Numberings Optimal for Learning

Sanjay Jain*[1] and Frank Stephan*[2]

[1] Department of Computer Science,
National University of Singapore, Singapore 117543, Republic of Singapore.
sanjay@comp.nus.edu.sg
[2] Department of Computer Science and Department of Mathematics,
National University of Singapore, Singapore 117543, Republic of Singapore.
fstephan@comp.nus.edu.sg

Abstract. This paper extends previous studies on learnability in non-acceptable numberings by considering the question: for which criteria which numberings are optimal, that is, for which numberings it holds that one can learn every learnable class using the given numbering as hypothesis space. Furthermore an effective version of optimality is studied as well. It is shown that the effectively optimal numberings for finite learning are just the acceptable numberings. In contrast to this, there are non-acceptable numberings which are optimal for finite learning and effectively optimal for explanatory, vacillatory and behaviourally correct learning. The numberings effectively optimal for explanatory learning are the K-acceptable numberings. A similar characterization is obtained for the numberings which are effectively optimal for vacillatory learning. Furthermore, it is studied which numberings are optimal for one and not for another criterion: among the criteria of finite, explanatory, vacillatory and behaviourally correct learning all separations can be obtained; however every numbering which is optimal for explanatory learning is also optimal for consistent learning.

1 Introduction

Consider the following model of learning. The learner receives, over time, more and more data about the concept to be learnt. From time to time, the learner conjectures a potential explanation for the data it is receiving. One can say that the learner learns the concept if the sequence of conjectures eventually converges to a correct explanation for the concept. This is essentially the notion of explanatory learning considered by Gold [10]. The concepts considered are usually recursively enumerable (r.e.) languages (subsets of natural numbers) or recursive functions. In this paper we will be concentrating on learning languages. The explanations thus take the form of grammars or indices from some hypothesis space or numbering of recursively enumerable languages.

Learning of just one r.e. language is not useful, as a learner which just conjectures a grammar for the language, on any data, will be successful on the language. Thus, it is more useful to

* Supported in part by NUS grant number R252-000-308-112.

consider learnability of a class of languages. A learner explanatorily learns a class of languages if it explanatorily learns each language in the class. Since Gold's paper [10], several other criteria of learnability have been explored and some of them will be considered in the current paper.

The learnability of the class depends not only on the class itself but also on the underlying numbering used as a hypothesis space. Angluin [1] initiated the systematic study of uniformly recursive hypothesis spaces; as such hypothesis spaces can contain only some but not all recursive sets, these spaces have to be selected in dependence of the class to be learnt. Lange and Zeugmann [14, 15, 24] investigated the topic thoroughly. De Jongh and Kanazawa [6] investigated to which extent one can generalize Angluin's characterization of learnability [1] from uniformly recursive to uniformly r.e. hypothesis spaces. Zilles [25, 26] studied the question of how to synthesize a learner from an index of a uniformly r.e. hypothesis space. Most of this and related work considered specialized hypothesis spaces, which permit only to learn some and not all classes; these specialized hypothesis spaces often do not even contain all r.e. sets.

In contrast to this, the focus of the present work lies on the question which hypothesis spaces are optimal for learning in the sense that every learnable class can be learnt using this hypothesis space. Therefore, a valid hypothesis space $A_0, A_1, A_2, \dots$ must be a universal numbering, that is, it must satisfy that $\{\langle e, x\rangle : x \in A_e\}$ is recursively enumerable and that, for every r.e. set B, there is an index e with $B = A_e$. In particular, acceptable, K-acceptable and Ke-numberings are considered, where K denotes the halting problem. (Here a numbering $A_0, A_1, A_2, \dots$ is acceptable (K-acceptable) if for every further numbering $B_0, B_1, B_2, \dots$ there is a recursive (K-recursive) function f such that $B_e = A_{f(e)}$ for all e. A Ke-numbering is a universal numbering for which the grammar equivalence problem is K-recursive.) A more restrictive notion is that of an effectively optimal hypothesis space where, additionally, one can effectively obtain a learner for the class using $A_0, A_1, A_2, \dots$ as hypothesis space from any learner for the class using another numbering B_0, $B_1, B_2, \dots$ as hypothesis space.

The optimality of the hypothesis space depends on the criterion of learning considered. The main criteria considered are finite, explanatory, vacillatory and behaviourally correct learning as defined below in Definition 1; but some interesting results are also obtained for other criteria of learning.

Intuitively, a learner M finitely learns [10] a language class if, for every language L in the class, for any order of presentation of elements of L, M outputs only one conjecture and the conjecture is an index for L. A learner M explanatorily learns [10] a language class if, for every language L in the class, for any order of presentation of elements of L, M outputs a sequence of conjectures which converges to an index for L. A learner M behaviourally correctly learns [2, 20] a language class if, for every language L in the class, for any order of presentation of elements of L, M outputs an infinite sequence of conjectures, all but finitely many of which are indices for L. Vacillatory learning [4] is a restriction of behaviourally correct learning, where the learner outputs only finitely many distinct conjectures (although some of them might be repeated infinitely often).

The most prominent numberings are the acceptable numberings and Friedberg numberings. Acceptable numberings are used by many authors as the standard hypothesis space [10] and every learnable class (according to most criteria) is also learnable using an acceptable numbering —

one exception is the criterion of learning with additional information, see Theorem 31. However, one-one numberings, also known as Friedberg numberings [8], are not optimal for learning [7, 12]. A central contribution of the present work is to show that there are many optimal numberings besides the acceptable numberings, but that it depends a lot on the underlying learning criterion which numberings are optimal for learning and which are not. For example, a nearly acceptable numbering (as defined in Definition 4) is effectively optimal for explanatory, vacillatory and behaviourally correct learning as well as optimal for finite learning (see Proposition 5).

In Theorem 6, we show characterizations for numberings which are effectively optimal for finite, explanatory and vacillatory learning. In particular, a numbering $A_0, A_1, A_2, \ldots$ is effectively optimal for finite learning iff the numbering is acceptable. A numbering $A_0, A_1, A_2, \ldots$ is effectively optimal for explanatory learning iff the numbering is K-acceptable. One can also similarly characterize effectively optimal numberings for vacillatory learning. We do not have a good characterization of numberings which are effectively optimal for behaviourally correct learning.

We show that there are numberings which are (non-effectively) optimal but not effectively optimal for various criteria of inference: Theorem 9 gives this result for finite learning; Theorem 13 gives this result for explanatory and vacillatory learning; Theorem 16 gives this result for behaviourally correct learning.

We also show that the set of optimal numberings for finite, explanatory, vacillatory and behaviourally correct learning are incomparable. Theorem 9 gives this result for finite learning versus explanatory, vacillatory and behaviourally correct learning. Theorem 12 gives this result for behaviourally correct learning versus finite, explanatory and vacillatory learning. Theorem 11 gives this result for explanatory and vacillatory learning versus finite learning and behaviourally correct learning. Theorem 10 gives this result for vacillatory learning versus explanatory learning. The numbering $A_0, A_1, A_2, \ldots$ in Theorem 13 gives this result for explanatory learning versus vacillatory learning.

In Section 4 we give special attention to consistent learning. Theorem 22 shows that optimal numberings for explanatory learning are optimal for consistent learning. This is one of the rare cases of an inclusion in the sense that every numbering optimal for a criterion I is also optimal for a different criterion J. The inclusion also holds with effective optimality in place of optimality. However, there are numberings which are effectively optimal for consistent learning but not optimal for finite, explanatory, vacillatory or behaviourally correct learning.

2 Preliminaries

Any unexplained recursion theoretic notion is from [21, 17, 18]. $\mathbb{N}$ denotes the set of natural numbers. Languages are recursively enumerable (r.e.) subsets of $\mathbb{N}$. We often identify L with its characteristic function, that is $L(x) = 1$ denotes that $x \in L$ and $L(x) = 0$ denotes that $x \notin L$.

For the ease of notation, learnability of r.e. subsets of $\mathbb{N}$ is studied (and other possible domains are ignored). The learners use some hypothesis space to represent their conjectures.

A *numbering* $A_0, A_1, \ldots$ is any listing of recursively enumerable sets such that $\{\langle e, x\rangle : x \in A_e\}$ is recursively enumerable. A numbering is called a *universal numbering* if it contains an index for all r.e. sets. The standard hypothesis space $W_0, W_1, W_2, \ldots$ is some fixed *acceptable numbering* [21], that is, for every further numbering $A_0, A_1, A_2, \ldots$ of r.e. sets, there is a recursive function f with $W_{f(e)} = A_e$ for all e. In general, every universal numbering can be a hypothesis space. A numbering $A_0, A_1, A_2, \ldots$ is called K-acceptable iff, for every further numbering $B_0, B_1, B_2, \ldots$, there is a K-recursive function f with $A_{f(e)} = B_e$ for all e. Here K denotes the halting problem $\{x : x \in W_x\}$. The concept of translating numberings in the limit is quite natural and occurs in the work of Lange and Zeugmann [15] as well as in the work of Case, Jain and Suraj [5]. $W_{e,s}$ denotes the set of $x < s$ which are enumerated into W_e within s steps. Similarly, $A_{e,s}$ denotes the set of $x < s$ which are enumerated into A_e within s steps.

A *text* T is a member of $(\mathbb{N} \cup \{\#\})^\infty$. $T(0)$, $T(1)$, $T(2)$ and so on denote the members of T; $T[n]$ denotes $T(0)T(1)\ldots T(n-1)$. A *sequence* σ is a member of $(\mathbb{N} \cup \{\#\})^*$. λ denotes the empty sequence. For a text T let $\text{content}(T) = \{T(n) : n \in \mathbb{N} \wedge T(n) \in \mathbb{N}\}$; similarly one defines $\text{content}(\sigma)$. The length of a sequence σ, denoted $|\sigma|$, is the number of elements in the domain of σ. One says that $\sigma \preceq T$ and $\sigma \preceq \tau$ iff σ is a prefix of T and τ, respectively. Furthermore, T is *a text for* L iff $L = \text{content}(T)$. Note that there is a uniformly recursive method for generating a text T_e for W_e; this text T_e is called a *canonical text* for W_e.

The general model of learning is that the learner M assigns, to every prefix $T[n]$ of a given text T for the set L to be learnt, an index $M(T[n])$ interpreted as M's conjecture for the language L; for finite learning, the learner M is allowed to output a special symbol "?" which denotes that the learner does not wish to make a conjecture at this point. One says that a learner M converges on a text T to an index e (denoted $M(T) = e$) iff $M(T[n]) = e$ for almost all n. Furthermore, one says that M outputs an index e on T iff there is an n with $M(T[n]) = e$. The following definition gives various criteria of learning.

Definition 1 (Bārzdins [2], Case [4], Gold [10], Osherson and Weinstein [20]). A learner M *finitely* learns a language L using a numbering $A_0, A_1, A_2, \ldots$ as hypothesis space [10] iff for every text T for L, M on T outputs exactly one index e, besides ?, and this index e satisfies $A_e = L$.

A learner M *explanatorily* learns a language L using a numbering $A_0, A_1, A_2, \ldots$ as hypothesis space [10] iff for every text T for L, M converges on T to an index e such that $A_e = L$.

A learner M *vacillatorily* learns a language L using a numbering $A_0, A_1, A_2, \ldots$ as hypothesis space [4] iff for every text T for L, M converges on T to an index d such that, for some $e \leq d$, $A_e = L$.

A learner M *behaviourally correctly* learns a language L using a numbering $A_0, A_1, A_2, \ldots$ as hypothesis space [2, 20] iff for every text T for L, $A_{M(T[n])} = L$ for almost all n. Note that it is permitted, but not required, that the $M(T[n])$ are syntactically different.

A learner M *finitely (explanatorily, vacillatorily, behaviourally correctly)* learns S iff it finitely (explanatorily, vacillatorily, behaviourally correctly) learns each $L \in S$.

A class S is *finitely (explanatorily, vacillatorily, behaviourally correctly)* learnable iff some learner finitely (explanatorily, vacillatorily, behaviourally correctly) learns S.

Note that the definition of vacillatorily learnable, as defined by Case [4], requires the learner to eventually output its conjecture only from finitely many correct indices for the input language

— that is, the learner eventually vacillates among only finitely many correct indices for the input language. The definition used above is equivalent to this definition and is a useful characterization of vacillatory learning (this follows using Proposition 16 in [11]). We defined vacillatory learning using this characterization mainly because of its ease of use in the proofs below.

For ease of notation below, if the hypothesis space is not specified, then the default numbering $W_0, W_1, W_2, \ldots$ is assumed as hypothesis space.

Definition 2 (Blum and Blum [3], Fulk [9]). A *stabilizing sequence* for M on L is a sequence σ such that $\mathrm{content}(\sigma) \subseteq L$ and $M(\sigma\tau) = M(\sigma)$ for all $\tau \in (L \cup \{\#\})^*$. A *locking sequence* for M on L is a stabilizing sequence σ for M on L such that $M(\sigma)$ is an index for L (in the hypothesis space used).

