambridge, Massachusetts.: MIT Press.

15 Sanjay Jain, Frank Stephan and Nan Ye. Prescribed learning of r.e. classes. *Theoretical Computer Science* 410(19) (2009): 1796–1806.

16 Sanjay Jain and Frank Stephan. Consistent partial identification. COLT 2009: 135–145.

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

18 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.

19 Hartley Rogers, Jr. 1987. *Theory of recursive functions and effective computability.* Cambridge, Massachusetts: MIT Press.

20 Rolf Wiehagen and Thomas Zeugmann. Learning and consistency. *Algorithmic Learning for Knowledge-Based Systems*, GOSLER Final Report, Springer LNAI 961 (1995): 1–24.

21 Thomas Zeugmann and Sandra Zilles. Learning recursive functions: a survey. *Theoretical Computer Science* 397(1-3) (2008): 4–56.