Learning in Friedberg Numberings

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. In this paper we consider learnability in some special numberings, such as Friedberg numberings, which contain all the recursively enumerable languages, but have simpler grammar equivalence problem compared to acceptable numberings. We show that every explanatorily learnable class can be learnt in some Friedberg numbering. However, such a result does not hold for behaviourally correct learning or finite learning. One can also show that some Friedberg numberings are so restrictive that all classes which can be explanatorily learnt in such Friedberg numberings have only finitely many infinite languages. We also study similar questions for several properties of learners such as consistency, conservativeness, prudence, iterativeness and non U-shaped learning. Besides Friedberg numberings, we also consider the above problems for programming systems with K-recursive grammar equivalence problem.

1 Introduction

Consider the following model of learning languages, first studied by Gold [14]. A learner receives, one element at a time, all and only the sentences of a language (such a presentation of data is called text of the language). As the learner receives the elements of the language, it conjectures hypotheses about what the input language might be. The conjecture about the input language may change over time, as more and more data becomes available. In inductive inference, we use indices from some underlying numbering or programming system as hypotheses. Following conventions from formal languages, we refer to these indices as grammars. One can say that the learner is successful if the sequence of grammars output as above converges to a grammar for the input language. This is essentially the model of **TxtEx**-learning (= explanatory learning) as proposed by Gold [14] and subsequently studied by several researchers [1, 5, 10, 16, 28, 33].

One of the important issues in learning has been the hypotheses space which a learner uses for making its conjectures. A natural hypotheses space, as considered by Gold [14], is an acceptable programming system. However, there have also been several studies which consider special programming systems [33]. For example, in the context of learning indexed families of languages

* Supported in part by NUS grant number R252-000-212-112 and R252-000-308-112.

(an indexed family is a uniformly recursive family of languages), the hypotheses space often considered are themselves indexed families (where the hypotheses space might be class-preserving or class-comprising; a class-preserving hypotheses space contains exactly the languages in the class being learnt while a class-comprising hypotheses space may contain some other languages in addition to the languages of the class being learnt). Furthermore, considering special hypotheses spaces have also been useful in obtaining various characterizations of learnability — see, for example, [17, 30, 31, 33].

Testing grammar equivalence in acceptable numberings is a difficult problem [26]. In this paper we consider learnability in some special numberings, which contain all the recursively enumerable languages, but with simpler grammar equivalence problem. Friedberg numberings [11] are numberings which contain exactly one grammar for each recursively enumerable language. Besides their historical importance, Friedberg numberings may be considered as a natural hypotheses space, as they do not contain any redundancy. Another natural class of numberings is the Ke-numberings in which grammar equivalence problem is recursive in the halting problem. Freivalds, Kinber and Wiehagen [12] considered learnability of recursive functions in Friedberg and other one-one numberings (for the criteria of explanatory and finite learning). We extend their study by considering how the learnability in various common criteria are effected when one uses hypotheses spaces as above.

We show (Theorem 10) that for **TxtEx**-model of learning, as described above, one can learn every **TxtEx**-learnable class in some Friedberg numbering. However, no Friedberg numbering is omnipotent. More precisely, for every Friedberg numbering η, there exists a **TxtEx**-learnable class which cannot be learnt using hypotheses space η. Furthermore, there are Friedberg numberings η which are trivial in the sense that any class **TxtEx**-learnable in η contains only finitely many infinite languages (Theorem 28).

In finite learning [14], denoted **TxtFin**, one requires that the learner outputs just one hypothesis, which must be correct. In contrast to the result for **TxtEx**-learning, there are **TxtFin**-learnable classes which cannot be learnt in any Friedberg numbering (Theorem 11). However, Ke-numberings are not so restrictive, as every **TxtFin**-learnable class can be learnt in some Ke-numbering (Theorem 15). Theorem 13 gives a characterization of the recursively enumerable classes which can be learnt in Friedberg numberings.

Several properties of learners have been considered in the literature. For example a consistent learner [1, 4] is a learner whose hypotheses always generate the data seen up to the point an hypothesis is made. A conservative learner does not change a hypothesis which is consistent with the input [2, 33]. A prudent learner [24] only outputs hypotheses for the languages which it is able to learn. A confident learner [24] always converges on any input text, even on texts for languages outside the class being learnt. A non U-shaped learner is a learner which does not have a sequence of hypotheses of form "…, correct hypothesis, …, wrong hypothesis, …, correct hypothesis, …" [3, 7, 8]. We denote the criteria of prudent, confident, consistent and non U-shaped learning with **PrudentTxtEx**, **ConfTxtEx**, **ConsTxtEx** and **NUShTxtEx**, respectively; accordingly for restricted variants. We show that, though confident and consistent learning are not restrictive for learning in Friedberg numberings (Theorems 16 and 26), non

U-shaped, conservative and prudent learning are restrictive (Theorems 19 and 20). On the other hand, none of the above properties are restrictive for learning in Ke-numberings (Theorems 21, 22 and 23 along with Theorems 16 and 26).

Behaviourally correct learning [10, 25] is similar to **TxtEx**-learning except that one does not require syntactic convergence, but only semantic convergence: the hypotheses conjectured by the learner are correct beyond some time. For Friedberg numberings, notion of **TxtBc** collapses to **TxtEx** due to trivial grammar equivalence problem. It is open at present whether every **TxtBc**-learnable class can be learnt in some Ke-numberings — though we can show that every class which can be **TxtFEx**-learnt can be **TxtBc**-learnt in some Ke-numbering (**TxtFEx**-learning [9] is **TxtBc**-learning where the learner only outputs finitely many distinct hypotheses). We can though show that there exists a non U-shaped behaviourally learnable class, which cannot be learnt in non U-shaped behaviourally correct manner in any Ke-numbering (Theorem 34).