Note that by the locking-sequence hunting construction [3, 9] there is a recursive enumeration of learners $M_0, M_1, M_2, \ldots$ such that (a) every explanatorily learnable class is learnt by one of these learners, (b) whenever M_e converges on some text for L, then M_e converges on all texts for L to the same index, (c) whenever M_e explanatorily learns L and T is a text for L, then for some n, $T[n]$ is a locking sequence for M_e on L.

Definition 3. A numbering $A_0, A_1, A_2, \ldots$ is called *optimal for explanatory learning* iff every explanatorily learnable class can be learnt using the numbering $A_0, A_1, A_2, \ldots$ as hypothesis space.

A numbering $A_0, A_1, A_2, \ldots$ is called *effectively optimal for explanatory learning* iff for every numbering $B_0, B_1, B_2, \ldots$, given a learner M, we can effectively find a learner M' such that if M witnesses explanatory learnability of a class S of languages using the numbering $B_0, B_1, B_2, \ldots$ as hypothesis space, then M' witnesses explanatory learnability of S using the numbering A_0, $A_1, A_2, \ldots$ as hypothesis space.

Similarly, one can also define optimality and effective optimality for other learning criteria.

As $W_0, W_1, W_2, \ldots$ is acceptable, for showing (effective) optimality of $A_0, A_1, A_2, \ldots$, it is sufficient to consider converting learners using $W_0, W_1, W_2, \ldots$ as hypothesis space to learners using A_0, $A_1, A_2, \ldots$ as hypothesis space.

Let C be the plain Kolmogorov complexity [16] and C^K be the Kolmogorov complexity relative to K. That is, for a fixed universal Turing Machine U, $C(x)$ is the length of the smallest string y such that $U(y) = x$, and $C^K(x)$ is the length of the smallest string y such that $U^K(y) = x$. Here U^K denotes computations using the oracle K for the halting problem. Note that C^K can be approximated from above relative to K. Let C^K_s denote an approximation of C^K relative to K. Approximations from above or from below relative to K have an easy characterization: A function g can be approximated from above (below) relative to K iff there are uniformly recursive functions g_s with $g(n) = \limsup_s g_s(n)$ for all n (respectively, $g(n) = \liminf_s g_s(n)$ for all n).

$\langle \cdot, \cdot \rangle$ denotes a recursive bijection from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$. Here we assume that $\langle \cdot, \cdot \rangle$ is increasing in both its arguments.

3 Optimality and Effective Optimality

The following notion generalizes the notion of acceptable numberings; it is an example of a natural class of numberings which goes beyond acceptable numberings but is still optimal for most of the learning criteria studied in the literature.

Definition 4. A numbering $A_0, A_1, A_2, \ldots$ is called *nearly acceptable* iff there is a recursive function f such that $A_{f(d,e)} = W_e$ whenever $d \in W_e$.

Proposition 5. *Let $A_0, A_1, A_2, \ldots$ be given by the equations*

$$A_0 = \emptyset;$$
$$A_{\langle d,e\rangle+1} = W_e \cup \{d\}.$$

The numbering $A_0, A_1, A_2, \ldots$ is nearly acceptable but not acceptable.

Furthermore, every nearly acceptable numbering is optimal for finite learning and effectively optimal for explanatory, vacillatory and behaviourally correct learning.

Proof. The numbering $A_0, A_1, A_2, \ldots$ is not acceptable as the Theorem of Rice [21] does not hold. In particular, the index set of $\emptyset$ is the recursive set $\{0\}$.

However, $A_0, A_1, A_2, \ldots$ is nearly acceptable via the function $(d,e) \mapsto \langle d,e\rangle + 1$: if $d \in W_e$, then $A_{\langle d,e\rangle+1} = W_e \cup \{d\} = W_e$.

Now assume that $B_0, B_1, B_2, \ldots$ is nearly acceptable and that this fact is witnessed by f. Let u be a fixed index of the empty set: $B_u = \emptyset$. For the criteria of explanatory or behaviourally correct learning, let a learner M be given. The new learner N is defined as

$$N(\sigma) = \begin{cases} u, & \text{if } \mathrm{content}(\sigma) = \emptyset; \\ f(d, M(\sigma)), & \text{if } d = \min(\mathrm{content}(\sigma)). \end{cases}$$

This learner clearly identifies $\emptyset$. Furthermore, if L is not empty and belongs to the class S learnt by M, then, for almost all n, $N(T[n]) = f(\min(L), M(T[n]))$. Thus, if $W_{M(T[n])} = L$, then $A_{N(T[n])} = L$. Furthermore, if M converges on a text T, then so does N. Hence $B_0, B_1, B_2, \ldots$ is effectively optimal for behaviourally correct and explanatory learning. Furthermore, by the implication in Theorem 6 below, $B_0, B_1, B_2, \ldots$ is also effectively optimal for vacillatory learning.

For finite learning, one has to do a case distinction. If $S = \{\emptyset\}$, then the new learner N outputs always the index u of the empty set. If $\emptyset \notin S$, then the new learner N waits for an element d to show up in the input and for M to output a hypothesis e; once d, e are known, N conjectures $f(d,e)$ and does not revise this hypothesis. The verification that this works is straightforward. □

The effectively optimal numberings for finite, explanatory and vacillatory learning are easy to characterize.

Theorem 6. *A numbering $A_0, A_1, A_2, \ldots$ of all r.e. sets is*
(a) *effectively optimal for finite learning iff it is acceptable;*
(b) *effectively optimal for explanatory learning iff it is K-acceptable;*
(c) *effectively optimal for vacillatory learning iff there is a limit-recursive function g such that, for all d, there is an $e \leq g(d)$ with $A_e = W_d$.*

Proof. We first consider necessity.

Suppose T_d is the canonical text for W_d. Let M^d be a learner which always conjectures d on any input. Then clearly M^d finitely, explanatorily and vacillatorily learns $\{W_d\}$ using $W_0, W_1, W_2, \ldots$ as hypothesis space.

Let M^d_{fin} (M^d_{ex}, M^d_{vac}) be a learner obtained effectively from M^d (and thus from d) to finitely (explanatorily, vacillatorily) learn $\{W_d\}$ using $A_0, A_1, A_2, \ldots$ as hypothesis space.

(a) For all d, M^d_{fin} outputs exactly one hypothesis on T_d, which is a grammar for W_d in the numbering $A_0, A_1, A_2, \ldots$; thus, there exists a recursive function h which maps d to the hypothesis output by M^d_{fin} on T_d. This recursive function witnesses that $A_0, A_1, A_2, \ldots$ is acceptable.

(b) For all d, M^d_{ex} converges on T_d to a grammar for W_d in the numbering $A_0, A_1, A_2, \ldots$; thus, there exists a limiting recursive function h which maps d to the hypothesis output by M^d_{ex}, in the limit, on T_d. This limiting recursive function h witnesses that $A_0, A_1, A_2, \ldots$ is K-acceptable.

(c) For all d M^d_{vac} converges on T_d to an upper bound on the grammar for W_d in the numbering $A_0, A_1, A_2, \ldots$; thus, there exists a limiting recursive function g which maps d to the hypothesis output by M^d_{vac}, in the limit, on T_d. This limiting recursive function g witnesses that (c) holds.

We now consider sufficiency.

(a) Suppose h is a recursive function which witnesses that $A_0, A_1, A_2, \ldots$ is acceptable. Then, given a finite learner M using $W_0, W_1, W_2, \ldots$ as hypothesis space, one can construct M' as follows. $M'(T[n]) = h(M(T[n]))$. It is easy to verify that M' finitely identifies each language which is finitely identified by M.

(b) Suppose h is a K-recursive function which witnesses that $A_0, A_1, A_2, \ldots$ is K-acceptable. Let N be an oracle Turing Machine which computes h using oracle K. Then, given an explanatory learner M using $W_0, W_1, W_2, \ldots$ as the hypothesis space, one can construct M' as follows. Let $K_0, K_1, \ldots$ be a recursive approximation to K. $M'(T[n]) = N^{K_n}(M(T[n]))$, if $N^{K_n}(M(T[n]))$ halts within n steps; $M'(T[n]) = 0$ otherwise. It is easy to verify that M' explanatorily learns each language which is explanatorily learnt by M.

(c) Suppose g is as given in the hypothesis. Let g' be a recursive function such that $g(x) = \lim_{t \to \infty} g'(x, t)$. Given a vacillatory learner M using $W_0, W_1, W_2, \ldots$ as the hypothesis space, one can construct M' as follows. $M'(T[n]) = \max(\{g'(i, n) : i \leq M(T[n])\})$. Suppose $M(T)$ converges. Then, $M'(T) = \max(\{g(i) : i \leq M(T)\})$. Thus, if exists an $i \leq M(T)$ such that $W_i = \mathrm{content}(T)$, then, there exists an $i' \leq g(i) \leq M'(T)$ such that $A_{i'} = \mathrm{content}(T)$. Thus, M' vacillatorily learns each language which is vacillatorily learnt by M. $\Box$

We now turn our attention to the separation of effectively and non-effectively optimal numberings, as well as the separation of optimal numberings for various criteria of inference. The following propositions are useful for showing some of our results.

Proposition 7. *If S is a finitely learnable class, then there is a number d such that almost all members of S have at least 2 non-elements below d.*

Proof. Suppose M finitely learns S. Fix an $L \in S$. Let σ be such that content$(\sigma) \subseteq L$ and $M(\sigma)$ is an index for L. Let $d_1 = \max(\text{content}(\sigma))$. Note that for $L' \in S$ with $L \neq L'$, content$(\sigma) \not\subseteq L'$. The reason is that otherwise M does not finitely learn $\{L, L'\}$. Thus, for all $L' \in S - \{L\}$, there exists an $i \leq d_1$, such that $i \notin L'$.

Let $S_i = \{L' \in S : i \notin L'\}$. For non-empty S_i, let L_i be a fixed member of S_i and let σ_i be such that content$(\sigma_i) \subseteq L_i$ and $M(\sigma_i)$ is an index for L_i. Let $d_2 = \max(\{\max(\text{content}(\sigma_i)) : i \leq d_1 \wedge S_i \neq \emptyset\})$. Note that if $S_i \neq \emptyset$, then for $L' \in S$ with $L_i \neq L'$, content$(\sigma_i) \not\subseteq L'$; thus, for all $L' \in S - \{L_i\}$, there exists a $j \leq d_2$, such that $j \neq i$ and $j \notin L'$.

Thus, for all $L' \in S - (\{L\} \cup \{L_i : S_i \neq \emptyset, i \leq d_1\})$, the set $(\mathbb{N} - L') \cap \{x : x \leq d_1 + d_2\}$ contains at least two elements. $\square$

For any n and n distinct numbers $a_1, a_2, \ldots, a_n$, let $D_e = \{a_1, a_2, \ldots, a_n\}$ iff $e = 2^{a_1} + 2^{a_2} + \ldots + 2^{a_n}$; furthermore, let $D_0 = \emptyset$. The number e is called the canonical index of D_e. Recall that C^K is the Kolmogorov complexity relative to K.

Proposition 8. *Let a uniformly K-recursive one-one listing $L_0, L_1, L_2, \ldots$ of cofinite sets be given such that $i = \min(\mathbb{N} - L_i)$ for all i. Then there is a numbering $H_0, H_1, H_2, \ldots$ of r.e. sets and a (non-recursive) function g such that for all i, j:*

- $\forall i > 0\, [D_i \subseteq H_i \subseteq D_i \cup \{\max(D_i), \max(D_i) + 1, \max(D_i) + 2, \ldots\}]$;
- $\forall i\, [H_{g(i)} = L_i]$;
- $\forall i\, [i \notin \{g(0), g(1), g(2), \ldots\} \Rightarrow H_i$ *is finite*$]$;
- $\forall i, j\, [\max(\{j : C^K(j) \leq 2^i\}) < g(i)]$.

The function g can be approximated from below relative to K.

Proof. The function g is defined by the following K-recursive approximation:

- in stage 0: choose $g_0(i)$ such that $D_{g_0(i)} = \{0, 1, 2, \ldots, i, i+1\} - \{i\}$;
- in stage $s+1$: if there is an x with
  - $[\max(D_{g_s(i)}) \leq x \leq s$ and $x \notin L_i]$ or
  - $[g_s(i) \leq x \leq s$ and $C_s^K(x) \leq 2^i]$

  then choose $g_{s+1}(i)$ such that $D_{g_{s+1}(i)} = \{s+1\} \cup (L_i \cap \{0, 1, 2, \ldots, s\})$
  else let $g_{s+1}(i) = g_s(i)$.

Note that whenever $g_{s+1}(i) \neq g_s(i)$, then $\max(D_{g_{s+1}(i)}) > s$ and hence $g_{s+1}(i) > s$. It follows that the set $G = \mathbb{N} - \{g(0), g(1), g(2), \ldots\}$ is K-r.e.; that is, there is a recursive approximation with $i \in G \Leftrightarrow \forall^\infty s\, [i \in G_s]$. Let $H_0 = \emptyset$; for $i > 0$, let

$$H_i = D_i \cup \{t : \exists s\, [\max(D_i) \leq t \leq s \wedge i \notin G_s]\}.$$

The sets $H_0, H_1, H_2, \ldots$ are uniformly r.e.; furthermore, H_i is finite iff $i \in G_s$ for almost all s. In the case that $i = g(j)$ it follows that $\max(D_i)$ is an upper bound on all non-elements of L_j and therefore $H_i = L_j$. This completes the proof. $\square$

Theorem 9. *There is a numbering which is optimal but not effectively optimal for finite learning. This numbering is not optimal for explanatory, vacillatory and behaviourally correct learning.*

Proof. Let $L_e = \mathbb{N} - \{e\}$ for all e; then choose the numbering $H_0, H_1, H_2, \ldots$ according to Proposition 8. Now let $A_{\langle 0,0\rangle} = \mathbb{N}$ and $A_{\langle 0,e+1\rangle} = H_e$ for all e. Furthermore, for every e and every $d > 0$, let

$$A_{\langle d,e\rangle} = \bigcup_{s:|\{0,1,2,\ldots,d\}-W_{e,s}|\geq 2} W_{e,s}.$$

Note that the resulting numbering covers all r.e. sets: first $\mathbb{N}$ and every set of the form $\mathbb{N} - \{a\}$ is covered by sets of the form $A_{\langle 0,e\rangle}$; second, a set W_e with least non-elements a, b is equal to $A_{\langle d,e\rangle}$ for all $d > a + b$.

Now let S be a finitely learnable class with learner M. Note that if $\mathbb{N} \in S$, then S contains no other languages and thus finite learnability in numbering A will be trivial. So assume $\mathbb{N} \notin S$. By Proposition 7 there is a number d such that all but finitely many members of S have at least 2 non-elements below d. Without loss of generality, d is so large that these exceptions are all of the form $\mathbb{N} - \{c\}$ with $c \leq d$. Now one builds a new learner N as follows:

- $N(\sigma)$ is an index $\langle 0, e_c\rangle$ for the set $\mathbb{N} - \{c\}$ whenever $\{0, 1, 2, \ldots, d\} - \{c\} \subseteq \text{content}(\sigma) \subseteq \mathbb{N} - \{c\}$ and $\mathbb{N} - \{c\} \in S$.
- $N(\sigma) = \langle d, M(\sigma)\rangle$ whenever $M(\sigma)$ is defined (that is, $M(\sigma) \neq ?$) and $c \in \text{content}(\sigma)$ for all c with $\mathbb{N} - \{c\} \in S$.
- $N(\sigma) = ?$, otherwise.

It is easy to see that N is a finite learner for S.

Note that, by the definition of $H_0, H_1, H_2, \ldots$ and $A_0, A_1, A_2, \ldots$, each of the sets $\mathbb{N} - \{c\}$ have exactly one index $\langle 0, e_c\rangle$ (with respect to $A_0, A_1, A_2, \ldots$), which also satisfies $C^K(e_c) > 2^c$. It follows that the class $\{\mathbb{N} - \{c\} : c \in \mathbb{N}\}$ is not behaviourally correctly learnable using A_0, $A_1, A_2, \ldots$ as hypothesis space, as otherwise $C^K(e_c)$ will, for every c, be bounded by c plus a constant independent of c. As $\{\mathbb{N} - \{c\} : c \in \mathbb{N}\}$ is explanatorily learnable, it follows that A_0, $A_1, A_2, \ldots$ is not optimal for explanatory, vacillatory and behaviourally correct learning.

Furthermore, $A_0, A_1, A_2, \ldots$ is not acceptable as $A_0, A_1, A_2, \ldots$ contains only one index for each set of the form $\mathbb{N} - \{c\}$. Thus, by Theorem 6, $A_0, A_1, A_2, \ldots$ is not effectively optimal for finite learning. □

Theorem 10. *There is a numbering $A_0, A_1, A_2, \ldots$ which is effectively optimal for vacillatory learning but not optimal for explanatory learning.*

Proof. Let $C_s^{K_s}$ be an approximation of C^K after s steps such that, for all x, $C^K(x) = \limsup C_s^{K_s}(x)$ and for all s and c, there are less than 2^c numbers y with $C_s^{K_s}(y) < c$. Now let

$$A_{\langle d,e\rangle} = \bigcup_{s:C_s^{K_s}(d)>2^e} W_{e,s}.$$

Then $A_{\langle d,e\rangle}$ is finite for those d and e where $C^K(d) \leq 2^e$. Furthermore, for every e and all sufficiently large d it holds that $C^K(d) > 2^e$.

No class S containing infinitely many infinite sets is explanatorily learnable using this numbering. The reason is that given the least index e of an infinite member of the class, the learner will converge on the canonical text of W_e to an index of Kolmogorov complexity (relative to K) at most a constant above that of e; however every index in the given numbering A for W_e will have Kolmogorov complexity (relative to K) at least 2^e minus a constant. Hence such an explanatory learner cannot exist. As there exist explanatorily learnable classes (such as $\{\{\langle e,x\rangle : x \in \mathbb{N}\} : e \in \mathbb{N}\}$) containing infinitely many infinite sets, it follows that $A_0, A_1, A_2, \ldots$ is not optimal for explanatory learning.

On the other hand, one can use Theorem 6 to obtain that the numbering considered is effectively optimal for vacillatory learning: the reason is that for every e there is a $d \leq 2^{e+1}$ with $C^K(d) > 2^e$ and $A_{\langle d,e\rangle} = W_e$. □

Theorem 11. *There is a numbering $A_0, A_1, A_2, \ldots$ which is effectively optimal for explanatory and vacillatory learning but not optimal for behaviourally correct learning or finite learning.*

Proof. Recall that a simple set [21] is r.e., co-infinite and intersects every infinite r.e. set. Let $a_0, a_1, a_2, \ldots$ be a recursive one-one enumeration of a simple set B and define

$$A_{\langle d,e\rangle} = \begin{cases} W_e, & \text{if } d \notin B \wedge d > e; \\ \{0,1,2,\ldots,2^{s+1} \cdot 3^d \cdot 5^e\}, & \text{if } d = a_s \wedge d > e; \\ \{0,1,2,\ldots,3^d \cdot 5^e\}, & \text{if } d \leq e. \end{cases}$$

This numbering is a K-acceptable numbering as, for every e, one can find the least $d \notin B \cup \{0,1,2,\ldots,e\}$ using the oracle K and then $A_{\langle d,e\rangle} = W_e$. By Theorem 6, the numbering is effectively optimal for explanatory and vacillatory learning.

It remains to show that the numbering is not optimal for behaviourally correct learning or finite learning.

Consider any class S of infinite languages which is behaviourally correctly learnable but not vacillatorily learnable using $W_0, W_1, W_2, \ldots$ as a hypothesis space [4]. Suppose that M, using the numbering $A_0, A_1, A_2, \ldots$ as hypothesis space, behaviourally correctly learns S. As S is not vacillatorily learnable, it follows by a result of Case [4] that there are $L \in S$ and a recursive text T for L on which the learner M outputs infinitely many distinct conjectures. For every pair $\langle d,e\rangle$ with $d \in \{0,1,2,\ldots,e\} \cup B$, the set $A_{\langle d,e\rangle}$ is a finite set and has a maximum which is a multiple of $3^d \cdot 5^e$. Hence M outputs only finitely many of these pairs on the text T. Now let E be the infinite r.e. set of all indices $\langle d,e\rangle$ output by M on T such that $d \notin \{0,1,2,\ldots,e\} \cup B$. The set $\{d : \exists e\,[\langle d,e\rangle \in E]\}$ is an r.e. set disjoint to B and hence finite. As for every e there are only pairs $\langle d,e\rangle$ with $d > e$ in E, it follows that E is finite as well in contradiction to the assumption.

From this contradiction it can be concluded that M is not a behaviourally correct learner for S and the numbering $A_0, A_1, A_2, \ldots$ is not optimal for behaviourally correct learning.

Consider $L_n = \{\langle n,x\rangle : x \in \mathbb{N}\}$. Clearly, $\{L_0, L_1, L_2, \ldots\}$ is finitely learnable using $W_0, W_1,$

$W_2, \ldots$ as hypothesis space. Suppose by way of contradiction that some learner finitely learns $\{L_0, L_1, L_2, \ldots\}$ using $A_0, A_1, A_2, \ldots$ as hypothesis space. Then, given n, one can effectively find an index $\langle d_n, e_n \rangle$ such that $A_{\langle d_n, e_n \rangle} = L_n$. In particular, $d_n \notin B$, $d_n > e_n$ and $W_{e_n} = L_n$. Note that all e_n are distinct. But then the set $\{d_n : n \in \mathbb{N}\}$ is an infinite r.e. set disjoint from B, a contradiction to B being a simple set. Thus, $\{L_0, L_1, L_2, \ldots\}$ is not finitely learnable using $A_0, A_1, A_2, \ldots$ as hypothesis space. □

Theorem 12. *The numbering $A_0, A_1, A_2, \ldots$ given by*

$$A_{\langle d,e \rangle} = \bigcup_{s : \exists m \, [m = \min(W_{e,s}) \wedge (d > |W_{m,s}| \vee |W_{e,s}| \leq |W_m|)]} W_{e,s}$$

is effectively optimal for behaviourally correct learning but not optimal for any of finite, explanatory or vacillatory learning.

Proof. The behaviourally correct learner N using $A_0, A_1, A_2, \ldots$ is effectively built by simulating a given learner M using the numbering $W_0, W_1, W_2, \ldots$ as hypothesis space and defining $N(\sigma) = \langle |\sigma|, M(\sigma) \rangle$. Given a text T for a set L learnt by M, use e_d as shorthand for $M(T[d])$ and note that $N(T[d]) = \langle d, e_d \rangle$. The learner N succeeds as shown in the following case distinction.

- $L = \emptyset$: then almost all e_d are indices of the empty set and hence $A_{\langle d, e_d \rangle}$ is empty for almost all d as well.
- $m = \min(L)$ exists and W_m is infinite: then $A_{\langle d, e_d \rangle} = W_{e_d}$ for all d where W_{e_d} is correct. Hence N behaviourally correctly learns L as well.
- $m = \min(L)$ exists and W_m is finite: then $A_{\langle d, e_d \rangle} = W_{e_d}$ for all d where W_{e_d} is correct and $d > |W_m|$. Hence N behaviourally correctly learns L as well.

Let p_n be the n-th prime number and let $L_n = \{n, p_n, p_n^2, p_n^3, p_n^4, p_n^5, \ldots\}$. Note that $p_n > n$ and L_n is the only set in $L_0, L_1, L_2, \ldots$ containing $\{n, p_n^m\}$ and $\{p_n^m, p_n^k\}$ as subsets for any different numbers m, k; hence one can identify L_n from any two of its elements and the class $\{L_0, L_1, L_2, \ldots\}$ is finitely learnable in any acceptable numbering. However $\{L_0, L_1, L_2, \ldots\}$ is not vacillatorily learnable in the numbering $A_0, A_1, A_2, \ldots$ — otherwise, for any n, one can produce a canonical text for L_n and then we will have that the largest hypothesis output by the learner on this text is an upper bound for $|W_n|$, whenever W_n is finite; this contradicts the fact that finiteness of r.e. sets cannot be decided in the limit. □

Theorem 13. *There are numberings $A_0, A_1, A_2, \ldots$ and $B_0, B_1, B_2, \ldots$ with the following properties.*

(a) *Both numberings are optimal for explanatory learning.*

(b) *Both numberings are neither effectively optimal for explanatory nor effectively optimal for vacillatory learning.*

(c) *Both numberings are not optimal for behaviourally correct learning.*

(d) *The numbering $A_0, A_1, A_2, \ldots$ is not optimal for vacillatory learning.*

(e) *The numbering $B_0, B_1, B_2, \ldots$ is optimal for vacillatory learning.*

Proof. The numberings $A_0, A_1, A_2, \ldots$ and $B_0, B_1, B_2, \ldots$ are obtained using two different versions of a K-recursive listing $L_0, L_1, L_2, \ldots$ such that

(PA) $\{\langle n, x\rangle : x \in L_n\} \leq_T K$;
(PB) $n = \min(\mathbb{N} - L_n)$ for all n;
(PC) each set L_n has at most $n + 1$ non-elements;
(PD) the class $\{L_0, L_1, L_2, \ldots\}$ has no infinite explanatorily learnable subclass.

The difference between these two numberings is that in the case of $B_0, B_1, B_2, \ldots$, the class $\{L_0, L_1, L_2, \ldots\}$ used has no infinite vacillatorily learnable subclass while in the case of $A_0, A_1, A_2, \ldots$, the class $\{L_0, L_1, L_2, \ldots\}$ itself is an infinite vacillatorily learnable class.

Now let $L_{m,s}(x)$ be a recursive approximation of $L_m(x)$ using s steps and let $W_{e,s}$ the set of all $x \leq s$ which are enumerated into W_e within s steps. It is assumed that the approximation also satisfies

$$\forall m\, \forall s\, \forall x \leq m\, [L_{m,s}(x) = L_m(x)].$$

The numbering $A_0, A_1, A_2, \ldots$ is built from the $H_0, H_1, H_2, \ldots$ assigned to $L_0, L_1, L_2, \ldots$ in Proposition 8 as follows:

$$A_{\langle d,e\rangle} = \begin{cases} \mathbb{N}, & \text{if } d = 0 \text{ and } e = 0; \\ H_{e-1}, & \text{if } d = 0 \text{ and } e > 0; \\ \bigcup_{s : \forall m \leq s\, \exists x \leq d\, [L_{m,s}(x) \neq W_{e,s}(x)]} W_{e,s}, & \text{if } d > 0. \end{cases}$$

Note that $A_0, A_1, A_2, \ldots$ is a universal numbering. In the case of $B_0, B_1, B_2, \ldots$, the only difference is that the parameter list $L_0, L_1, L_2, \ldots$ is chosen differently. Furthermore, let g denote the function g corresponding to $H_0, H_1, H_2, \ldots$ from Proposition 8.

(a): Now assume that a class S is explanatorily learnable using $W_0, W_1, W_2, \ldots$; it is shown that S is also explanatorily learnable using $A_0, A_1, A_2, \ldots$ and $B_0, B_1, B_2, \ldots$ as hypothesis space. The choice of $L_0, L_1, L_2, \ldots$ is not yet fixed, but only the properties (PA) to (PD) are used. Hence it is sufficient to show the learnability using $A_0, A_1, A_2, \ldots$ as hypothesis space; the learnability using $B_0, B_1, B_2, \ldots$ follows along the same lines. Assume that M is an explanatory learner for S using $W_0, W_1, W_2, \ldots$ as hypothesis space. Note that this learner can only learn finitely many members of $\{L_0, L_1, L_2, \ldots\}$, as no learner can explanatorily learn infinitely many members of $\{L_0, L_1, L_2, \ldots\}$. Let $I = \{n : L_n \in S\}$. If $n \in I$, then let F_n be a corresponding tell-tale set [1] for L_n, that is, F_n is a finite subset of L_n such that, for all $B \in S - \{L_n\}$, $\neg[F_n \subseteq B \subseteq L_n]$. Furthermore, in the case that $\mathbb{N} \in S$, let E be a corresponding tell-tale set, that is, E is a finite set such that for all $B \in S - \{\mathbb{N}\}$, $\neg[E \subseteq B]$.

Now, a new learner N, using $A_0, A_1, A_2, \ldots$ as hypothesis space, on input σ is defined as follows. Let emp be such that $A_{emp} = \emptyset$.

$$N(\sigma) = \begin{cases} emp, & \text{if } \mathrm{content}(\sigma) = \emptyset; \\ \langle 0, 0 \rangle, & \text{if } E \subseteq \mathrm{content}(\sigma); \\ \langle 0, g(n)+1 \rangle, & \text{if } n \in I \text{ and } [F_n \subseteq \mathrm{content}(\sigma) \subseteq L_n]; \\ \langle d, M(\sigma) \rangle, & \text{otherwise, where, for } m = \min(\mathbb{N} - \mathrm{content}(\sigma)), \\ & d = \min(\{c : c > m + |\sigma| \vee \\ & \qquad c \in (L_{m,|\sigma|} - \mathrm{content}(\sigma)) \vee \\ & \qquad c \in (\mathrm{content}(\sigma) - L_{m,|\sigma|})\}). \end{cases}$$

The learner N is recursive. It is clear that N learns all sets in $\{\emptyset, \mathbb{N}, L_0, L_1, L_2, \ldots\} \cap S$ using the first three cases.

Let T be a text of a set $L \in S - \{\emptyset, \mathbb{N}, L_0, L_1, L_2, \ldots\}$. If the initial segment σ of T currently processed by N is sufficiently large, then $e = M(\sigma)$ is the hypothesis to which M converges on T, the value m in the above algorithm is the least non-element of L and d is the least number with the property that $L(d) \neq L_m(d)$ (note that $d > m$). Thus N converges on T to $\langle d, e \rangle$. Furthermore, for all $n \neq m$, $W_e \cap \{0, 1, 2, \ldots, m\} \neq L_n \cap \{0, 1, 2, \ldots, m\}$. Thus, $W_e \cap \{0, 1, 2, \ldots, d\} \neq L_n \cap \{0, 1, 2, \ldots, d\}$ for all n. It follows that $A_{\langle d,e \rangle} = W_e$. Hence N explanatorily learns S using the numbering $A_0, A_1, A_2, \ldots$ as hypothesis space. It follows that $A_0, A_1, A_2, \ldots$ is optimal for explanatory learning.

(b): Now it is shown that $A_0, A_1, A_2, \ldots$ and $B_0, B_1, B_2, \ldots$ are not effectively optimal for explanatory or vacillatory learning. Note that if $A_0, A_1, A_2, \ldots$ is effectively optimal for either explanatory or vacillatory learning (or both), then it follows from Theorem 6 that there is a K-recursive function h such that

$$\forall d \, \exists e \leq h(d) \, [A_e = \mathbb{N} - D_d].$$

It is now shown that this property will lead to a contradiction. Suppose n and the cardinality $m = |\mathbb{N} - L_n|$ are given. Recall that $m \leq n+1$. Then one can find, using the oracle K, the unique index d with $D_d = \mathbb{N} - L_n$, by searching for these m non-elements. Then one can compute, using the oracle K, the upper bound $h(d)$ of an e with $A_e = \mathbb{N} - D_d$. Due to Kolmogorov complexity arguments, the complexity relative to K of $h(d)$ is at most $c + 2\log(n)$, for some constant c, as one can describe n and m both by two binary numbers having $1 + \log(n)$ bits. But by construction, the only index $\langle 0, g(n) \rangle$ of L_n in $A_0, A_1, A_2, \ldots$ has a second component, which is larger than all numbers with Kolmogorov complexity at most 2^n; a contradiction. It follows that $A_0, A_1, A_2, \ldots$ is neither effectively optimal for explanatory nor effectively optimal for vacillatory learning. Similarly, $B_0, B_1, B_2, \ldots$ is neither effectively optimal for explanatory nor effectively optimal for vacillatory learning.

(c): Now it is shown that the numberings $A_0, A_1, A_2, \ldots$ ($B_0, B_1, B_2, \ldots$) are not optimal for behaviourally correct learning. This can be seen as follows.

The numbering $A_0, A_1, A_2, \ldots$ has exactly one index for each set L_n. Hence every behaviourally correct learner for $\{L_0, L_1, L_2, \ldots\}$ using $A_0, A_1, A_2, \ldots$ as hypothesis space is also explanatorily learning $\{L_0, L_1, L_2, \ldots\}$ using $A_0, A_1, A_2, \ldots$ as hypothesis space. As $\{L_0, L_1, L_2, \ldots\}$ is not explanatorily learnable, $\{L_0, L_1, L_2, \ldots\}$ is not behaviourally correctly learnable using A_0, $A_1, A_2, \ldots$ as hypothesis space.

On the other hand, the class $\{L_0, L_1, L_2, \ldots\}$ is behaviourally correctly learnable using W_0, $W_1, W_2, \ldots$ as hypothesis space. To see this consider a learner which, on sequence $T[s]$, outputs an index (in $W_0, W_1, W_2, \ldots$) for $\bigcup_{t>s} L_{n,t}$, where n is the minimal element not in content$(T[s])$. Note that on any text T for L_m, for all but finitely many s, the n found as above is m. Furthermore, for all but finitely many s, $\bigcup_{t>s} L_{n,t} = L_n$ — as $L_{n,t}$ converge pointwise to L_n, for sufficiently large t, $L_{n,t}$ does not contain any of the finitely many non-elements of L_n, whereas every element of L_n is contained in almost all $L_{n,t}$.

Thus, $\{L_0, L_1, L_2, \ldots\}$ is behaviourally correctly learnable but not using $A_0, A_1, A_2, \ldots$ as hypothesis space. It follows that $A_0, A_1, A_2, \ldots$ is not an optimal numbering for behaviourally correct learning. It can be similarly shown that $B_0, B_1, B_2, \ldots$ is not optimal for behaviourally correct learning.

(d): Now it is shown that one can choose $L_0, L_1, L_2, \ldots$ such that the resulting numbering A_0, $A_1, A_2, \ldots$ is not optimal for vacillatory learning. Let $M_0, M_1, M_2, \ldots$ be a listing of total learners such that every class which is explanatorily learnable is explanatorily learnable by one of these machines using $W_0, W_1, W_2, \ldots$ as hypothesis space. Additionally we assume that for each M_i, for all texts T for a language L which M_i explanatorily learns, there is a prefix of T which is a locking sequence for M_i on L (see [9]).

For ease of notation we use $\diamond$ to denote concatenation of strings. We say that T is a characteristic-text if $T(i) \in \{i, \#\}$ for all i. We say that a sequence σ is a characteristic-sequence if $\sigma(i) \in \{i, \#\}$, for all $i < |\sigma|$. We will now define L_n. The construction below can be easily seen to be uniform in n.

Define a recursive function F_n as follows. For each binary string η of length at most n, $F_n(\eta, t)$ is a characteristic-sequence defined as follows. Let $\sigma_{init} = 0 \diamond 1 \diamond 2 \diamond \ldots \diamond (n-1) \diamond \# \diamond n+1$, be the characteristic-sequence of length $n+2$ with content $\{0, 1, 2 \ldots, n-1, n+1\}$.

For $t \leq n+1$, let $F_n(\eta, t) = \sigma_{init}[t+1]$. For $t = n+2, n+3, n+4, \ldots$, the value $F_n(\eta, t)$ is defined inductively in stage t.

Stage t: Definition of $F_n(\eta, t)$.

- (1) If $\eta = \lambda$:
  - Let $m = |F_n(\eta, t-1)|$.
  - (1.1) If there exists a set $X \subseteq \{i : i < n\}$ with $|X| \geq |\eta| + 1$ such that, for all $i \in X$ and for all $\sigma \in (\text{content}(F_n(\eta, t-1)) \cup \{x : x \geq m\} \cup \{\#\})^*$ with $|\sigma| \leq t$, it holds that $M_i(F_n(\eta, t-1)) = M_i(F_n(\eta, t-1) \diamond \sigma)$
    Then let $F_n(\eta, t) = F_n(\eta, t-1)$
    Else let $F_n(\eta, t) = F_n(\eta, t-1) \diamond m \diamond m+1 \diamond \ldots \diamond t$.
- (2) If $\eta \neq \lambda$:
  - Let $\eta = \beta a$, where $a \in \{0, 1\}$.
  - (2.1) If $|F_n(\beta, t)| = t+1$, then let $F_n(\eta, t) = F_n(\beta, t)$.
  - (2.2) If $|F_n(\beta, t)| = t$, then define $F_n(\eta, t) = F_n(\beta, t) \diamond w$, where $w = t$, if $a = 1$, and $w = \#$, otherwise.

– (2.3) If $|F_n(\beta,t)| < t$:
  – Let $m = |F_n(\eta, t-1)|$.
  – (2.3.1) If there exists a set $X \subseteq \{i : i < n\}$ with $|X| \geq |\eta| + 1$ such that for all $i \in X$, for all $\sigma \in (\text{content}(F_n(\eta, t-1)) \cup \{x : x \geq m\} \cup \{\#\})^*$ with $|\sigma| \leq t$, $M_i(F_n(\eta, t-1)) = M_i(F_n(\eta, t-1) \diamond \sigma)$
    Then let $F_n(\eta, t) = F_n(\eta, t-1)$
    Else let $F_n(\eta, t) = F_n(\eta, t-1) \diamond m \diamond m+1 \diamond \ldots \diamond t$.

End Stage t.

The following claim follows easily by induction on the stages.

Claim 14. *The following hold for all η of length at most n and all t.*
(i) *$F_n(\eta, t+1) = F_n(\eta, t)$ or $F_n(\eta, t+1)$ is of length $t+2$.*
(ii) *$F_n(\eta, t)$ is a string of length at most $t+1$.*
(iii) *$F_n(\eta, t) \subseteq F_n(\eta a, t)$, for $a \in \{0, 1\}$, and η of length $< n$.*
(iv) *content$(F_n(\eta, t)) \subseteq$ content$(F_n(\eta, t+1))$.*
(v) *If $F_n(\eta, t') = F_n(\eta, t)$, for all $t' > t$, then $F_n(\eta, t)$ is a stabilizing sequence for at least $|\eta|+1$ machines among $M_0, M_1, M_2, \ldots, M_{n-1}$ on content$(F_n(\eta, t)) \cup \{x : x \geq |F_n(\eta, t)|\}$.*
(vi) *If $F_n(\eta, t) = F_n(\eta, t+1)$, then for all $t' \geq t$, either $F_n(\eta, t') = F_n(\eta, t)$ or $F_n(\eta, t')(t+1) = t+1$.*

Properties (i) – (v) are easy to verify. We can show (vi) by induction on length of η. Note that if $F_n(\eta, t+1) = F_n(\eta, t)$ and $F_n(\eta, t') \neq F_n(\eta, t)$, where t' is minimal such number greater than t, then it must be the case that $F_n(\beta, t) = F_n(\beta, t'')$, for all t'' such that $t \leq t'' < t'$ and $\beta \subseteq \eta$. Thus, $F_n(\eta, t')$ is defined via either 1.1 or 2.3.1 (in which case $F_n(\eta, t')(t+1) = t+1$) or $F_n(\eta, t')$ is defined via 2.1 and thus $F_n(\eta, t')(t+1) = F_n(\beta, t')(t+1) = t+1$, where β is the longest proper prefix of η. Note that $F_n(\eta, t')$ cannot be defined via 2.2, as otherwise $F_n(\eta, t'-1)$ must also be different from $F_n(\eta, t)$.

Claim 15. *For all η with $|\eta| \leq n$, for all t, $F_n(\eta, t)$ is $\#$ for at most $|\eta|+1$ inputs.*

Note that # is introduced in $F_n(\eta, t)$ only via step 2.2. or by initialization σ_{init}. It thus follows by induction on length of η that $F_n(\eta, t)$ has at most $(|\eta|+1)$ #s.

Now define $H_{n,\eta}$ to be $\bigcup_{t \in \mathbb{N}} \text{content}(F_n(\eta, t)) \cup \{t+1 : F_n(\eta, t) = F_n(\eta, t+1)\}$. It follows from above claim that one can effectively find an index for $H_{n,\eta}$. It thus follows using Claim 15 that $H_{n,\eta}$ has at most $n+1$ non-elements. Also, by definition of σ_{init}, $\min(\{\mathbb{N} - H_{n,\eta}\}) = n$.

Now we define L_n. L_n will be one of $H_{n,\eta}$, with η a binary string of length at most n. We give below a procedure for defining $L_n(t)$, using the oracle K. Thus, $L_0, L_1, \ldots$ satisfy the requirements (PA), (PB) and (PC). Initially let $\eta = \lambda$ and $Q = \emptyset$. Intuitively, Q will denote the set of machines which have been diagonalized against explicitly (by diagonalizing against the learner's conjecture on a stabilizing sequence for it on L_n).

Stage t: Definition of $L_n(t)$.

– If $|F_n(\eta, t)| < t+1$, then let $L_n(t) = 1$.

- If $|F_n(\eta, t)| = t + 1$, then let $L_n(t) = 1$ if and only if $t \in \text{content}(F_n(\eta, t))$.
- If for all $t' > t$, $F_n(\eta, t') = F_n(\eta, t)$, then
  - (* Here $F_n(\eta, t)$ is a stabilizing sequence for at least $|\eta| + 1$ machines among $M_0, M_1, M_2, \ldots, M_{n-1}$ on $\text{content}(F_n(\eta, t)) \cup \{x : x \geq |F_n(\eta, t)|\}$. Furthermore, t is the point of convergence for $F_n(\eta, \cdot)$. *)
  - Let $j \in \{0, 1, 2, \ldots, n-1\} - Q$ be such that $F_n(\eta, t')$ is a stabilizing sequence for M_j on the set $\text{content}(F_n(\eta, t)) \cup \{x : x \geq |F_n(\eta, t)|\}$.
  - Update η to $\eta \diamond (1 - W_{M_j(F_n(\eta,t))}(t + 1))$.
  - Update Q to $Q \cup \{j\}$.
  - (* Note that we will explicitly diagonalize against M_j, in stage $t + 1$, as for updated η, $F_n(\eta, t + 1)(t + 1)$ is different from $W_{M_j(F_n(\eta,t))}(t + 1)$ — note that $F_n(\eta, t + 1)$, for the updated η, is defined via step 2.2. *)

End Stage t.

Let η, Q be the limiting value for η and Q in the above construction (note that there exists such a limiting value, as $F_n(\beta, \cdot)$, does not converge for all β of length n). It is easy to verify that L_n above is $H_{n,\eta}$.

Now, L_n is not explanatorily learnt using numbering $W_0, W_1, W_2, \ldots$ by all M_j, $j \in Q$ due to explicit diagonalization above. Furthermore, for all $j \in \{0, 1, 2, \ldots, n-1\} - Q$, no prefix of the characteristic text T for L_n is a stabilizing sequence for M_j on L_n (as otherwise, $F_n(\eta, t)$ will converge). It follows that M_j, for $j < n$, do not explanatorily learn L_n using $W_0, W_1, W_2, \ldots$ as hypothesis space. Thus, $L_0, L_1, \ldots$ satisfy property (PD) also.

Furthermore, as $H_{n,\eta}$ is equal to L_n, one has that $H_{n,\beta} = L_n$ for some binary string β of length at most n. Thus, from n, one can effectively find $2^{n+1} - 1$ indices, one of which is an index for L_n. Thus, one can vacillatorily learn $\{L_0, L_1, L_2, \ldots\}$ using $W_0, W_1, W_2, \ldots$ as hypothesis space. However $L_0, L_1, L_2, \ldots$ is not vacillatorily learnable using hypothesis space $A_0, A_1, A_2, \ldots$ as can be proved along the lines of part (c). It follows that $A_0, A_1, A_2, \ldots$ is not optimal for vacillatory learning.

(e): Now it is shown that one can choose $L_0, L_1, L_2, \ldots$ such that the resulting numbering $B_0, B_1, B_2, \ldots$ is optimal for vacillatory learning. We will have that no learner vacillatorily learns more than finitely many languages in $\{L_0, L_1, L_2, \ldots\}$ using $W_0, W_1, W_2, \ldots$ as hypothesis space. We use a variable u below which will change its value at most $2n$ times. Initially $u = 0$. We now define L_n in stages $s = 0, 1, \ldots$, starting with stage $s = 0$.

Stage s: Definition of $L_n(s)$. Take the first case which applies.

- If $s < n$ or $s = n + 1$, then let $L_n(s) = 1$ and go to stage $s + 1$.
- If $s = n$, then let $L_n(s) = 0$ and go to stage $s + 1$.
- If $u > 0$ and, for all $e < u$, $L_n \cap \{0, 1, 2, \ldots, s-1\} \neq W_e \cap \{0, 1, 2, \ldots, s\}$, then let $L_n(s) = 0$, let $u = 0$ and go to stage $s + 1$.
- If there is a $k < n$ such that
  - in no earlier stage M_k was dealt with,

– there is a $\sigma \in (L_n \cap \{0, 1, 2, \ldots, s-1\})^s$ such that $M_k(\sigma\tau) = M_k(\sigma)$ for all $\tau \in (L_n \cup \{s, s+1, s+2, \ldots\})^*$,

then M_k (for least such k) is dealt with in this stage, let $u = M_k(\sigma) + 1$, let $L_n(s) = 1$ and go to stage $s + 1$.

– Otherwise let $L_n(s) = 1$ and go to stage $s + 1$.

End Stage s.

It is easy to see that $L_0, L_1, \ldots$ satisfy (PA) and (PB). Furthermore, it is easy to verify that u changes from 0 to a non-zero value at most n times as at each such stage, the algorithm deals with a machine M_k with $k < n$ and will later not deal with the same machine again. Thus, L_n has at most n non-elements s, except for $s = n$, where in stage s, u is changed from a non-zero value to 0. Thus, $L_0, L_1, \ldots$ satisfy (PC). Whenever $k < n$ and M_k has a stabilizing sequence for L_n, then the algorithm will eventually deal with M_k on some stabilizing sequence σ. In particular it will set u to an upper bound of $M_k(\sigma)$. At each subsequent stage $t > s$, there is

– either an index $e \le u$ such that $L_n \cap \{0, 1, 2, \ldots, t-1\}$ equals $W_e \cap \{0, 1, 2, \ldots, t\}$ and L_n is made different from W_e by letting $L_n(t) = 1$
– or none of the W_e with $e \le u$ agrees with $L_n \cap \{0, 1, 2, \ldots, t-1\}$ and L_n is ensured to be different from all W_e with $e \le u$ by letting $L_n(t) = 0$.

It is easy to see that the latter happens latest at the stage $t = u + s + 1$ and hence u goes back to 0 eventually. Hence every machine M_k can vacillatorily learn only the sets $L_0, L_1, L_2, \ldots, L_k$ but not any L_n with $n > k$. It follows that $L_0, L_1, \ldots$ satisfy (PD) too.

The above can then be used to show that, for every class S having a vacillatory learner M using $W_0, W_1, W_2, \ldots$, there is a further vacillatory learner N using $B_0, B_1, B_2, \ldots$; the translation of the learners is the same as in part (a) with the only difference that now the learners converge to upper bounds of correct indices instead of converging to the correct indices themselves. To see this, note that if b is an upper bound of e, then $\langle d, b\rangle$ is an upper bound of $\langle d, e\rangle$ by the monotonicity of the pairing functions. Hence $B_0, B_1, B_2, \ldots$ is optimal for vacillatory learning. □

Theorem 16. *There is a numbering which is optimal but not effectively optimal for behaviourally correct learning.*

Proof. The idea is to construct a uniformly K-r.e. listing $L_0, L_1, L_2, \ldots$ of cofinite sets such that, for every m,

– $\min(\mathbb{N} - L_m)$ exists and is m;
– the machines $M_0, M_1, M_2, \ldots, M_m$ do not behaviourally correctly learn L_m.

Each set L_m is obtained using movable markers $a_0, a_1, a_2, \ldots, a_m$: One constructs a text $T_m \le_T K$ for language L_m, which enumerates all numbers except m and the final values of those markers which move only finitely often. Each marker a_k is initialized as $m + k + 1$. $T_m[s]$ contains only values below $s + m + 2$. In the case that the current value of a_k is not in $W_{M_k(T_m[s])}$, move a_k to the value $(s + 1)(m + 1) + k + 1$. Furthermore, $T_m(s)$ is the least number x neither in $\{m\} \cup \mathrm{content}(T_m[s])$ nor a current value of any marker. In the case that the value of a_k

changes infinitely often, M_k does not converge on T_m semantically to L_m, as M_k infinitely often conjectures a set not containing some intermediate value of a_k, even though this intermediate value belongs to L_m. In the case that the value of a_k changes only finitely often, the final value of a_k does not belong to L_m, but belongs to almost all of the conjectures output by M_k on T_m.

The reader should note that there are uniformly recursive approximations $L_{m,s}$ satisfying for all m that

- $\forall x \leq m \, \forall s \, [x \in L_{m,s} \Leftrightarrow x < m]$;
- $\forall x > m \, [x \in L_m \Leftrightarrow \forall^\infty s \, [x \in L_{m,s}]]$.

Using a construction similar to Proposition 8, one can construct a numbering $H_0, H_1, H_2, \ldots$ with the following property: For every k, the cofinite set $\mathbb{N} - D_k$ has exactly one index $g(k)$ and this $g(k)$ satisfies $C^K(g(k)) > 2^k$. Thus no infinite class of cofinite sets can be behaviourally correctly learnt using $H_0, H_1, H_2, \ldots$ as a hypothesis space.

Now define, for all e and $d > 0$, that $A_{\langle 0,e \rangle} = H_e$ and $A_{\langle d,e \rangle}$ is the union of all $W_{e,s}$ for which there are m, x such that

- $m < x \leq d \leq s$ and
- $m = \min(\mathbb{N} - W_{e,s})$ and
- either $x \in \bigcap_{t=d,d+1,d+2,\ldots,s} L_{m,t} - W_{e,s}$ or $x \in W_{e,s} - L_{m,s}$.

Note that $A_{\langle d,e \rangle}$ is finite if $\{0, 1, 2, \ldots, d\} \subseteq W_e$ or there exists a number $m < d$ with $L_m \cap \{0, 1, 2, \ldots, d\} = W_e \cap \{0, 1, 2, \ldots, d\}$. Furthermore, $H_0, H_1, H_2, \ldots$ covers all cofinite sets and hence $A_0, A_1, A_2, \ldots$ also covers all cofinite sets. The coverage of the coinfinite sets is now based on the following claim.

Claim 17. *Let B be a given r.e. set such that $B \notin \{\emptyset, \mathbb{N}, L_0, L_1, L_2, \ldots\}$. Then there is a constant c such that, for all e with $W_e = B$ and all $d > c$, it holds that $A_{\langle d,e \rangle} = B$.*

To see this claim, let $m = \min(\mathbb{N} - B)$ and $x = \min((L_m - B) \cup (B - L_m))$. Note that $x > m$. If $x \notin L_m$, then let $c = x + 1$, else choose c so large that $\forall s \geq c \, [c > x \wedge x \in L_{m,s}]$. Let e be such that $W_e = B$. Assume that $d > c$. Note that $x \leq d$. There are two cases.

First $x \in L_m \wedge x \notin B$. Then it holds, for all $s \geq d$, that $x \in \bigcap_{t:d \leq t \leq s} L_{m,t} - W_{e,s}$ and hence $A_{\langle d,e \rangle} = \bigcup_{s:s \geq d} W_{e,s} = W_e$.

Second $x \notin L_m \wedge x \in B$. Then there are infinitely many s with $x \in W_{e,s} - L_{m,s}$ and $A_{\langle d,e \rangle}$ is the union of the sets $W_{e,s}$ for these s; hence $A_{\langle d,e \rangle} = W_e = B$. This completes the proof of the claim.

Let S be a behaviourally correctly learnable class with learner M and let $I = \{i : H_i \in S \cap \{\mathbb{N}, L_0, L_1, L_2, \ldots\}\}$. By choice of $L_0, L_1, L_2, \ldots$ and $H_0, H_1, H_2, \ldots$, I is finite. For each $i \in I$, let F_i be a tell-tale set for H_i with respect to S. That is, F_i is a finite subset of H_i such that, for all $B \in S - \{H_i\}$, $\neg[F_i \subseteq B \subseteq H_i]$. One now defines a new learner N as follows:

$$N(\sigma) = \begin{cases} \langle 0, i \rangle, & \text{if } i \in I \text{ and } F_i \subseteq \text{content}(\sigma) \subseteq H_i; \\ \langle |\sigma|, M(\sigma) \rangle, & \text{if such an } i \in I \text{ does not exist.} \end{cases}$$

If there are several $i \in I$ qualifying, one just takes the least of these i. The new learner N clearly learns $\{H_i : i \in I\}$. Now consider any text T for a set $B \in S - \{\mathbb{N}, L_0, L_1, L_2, \ldots\}$. Then, for all sufficiently large s, $W_{M(T[s])} = B$, $s > c$ for the constant c from the claim and there is no $i \in I$ with $F_i \subseteq \text{content}(T[s]) \subseteq H_i$. It follows that $N(T[s]) = \langle s, M(T[s])\rangle$ and $A_{N(T[s])} = A_{\langle s,M(T[s])\rangle} = B$. Hence N behaviourally correctly learns B using $A_0, A_1, A_2, \ldots$ as hypothesis space and $A_0, A_1, A_2, \ldots$ is optimal for behaviourally correct learning.

Now assume by way of contradiction that $A_0, A_1, A_2, \ldots$ is effectively optimal for behaviourally correct learning. Thus, one can effectively find a learner N_d for $\{\mathbb{N} - D_d\}$ (using the numbering $A_0, A_1, A_2, \ldots$ as hypothesis space). Let T_d be a text for $\mathbb{N} - D_d$, obtained effectively from d. Let h be a partial K-recursive function such that $h(d) = e$, if N_d on T_d converges to e; otherwise, $h(d)$ is undefined. Note that $h(d) = g(d)$ for all d such that $\mathbb{N} - D_d = L_n$ for some n. Furthermore, $C^K(h(d)) \leq d + c$, for some constant c, whenever $h(d)$ is defined. However, recall that $C^K(g(d)) \geq 2^d$ for all d. This leads to contradiction, as there exist infinitely many distinct d such that $\mathbb{N} - D_d = L_n$ for some n. It follows that $A_0, A_1, A_2, \ldots$ is not effectively optimal for behaviourally correct learning. $\square$

4 Consistent and Confident Learning

There are various versions of requiring consistency for learning. For example, one can either require that consistency holds only for texts for sets from the class to be learnt or for all texts. Furthermore, one might either require that a learner is partial or that a learner is total. In the following, the version is chosen which Wiehagen and Zeugmann [23] called "totally consistent" and where the learner has to be total and always outputs hypotheses containing all data seen so far (even on data not belonging to any set to be learnt).

Definition 18 (Wiehagen and Liepe [22]). A learner M is consistent iff for every sequence σ it holds that $M(\sigma)$ is defined and $\text{content}(\sigma) \subseteq W_{M(\sigma)}$. A class S is consistently learnable iff there is a consistent learner which explanatorily learns S.

Proposition 19. *If a numbering is effectively optimal for explanatory learning then it is also effectively optimal for consistent learning.*

Proof. Let $A_0, A_1, A_2, \ldots$ be a numbering which is effectively optimal for explanatory learning. Then there is, by Theorem 6, a recursive function f such that, for all e, $d = \lim_s f(e, s)$ exists and $A_d = W_e$. Now let S be a consistently learnable class and let M be a consistent learner for S using $W_0, W_1, W_2, \ldots$ as hypothesis space. The new learner for S, using $A_0, A_1, A_2, \ldots$ as hypothesis space, is given as

$$N(\sigma) = f(M(\sigma), s) \text{ for the least } s \text{ with } s > |\sigma| \wedge \text{content}(\sigma) \subseteq A_{f(M(\sigma),s),s}.$$

As M is consistent, $\text{content}(\sigma) \subseteq W_{M(\sigma)}$. Furthermore, $f(M(\sigma), s)$ converges to a fixed value d as s goes to infinity; this d satisfies $\text{content}(\sigma) \subseteq A_{d,s}$ for almost all s. Hence, if s is sufficiently large, $\text{content}(\sigma) \subseteq A_{f(M(\sigma),s),s}$ as well. It follows that above new learner N is total and consistent.

Furthermore, when M converges on a text T to e, then N converges to a value $d = \lim_s f(e, s)$.

The reason is that there are only finitely many s for which $f(e,s)$ differs from d; thus if the initial segment $\sigma \preceq T$ processed by M is sufficiently large, then $M(\sigma) = e$ and all $s > |\sigma|$ satisfy $f(e,s) = d$ — hence $N(\sigma) = d$. By the definition of f, $A_d = W_e$. So it follows that N using A_0, $A_1, A_2, \ldots$ explanatorily learns S. $\square$

Definition 20 (Osherson, Stob and Weinstein [19], Fulk [9]). A learner is called *prudent* if it learns (according to the relevant criterion) every set for which it outputs a hypothesis on some data.

The next result shows that every consistently learnable class can be learnt by a consistent and prudent learner.

Theorem 21. *If M consistently learns a class S, then there is also a consistent and prudent learner N for S.*

Proof. Without loss of generality, one can assume that, for all L, if M converges on some text for L to i, then M converges on all texts for L to i. Furthermore, if M has a stabilizing sequence for L, then every text for L starts with a stabilizing sequence for M on L. This can be shown essentially using the same proof as Fulk [9] for explanatory learning.

Without loss of generality, we assume that S contains all sets consistently learnt by M. Now make a recursive function f such that

$$W_{f(\sigma)} = \begin{cases} W_{M(\sigma)}, & \text{if } \sigma \text{ is a stabilizing sequence for } M \text{ on } W_{M(\sigma)}; \\ \mathbb{N}, & \text{if } \mathbb{N} \in S \text{ and } \sigma \text{ is not a stabilizing sequence} \\ & \text{for } M \text{ on } W_{M(\sigma)}; \\ \{0,1,2,\ldots,x\}, & \text{if } \mathbb{N} \notin S \text{ and } x \text{ is the least number such that} \\ & x \geq \max(\{|\sigma|\} \cup \text{content}(\sigma)) \text{ and it is verified in} \\ & \text{time } x \text{ that } \sigma \text{ is not a stabilizing sequence for} \\ & M \text{ on } W_{M(\sigma)}. \end{cases}$$

Note that whenever it is not disproved within time x that σ is a stabilizing sequence for $W_{M(\sigma)}$, then

$$W_{M(\sigma)} \cap \{0,1,2,\ldots,x\} \subseteq W_{f(\sigma)}.$$

This property is useful and will go into the construction of the new learner N.

On input σ, one defines $N(\sigma)$ according to the first case which applies:

- If $\mathbb{N} \notin S$ and $\text{content}(\sigma) = \{0,1,2,\ldots,y\}$ for some y, then $N(\sigma)$ is a canonical index for this set.
- If there is some $\tau \preceq \sigma$ such that $M(\tau) = M(\sigma)$ and, for the parameter $x = \max(\{|\sigma|\} \cup \text{content}(\sigma))$, it cannot be verified in time x that τ is not a stabilizing sequence for $W_{M(\tau)}$, then $N(\sigma) = f(\tau)$ for the smallest such τ.
- Otherwise $N(\sigma) = f(\sigma)$.

Note that the conditions on τ in the second item imply that $\mathrm{content}(\sigma) \subseteq W_{f(\tau)}$. Furthermore, $\mathrm{content}(\sigma) \subseteq W_{f(\sigma)}$ for all σ. Hence N is consistent.

In the case that $\mathbb{N} \notin S$, one can see that N explanatorily learns all sets of the form $\{0, 1, 2, \ldots, y\}$. Furthermore, if $L \in S$ and T is a text for L, then there is a smallest stabilizing sequence $\sigma \preceq T$ for M on L. Now N converges to $f(\sigma)$ on T as all $\tau \prec \sigma$ eventually disqualify. By definition, $W_{f(\sigma)} = W_{M(\sigma)}$ and so N explanatorily learns L as well. Hence N explanatorily learns all sets consistently learnt by M. Furthermore, whenever N outputs a hypothesis, it is either a member of S or it can be, in the case of $\mathbb{N} \notin S$, a set of the form $\{0, 1, 2, \ldots, y\}$. N explanatorily learns all these sets and hence N is prudent. $\Box$

Theorem 22. *If $A_0, A_1, A_2, \ldots$ is optimal for explanatory learning, then $A_0, A_1, A_2, \ldots$ is also optimal for consistent learning.*

Proof. Let T_e be the canonical text for W_e; note that the T_e are all uniformly recursive. Assume that $A_0, A_1, A_2, \ldots$ is optimal for explanatory learning and let S be a consistently learnable class. By Theorem 21 there is a prudent and consistent learner M for S using $W_0, W_1, W_2, \ldots$ as hypothesis space. As $A_0, A_1, A_2, \ldots$ is optimal for explanatory learning, there is also a further explanatory learner P using $A_0, A_1, A_2, \ldots$ for the class consistently learnt by M. The new consistent learner N using $A_0, A_1, A_2, \ldots$ is defined as follows:

$$N(\sigma) = P(T_{M(\sigma)}[n]) \text{ for the least } n \text{ with } n > |\sigma| \text{ and } \mathrm{content}(\sigma) \subseteq A_{P(T_{M(\sigma)}[n]),n}.$$

The learner N uses $A_0, A_1, A_2, \ldots$ and is partial-recursive. As $M(\sigma)$ is the index of a set containing $\mathrm{content}(\sigma)$, the learner P converges on the text $T_{M(\sigma)}$ to an index c with $\mathrm{content}(\sigma) \subseteq W_{M(\sigma)} = A_c$. Hence the parameter n in the algorithm to compute $N(\sigma)$ is always found; so the learner N is total and consistent. Furthermore, if M converges on a text to e, then P is, from some time onwards, always simulated on T_e. As P converges on T_e to an index d with $A_d = W_e$ and as N always chooses a parameter $n > |\sigma|$, it follows that N converges to this d as well. Hence N explanatorily learns all the sets consistently learnt by M; in particular, N explanatorily learns the class S. This shows that $A_0, A_1, A_2, \ldots$ is optimal for consistent learning. $\Box$

The converse is not true. There is a numbering which is effectively optimal for consistent learning but not optimal for explanatory learning.

Theorem 23. *There is a numbering $A_0, A_1, A_2, \ldots$ such that:*
(a) *$A_0, A_1, A_2, \ldots$ is effectively optimal for consistent learning;*
(b) *$A_0, A_1, A_2, \ldots$ is not optimal for finite, explanatory, vacillatory or behaviourally correct learning.*

Proof. The basic idea is to make a numbering $A_0, A_1, A_2, \ldots$ such that, for every recursive set W_e, one can find in the limit a parameter d such that $A_{\langle d,e\rangle} = W_e$; however, no infinite subclass of $\{L_0, L_1, L_2, \ldots\}$, where $L_n = \{2x : x \in K\} \cup \{2n+1\}$, is learnable using $A_0, A_1, A_2, \ldots$ under any of the criteria mentioned in (b). As the class $\{L_0, L_1, L_2, \ldots\}$ is finitely learnable using W_0, $W_1, W_2, \ldots$, it follows that $A_0, A_1, A_2, \ldots$ is not optimal for the criteria given under (b).

The numbering $A_0, A_1, A_2, \ldots$ is constructed as follows: Let $H_0, H_1, H_2, \ldots$ be a Friedberg numbering [8] of all r.e. sets such that no infinite class of infinite sets is learnable using H_0, $H_1, H_2, \ldots$ under any of the criteria of finite, explanatory, vacillatory and behaviourally correct learning [12]. Now let $A_{\langle 0,e\rangle} = H_e$. For $d > 0$ let $A_{\langle d,e\rangle}$ be the union of all $W_{e,s}$ where there is an $x < d$ with $W_{e,s}(2x) \neq K_s(x)$. It is easy to see that whenever $\{x : 2x \in W_e\}$ differs from K, then there is an x with $W_e(2x) \neq K(x)$ and thus $A_{\langle d,e\rangle} = W_e$ for all $d > x$. Now let $f(e,s) = \langle x+1, e\rangle$ for the minimal x with either $W_{e,s}(2x) \neq K_s(x)$ or $x = s$. The function f is recursive and whenever $\{x : 2x \in W_e\}$ differs from K, then $\lim_s f(e,s)$ exists and is $\langle d,e\rangle$ with $A_{\langle d,e\rangle} = W_e$.

(a): Assume that M consistently learns a class S using $W_0, W_1, W_2, \ldots$ as hypothesis space. Let L be a set explanatorily learnt by M and let σ be a locking sequence for M on L. Note that due to the totalness and consistency of M, it holds that $x \in L$ iff $M(\sigma x) = M(\sigma)$. Hence L is recursive and M does not explanatory learn any nonrecursive sets. Let u be a fixed index with $A_u = \mathbb{N}$.

Now the new learner N is built as follows: Let σ be the input and $e = M(\sigma)$. Then N searches for the least $s > |\sigma|$ satisfying one of the two conditions below and continues according to the case which qualifies first.

- $W_{e,s}(2x) = K_s(x)$ for all $x \leq |\sigma|$: then $N(\sigma) = u$.
- content$(\sigma) \subseteq A_{f(e,s),s}$: then $N(\sigma) = f(e,s)$.

Note that the search for s always terminates as content$(\sigma) \subseteq W_{M(\sigma)}$ for all σ and either $K = \{x : 2x \in W_e\}$ or $A_{\langle d,e\rangle} = W_e$ for all sufficiently large d. In the second case, the limit $\lim_s f(e,s)$ converges to such a $\langle d,e\rangle$; thus content$(\sigma) \subseteq A_{f(e,s),s}$ for all sufficiently large s.

Furthermore, one can easily see that N is consistent as whichever case the search terminates in, $N(\sigma)$ is an index satisfying content$(\sigma) \subseteq A_{N(\sigma)}$.

Furthermore, if M converges on a text T, for a language it consistently learns, to an index e with W_e = content(T), then there is a least x such that $W_e(2x) \neq K(x)$. Let $d = x+1$. For all sufficiently long $\sigma \preceq T$ and all $s > |\sigma|$, $f(e,s) = \langle d,e\rangle$ and $W_{e,s}(x) \neq K_s(x)$. Hence $N(\sigma) = \langle d,e\rangle$ and N converges on T to the index $\langle d,e\rangle$ with $A_{\langle d,e\rangle} = W_e$. Thus N explanatorily learns all sets explanatorily learnt by M and N is a consistent learner for S. This implies that A_0, $A_1, A_2, \ldots$ is effectively optimal for consistent learning.

(b): The class $L_0, L_1, L_2, \ldots$ is finitely learnable as one needs only to find the unique odd number $2n+1$ in the text and then one knows that the set to be learnt is L_n. For each L_n there is exactly one index e_n with $H_{e_n} = L_n$. Then $A_{\langle 0,e_n\rangle}$ is the only member of $A_0, A_1, A_2, \ldots$ which equals L_n and any behaviourally correct learner, on a text for L_n, has to syntactically converge to $\langle 0, e_n\rangle$. By choice of the numbering $H_0, H_1, H_2, \ldots$ this is impossible and hence $\{L_0$, $L_1, L_2, \ldots\}$ is not behaviourally correctly learnable using $A_0, A_1, A_2, \ldots$; this non-learnability result transfers also to the criteria of finite, explanatory and vacillatory learning. □

Note that the proof of Theorem 9 gives a numbering which is optimal for finite learning but not optimal for consistent learning. The proof of Theorem 10 gives a numbering which is effectively optimal for vacillatory learning but not optimal for consistent learning. The proof of Theorem 12 gives a numbering which is effectively optimal for behaviourally correct learning but not optimal for consistent learning. Separation of non-effective and effective optimality for consistent learning can be obtained using the numbering $A_0, A_1, A_2, \dots$ in Theorem 13: using part (a) of Theorem 13 and Theorem 22, one has that $A_0, A_1, A_2, \dots$ is optimal for consistent learning. Note that, given a finite set D, one can effectively find a consistent learner for $\mathbb{N} - D$ using $W_0, W_1, W_2, \dots$ as hypothesis space. Using this one can modify the proof of part (b) of Theorem 13 to show that A_0, $A_1, A_2, \dots$ cannot be effectively optimal for consistent learning.

The results and the proofs of confident learning are similar to the ones of consistent learning. In the following, the definition of confidence is, as originally done, based on syntactic convergence and hence confident learners are by definition explanatory learners.

Definition 24 (Osherson, Stob and Weinstein [19]). A learner M is *confident* iff it converges syntactically on every text. A class is confidently learnable iff it has a confident explanatory learner.

The next remark gives all known implications for optimality and effective optimality which can directly be derived from previous results.

Remark 25. Only finite subclasses of $\{\mathbb{N}-\{c\} : c \in \mathbb{N}\}$ are confidently learnable. A modification of the proof of Theorem 9 can be used to show that the numbering from there is optimal for confident learning but not for explanatory, consistent, vacillatory and behaviourally correct learning.

The numbering from Theorem 10 is an example of a numbering which is effectively optimal for vacillatory learning but not for confident learning.

The numbering from Theorem 12 is an example of a numbering which is effectively optimal for behaviourally correct learning but not optimal for confident learning.

The numbering from Theorem 23 is effectively optimal for consistent learning but not optimal for confident learning. The reason is that the class of all $L_n = \{2n+1\} \cup \{2x : x \in K\}$ is confidently learnable using $W_0, W_1, W_2, \dots$ but not confidently learnable using the numbering in Theorem 23.

Theorem 28 below shows that every numbering which is optimal for explanatory learning is also optimal for confident learning. It therefore follows that there are numberings which are optimal for confident learning but not for finite, vacillatory and behaviourally correct learning, respectively.

The numberings which are effectively optimal for confident learning are K-acceptable numberings. Note that it is an immediate consequence of this characterization that a numbering is effectively optimal for confident learning iff it is effectively optimal for explanatory learning. The proof of following proposition is exactly the same as the proof of Theorem 6(b) and hence the proof is omitted.

Proposition 26. *A numbering is effectively optimal for confident learning iff it is a K-acceptable numbering.*

In the non-effective case only one inclusion holds. The proof needs the following result.

Proposition 27. *Every confidently learnable class has a prudent and confident learner which also explanatorily learns* $\mathbb{N}$.

Proof. Let M be a confident learner for a given class S. Recall that a learner is order independent [3], if for every language L, it either diverges on all texts for L or it converges on all texts for L to the same index. Using a proof similar to the locking sequence hunting construction for explanatory learning [3, 9], one may assume without loss of generality that, M is order independent and, for all L, every text for L starts with a stabilizing sequence for M on L. Thus, if σ is a stabilizing sequence for M on $W_{M(\sigma)}$, then M explanatorily learns L. Furthermore, define

$$W_{f(\tau)} = \begin{cases} W_{M(\tau)}, & \text{if } M(\tau\sigma) = M(\tau) \text{ for all } \sigma \in (W_{M(\tau)} \cup \{\#\})^*; \\ \mathbb{N}, & \text{otherwise.} \end{cases}$$

Note that, if M explanatorily learns L, then $W_{f(\tau)} = W_{M(\tau)} = L$ for all stabilizing sequences τ for M on L. Let η be a stabilizing sequence for M on $\mathbb{N}$. Let $H = W_{M(\eta)}$. If $H \neq \mathbb{N}$, then let $x = \min(\mathbb{N}-H)$, else let $x = 0$. Let u be a fixed index for $\mathbb{N}$. Let $P(\tau)$ denote the smallest prefix of τ such that, for all $\sigma \in (\text{content}(\tau) \cup \{\#\})^*$ with $|P(\tau)\sigma| \leq |\tau|$, $M(P(\tau)\sigma) = M(P(\tau)) = M(\tau)$. Now define a new learner N as follows:

$$N(\tau) = \begin{cases} f(P(\tau)) & \text{if content}(\eta x) \not\subseteq \text{content}(\tau) \text{ and} \\ & \text{content}(P(\tau)) \subseteq W_{M(P(\tau)),|\tau|}; \\ u & \text{otherwise.} \end{cases}$$

Given a set L and a text T for L, P converges on T to the smallest prefix of T which is a stabilizing sequence for M on L. Call this smallest prefix $P(T)$. If $\text{content}(P(T)) \subseteq W_{M(P(T))}$ and $\text{content}(\eta x) \not\subseteq L$, then N converges on T to $f(P(T))$, else N converges on T to u.

As the learner N converges on every text, N is confident. It can easily be seen that N explanatorily learns $\mathbb{N}$. Furthermore, N explanatorily learns all sets L such that M explanatorily learns L. Thus, N explanatorily learns S.

In the case that N outputs a conjecture of the form $f(P(\tau))$, $W_{M(P(\tau))}$ contains $\text{content}(P(\tau))$. If $P(\tau)$ a stabilizing sequence for M on $W_{M(P(\tau))}$, then both M and N explanatorily learn $W_{N(f(P(\tau)))} = W_{M(P(\tau))}$, else $W_{f(P(\tau))} = \mathbb{N}$ and N explanatorily learns $\mathbb{N}$ as well. Hence N is prudent. $\square$

Theorem 28. *Every numbering which is optimal for explanatory learning is also optimal for confident learning.*

Proof. Assume that $A_0, A_1, A_2, \ldots$ is optimal for explanatory learning and that S is a class containing $\mathbb{N}$ with a prudent and confident learner M using $W_0, W_1, W_2, \ldots$ as hypothesis space. Furthermore, let P be an explanatory learner for S using $A_0, A_1, A_2, \ldots$ as hypothesis space; P exists by the assumption that $A_0, A_1, A_2, \ldots$ is optimal for explanatory learning. Recall that T_e denotes the canonical text for W_e. Now a new learner N is defined by

$$N(\sigma) = P(T_{M(\sigma)}[|\sigma|]).$$

Given a text T, M converges on T to some index d. As M is prudent, M, and thus P, explanatorily learns W_d. Hence P converges on T_d to some index e with $A_e = W_d$. It follows that $N(T[n])$ outputs, for almost all n, the value $P(T_d[n])$ and hence N converges on T to e. Hence N is confident. Furthermore, whenever M explanatorily learns a set L, then M converges (on a text for L) to an index d with $W_d = L$. It follows using above analysis that M and N explanatorily learn $W_d = L$ using $A_0, A_1, A_2, \ldots$ and hence N explanatorily learns S. Thus, N is a confident learner for S. □

5 Learning with Additional Information

Learning with additional information is a scenario in which a learner receives, besides the text of the set to be learnt, also an upper bound on an index (in the numbering used as hypothesis space) for the set to be learnt. We can consider the learner as receiving two items as input: first an upper bound on an index for the input language and second the text for the language to be learnt.

Definition 29. A class S is explanatorily learnable with additional information using $A_0, A_1, A_2, \ldots$ as hypothesis space iff there is a learner M such that, for every d, e with $d \geq e \wedge A_e \in S$ and for every text T for A_e, $\lim_{n\to\infty} M(d, T[n])$ converges to an index c with $A_c = A_e$.

Remark 30. Jain and Sharma [11] considered also the notion of vacillatory learning with additional information (in Case's original definition [4]) and showed that the class of all r.e. sets is vacillatorily learnable using additional information using $W_0, W_1, W_2, \ldots$ as hypothesis space. More precisely, they showed that there is a recursive learner M such that, on every text T for an r.e. set W_e and every $b \geq e$, for almost all n, $M(b, T[n]) \leq b \wedge W_{M(b,T[n])} = W_e$. The proof of Jain and Sharma [11] works for every universal numbering $A_0, A_1, A_2, \ldots$ and hence every universal numbering is optimal for vacillatory learning. As one can use the same learner M for every class of r.e. sets, every universal numbering is even effectively optimal for vacillatory learning with additional information.

Note that the additional information d must be chosen according to the hypothesis space A_0, $A_1, A_2, \ldots$ used and not according to any other numbering.

Recall that Jain and Stephan [12] called a universal numbering $A_0, A_1, A_2, \ldots$ a Ke-numbering iff $\{\langle i, j\rangle : A_i = A_j\} \leq_T K$. Ke-numberings are generalizations of Friedberg numberings and can never be acceptable or K-acceptable.

Theorem 31. *A numbering is optimal for learning with additional information iff it is effectively optimal for learning with additional information iff it is a Ke-numbering.*

Proof. Assume that a Ke-numbering $A_0, A_1, A_2, \ldots$ is given; then there is a K-recursive function f with $f(e) = \min(\{d \leq e : A_d = A_e\})$. This f can be approximated using a recursive sequence of recursive functions $(f_s)_{s\in\mathbb{N}}$. By Remark 30 there is a recursive M such that, for every index e and for every text T for A_e and every $b \geq e$, almost all n satisfy $M(b, T[n]) \leq b \wedge W_e = A_{M(b,T[n])}$. Given a set A_e, a text T of A_e and a bound $b \geq e$, the new learner N given as

$N(b, T[n]) = f_n(M(b, T[n]))$ converges syntactically to the minimal index of the given set A_e. This is so, as almost all hypotheses of M are from the finitely many indices of A_e below b and f_n coincides on these indices with f for almost all n. Hence the class of all r.e. sets is explanatorily learnable with additional information using $A_0, A_1, A_2, \ldots$ as a hypothesis space. As one can use the learner N also for every subclass of the class of all r.e. sets, it follows that $A_0, A_1, A_2, \ldots$ is effectively optimal for learning with additional information.

On the other hand, if an optimal numbering is given, one can do the following to check in the limit whether $A_i = A_j$: Suppose a learner learning the class of all r.e. sets using $A_0, A_1, A_2, \ldots$ is given. One can find in the limit the least stabilizing sequences for the learner on A_i and A_j respectively, with respect to the upper bound $i + j + 1$. If the stabilizing sequence found for A_i equals to that found for A_j, then $A_i = A_j$, else $A_i \neq A_j$. This completes the proof. □

Remark 32. One can similarly show that a numbering is a Ke-numbering iff it is optimal for confident learning with additional information iff it is effectively optimal for confident learning with additional information. On one hand, one can confidently learn the class of all r.e. sets using additional information in all Ke-numberings. To see this, note that the M constructed by Jain and Sharma [11, Proposition 16] and referred to in Remark 30 also satisfies the following: given a bound b and text T which is not a text for any A_e with $e \leq b$, the algorithm converges to the least index $e \leq b$ such that $\max(\{x : \forall y < x\,[y \in A_e \Leftrightarrow y \text{ occurs in } T]\})$ is maximal. It follows that the translation $f_n(M(b, T[n]))$ from Theorem 31 converges on all texts. Thus, N is confident as well. On the other hand, by definition, confident learners with additional information are explanatory learners with additional information. Thus, a numbering is optimal only if it is a Ke-numbering.

A natural question is whether there are numberings which are optimal for finite learning with additional information. The following theorem answers this question negatively.

Theorem 33. *There is no numbering which is optimal or effectively optimal for finite learning with additional information.*

Proof. Suppose a universal numbering $A_0, A_1, A_2, \ldots$ is given. Let $M_0, M_1, M_2, \ldots$ be a numbering of all finite learners with additional information (where $A_0, A_1, A_2, \ldots$ is the numbering used by these learners). Here one may assume that for any text T and any additional information e, any learner M_i with additional information e outputs at most one conjecture on the text T. Let g be a (non-recursive) function such that $g(n)$ is the sum of the least indices of $\{2n\}$ and $\{2n, 2n+1\}$ in $A_0, A_1, A_2, \ldots$; note that $g \leq_T K$ (as using the oracle K, one can test for every e whether $A_e = \{2n\}$ or $A_e = \{2n, 2n+1\}$). Let T_n be a recursive text for $\{2n\}$, say, $T_n(m) = 2n$ for all m. Let $f(n) = 1$, if M_n, with additional information $g(n)$, outputs on text T_n some index e_n such that $A_{e_n} = \{2n\}$; otherwise let $f(n) = 0$. Note that $f \leq_T K$. Let $L_n = \{2n\}$, if $f(n) = 0$; $L_n = \{2n, 2n+1\}$ otherwise. It is now easy to verify that $\{L_0, L_1, L_2, \ldots\}$ is not finitely learnable with additional information using the numbering $A_0, A_1, A_2, \ldots$, as M_n, with additional information $g(n)$, does not finitely learn L_n.

We now show that there is a universal numbering $B_0, B_1, B_2, \ldots$ for which $\{L_0, L_1, L_2, \ldots\}$ is finitely learnable with additional information. Let f_s be a recursive approximation to f. One can then easily construct a universal numbering $B_0, B_1, B_2, \ldots$ along with its approximations from below $B_{i,s}$, such that for each n there is a number t_n with $B_{t_n} = B_{t_n,t_n+1} = \{2n\}$, $B_{t_n+1} = B_{t_n+1,t_n+1} = \{2n, 2n+1\}$, $f_s(n) = f_{t_n}(n)$ for all $s > t_n$ and

$$\forall m < t_n \, [2n \in B_m \Rightarrow B_{m,t_n+1} \not\subseteq \{2n, 2n+1\}].$$

Now the class $\{L_0, L_1, L_2, \ldots\}$ is finitely learnable with additional information using $B_0, B_1, B_2, \ldots$ as hypothesis space. The learner, with additional information s, waits for the first number of form $2n$ to occur in the input and then outputs the least index $e \leq s+1$ such that $2n \in B_{e,s+1}, 2n + f_s(n) \in B_{e,s+1}$, and no $x \in \mathbb{N} - \{2n, 2n+1\}$ belongs to $B_{e,s+1}$. Note that by the definition of the numbering $B_0, B_1, B_2, \ldots$ and by $s \geq t_n$ the index e is an index (in $B_0, B_1, B_2, \ldots$) for the language L_n.

In summary, it has been shown that for every universal numbering $A_0, A_1, A_2, \ldots$ there is a further universal numbering $B_0, B_1, B_2, \ldots$ and a class $\{L_0, L_1, L_2, \ldots\}$ such that $\{L_0, L_1, L_2, \ldots\}$ can be learnt finitely with additional information using $B_0, B_1, B_2, \ldots$ but not using $A_0, A_1, A_2, \ldots$; hence $A_0, A_1, A_2, \ldots$ cannot be optimal or effectively optimal for finite learning with additional information. $\square$

6 Open Problems

Not fully characterized is the optimality of Ke-numberings. First some facts.

Remark 34. It follows along the lines of previous work [12] that the classes $\{L_0, L_1, L_2, \ldots\}$ given by $L_n = \{2m : m \in \mathbb{N}\} \cup \{2n+1\}$ and $\{H_0, H_1, H_2, \ldots\}$ given by $H_n = \{2m : m \leq |W_n|\} \cup \{2n+1\}$ are both finitely learnable using $W_0, W_1, W_2, \ldots$, but for every Ke-numbering $A_0, A_1, A_2, \ldots$ at least one of these classes is not vacillatorily learnable using this numbering. Hence Ke-numberings are not optimal for finite, explanatory and vacillatory learning.

An open question by Jain and Stephan [12] asks whether every behaviourally correctly learnable class has a learner which uses a Ke-numbering as hypothesis space. The natural counterpart of this question is to ask for the existence of a Ke-numbering which is optimal for behaviourally correctly learning.

Open Problem 35. *Is every behaviourally correct learnable class learnable using some Ke-numbering* [12]*? Is there a Ke-numbering which is optimal for behaviourally correct learning?*

Optimality of Ke-numberings for consistent learning is open as well.

Open Problem 36. *Is there a Ke-numbering which is optimal for consistent learning?*

7 Conclusion

Acceptable numberings are quite convenient hypothesis spaces as they permit the learning of all classes which are learnable with respect to any hypothesis space (for most learning criteria).

Freivalds, Kinber and Wiehagen [7] investigated the one-one numberings as an alternative hypothesis space. They established that, on the one hand, every explanatorily learnable class of functions can be learnt using such a hypothesis space, but on the other hand, the hypothesis space has to be tailored for the class to be learned — there is no single one-one hypothesis space using which one can explanatorily learn every learnable class of functions. Jain and Stephan [12] transferred this result into the setting of learning languages. Based on this result, one might ask whether, except for the acceptable numberings, any other numbering is optimal for learning at all, that is, any other numbering can be used to learn all learnable classes.

The starting point of the present work is the observation that not only acceptable numberings but also nearly acceptable numberings are optimal for the criteria of finite, explanatory, consistent, vacillatory and behaviourally correct learning. Based on this observation, it is investigated which numberings are optimal for which learning criterion. In particular, it is shown that it depends heavily on the learning criterion whether a numbering is optimal for this criterion or not. Most distinct learning criteria I, J can be separated in the sense that there is a numbering optimal for I learning but not optimal for J learning. But there is one notable exception: numberings which are optimal for explanatory learning are also optimal for consistent learning. Furthermore, the notion of learning with additional information is different from all others as there the Ke-numberings are optimal for learning while the acceptable numberings are not. The reason is that the additional information is numbering-dependent and in a Ke-numbering the upper bound on the least index can be used much better than in an acceptable numbering. While it is known that Ke-numberings are not optimal for explanatory or vacillatory learning, it remains an open problem whether they are optimal for behaviourally correct learning or consistent learning.

Besides optimality, also the notion of effective optimality has been considered. This notion turned out to be much more regular than optimality itself. For example, a numbering is effectively optimal for finite learning iff it is acceptable and effectively optimal for explanatory learning iff it is K-acceptable. Therefore, there are also more implications than in the case of optimality: for example, every numbering effectively optimal for explanatory learning is also effectively optimal for vacillatory learning, but not vice versa.

References

1. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
2. Janis Bārzdiņš. Two theorems on the limiting synthesis of functions. *Theory of Algorithms and Programs*, Volume 1, pages 82–88, Latvian State University, Riga, Latvia, 1974.
3. Lenore Blum and Manuel Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.
4. John Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28:1941–1969, 1999.
5. John Case, Sanjay Jain and Mandayam Suraj. Control Structures in Hypothesis Spaces: The Influence on Learning. *Theoretical Computer Science*, 270:287–308, 2002.

6. Dick de Jongh and Makoto Kanazawa. Angluin's theorem for indexed families of r.e. sets and applications. *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, ACM Press, pages 193–204, 1996.
7. Rūsiņš Freivalds, Efim Kinber and Rolf Wiehagen. Inductive inference and computable one-one numberings. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 28:463–479, 1982.
8. Richard Friedberg. Three theorems on recursive enumeration. *The Journal of Symbolic Logic*, 23(3):309–316, 1958.
9. Mark Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.
10. E. Mark Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
11. Sanjay Jain and Arun Sharma. Learning with the knowledge of an upper bound on program size. *Information and Computation*, 102:118–166, 1993.
12. Sanjay Jain and Frank Stephan. Learning in Friedberg numberings. *Information and Computation*, 206:776–790, 2008.
13. Sanjay Jain, Daniel Osherson, James S. Royer and Arun Sharma. *Systems That Learn: An Introduction to Learning Theory.* Second Edition. MIT-Press, Boston, MA., 1999. Second Edition. MIT-Press, 1999.
14. Steffen Lange. *Algorithmic Learning of Recursive Languages.* Habilitationsschrift, Fakultät für Mathematik und Informatik, Universität Leipzig, Mensch und Buch Verlag, Berlin, 2000.
15. Steffen Lange and Thomas Zeugmann. Language learning in dependence on the space of hypotheses. *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, Santa Cruz, California, United States, pages 127–136, 1993.
16. Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Second Edition, Springer, 1997.
17. Piergiorgio Odifreddi. *Classical Recursion Theory*, *Studies in Logic and the Foundations of Mathematics*, volume 125. North-Holland, Amsterdam, 1989.
18. Piergiorgio Odifreddi. *Classical Recursion Theory II*, *Studies in Logic and the Foundations of Mathematics*, volume 143. Elsevier, Amsterdam, 1999.
19. Daniel Osherson, Michael Stob and Scott Weinstein. *Systems That Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists.* Bradford — The MIT Press, Cambridge, Massachusetts, 1986.
20. Daniel N. Osherson and Scott Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.
21. Robert I. Soare. *Recursively enumerable sets and degrees.* Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1987.
22. Rolf Wiehagen and Walter Liepe. Charakteristische Eigenschaften von erkennbaren Klassen rekursiver Funktionen. *Journal of Information Processing and Cybernetics (EIK)*, 12:421–438, 1976.
23. Rolf Wiehagen and Thomas Zeugmann. Learning and consistency. *Algorithmic Learning for Knowledge-Based Systems*, Springer LNAI, 961:1–24, 1995.

24. Thomas Zeugmann. *Algorithmisches Lernen von Funktionen und Sprachen.* Habilitationsschrift, Technische Hochschule Darmstadt, 1993.
25. Sandra Zilles. Separation of uniform learning classes. *Theoretical Computer Science*, 313:229–265, 2004.
26. Sandra Zilles. Increasing the power of uniform inductive learners. *Journal of Computer and System Sciences*, 70:510–538, 2005.