Partial identification [24] is a very general criterion which permits to learn the class of all r.e. sets in acceptable numberings. We show that this learnability result carries over to learning with respect to any given Ke-numbering (Theorem 35) although it does not carry over to all universal numberings (Theorem 36).

The next table summarizes for which major criteria the learning with respect to Friedberg numberings or Ke-numberings is restrictive.

Summary of Major Results.

In Friedberg numberings	In Ke-numberings	In acceptable numberings
FrTxtFin	$\subset$ **KeTxtFin**	$=$ **TxtFin**
FrTxtEx	$=$ **KeTxtEx**	$=$ **TxtEx**
ConfFrTxtEx	$=$ **ConfKeTxtEx**	$=$ **ConfTxtEx**
TConsFrTxtEx	$=$ **TConsKeTxtEx**	$=$ **TConsTxtEx**
PrudentFrTxtEx	$\subset$ **PrudentKeTxtEx**	$=$ **PrudentTxtEx**
NUShFrTxtEx	$\subset$ **NUShKeTxtEx**	$=$ **NUShTxtEx**
FrTxtBc	$\subset$ **KeTxtBc**	$\subseteq$ **TxtBc**
NUShFrTxtBc	$\subset$ **NUShKeTxtBc**	$\subset$ **NUShTxtBc**

2 Notation and Preliminaries

Any unexplained recursion-theoretic notions can be found in the textbooks of Odifreddi [23] and Rogers [26].

$\mathbb{N}$ denotes the set of natural numbers, $\{0,1,2,\ldots\}$. $\emptyset$ denotes the empty set. $card(S)$ denotes the cardinality of set S. $\max(S)$ and $\min(S)$, respectively, denote the maximum and minimum of a set S, where $\max(\emptyset)$ is 0 and $\min(\emptyset)$ is ∞. The symbols $\subseteq, \supseteq, \subset, \supset$ respectively denote the subset, superset, proper subset and proper superset relation between sets. $A \triangle B$ denotes the symmetric difference of A and B: $(A \cup B) - (A \cap B)$. The quantifiers $\forall^\infty$ and $\exists^\infty$ mean "for all but finitely many" and "there exist infinitely many", respectively. So

$$(\forall^\infty n)\,[P(n)] \Leftrightarrow (\exists m)\,(\forall n > m)\,[P(n)] \quad \text{and} \quad (\exists^\infty n)\,[P(n)] \Leftrightarrow (\forall m)\,(\exists n > m)\,[P(n)].$$

A pair $\langle i, j\rangle$ stands for an arbitrary, computable one-to-one encoding of all pairs of natural numbers onto $\mathbb{N}$ [26]. Similarly we can define $\langle\cdot, \ldots, \cdot\rangle$ for encoding n-tuples of natural numbers, for $n > 1$, onto $\mathbb{N}$.

Any partial recursive function of two arguments is called a numbering. For a numbering ψ, $\psi_i(x)$ denotes $\psi(i, x)$. We let Ψ denote a Blum complexity measure [6] associated with the numbering ψ. We let $\psi_{i,s}(x) = \psi_i(x)$, if $x < s$ and $\Psi_i(x) < s$; $\psi_{i,s}(x)$ is undefined if $x \geq s$ or $\Psi_i(x) \geq s$. We let $W_i^\psi = domain(\psi_i)$ and $W_{i,s}^\psi = domain(\psi_{i,s})$. We call i a ψ-grammar for W_i^ψ.

For numberings ψ and η, $\psi \leq \eta$ denotes that there exists a recursive function g such that $W_i^\psi = W_{g(i)}^\eta$ for all i. $\psi \leq^A \eta$ denotes that there exists an A-recursive function g such that $W_i^\psi = W_{g(i)}^\eta$ for all i.

$\mathcal{E}$ denotes the class of all recursively enumerable (r.e.) subsets of the natural numbers [26]; an r.e. set is also called a *language*. $\mathcal{F}$ is the class of all finite sets and $\mathcal{I}$ is the class $\{\emptyset, \{0\}, \{0,1\}, \{0,1,2\}, \ldots, \{0,1,\ldots,n\}, \ldots\}$. A *universal numbering* [26] ψ is a numbering such that, for all $L \in \mathcal{E}$, there exists a ψ-grammar for L. An *acceptable numbering* [26] ψ is a numbering such that, for all numberings η, $\eta \leq \psi$. Acceptable numberings are also called Gödel numberings.

φ denotes a fixed acceptable programming system for the partial computable functions [26]. We let $W_e = W_e^\varphi = domain(\varphi_e)$. $K = \{e : e \in W_e\}$, the diagonal halting problem, is a standard example for a nonrecursive r.e. set.

Friedberg [11] showed that there exist numberings in which every r.e. language has exactly one index (grammar). Hence the equivalence problem for grammars is obviously recursive in such numberings; furthermore, one can easily translate every numbering with a recursive equivalence problem into a Friedberg numbering. It might be important to relax this condition and to consider numberings where the equivalence problem is only K-recursive. K-recursive equivalence and translations have already received some attention; for example Goncharov [15] showed that if two Friedberg numberings of a given family of r.e. sets are not equivalent but can be K-recursively translated into each other, then this family has infinitely many non-equivalent numberings.

We are not aware of any common name for numberings with a K-recursive equivalence problem; thus we refer to them as Ke-numberings, "Ke" standing for "K-recursive equivalence".

Definition 1. A *Friedberg-numbering* is a universal numbering in which every recursively enumerable set has exactly one grammar. A *Ke-numbering* is a universal numbering for which the grammar equivalence problem is K-recursive.

A class $\mathcal{L}$ is said to be recursively enumerable if there exists an r.e. set S such that $\mathcal{L} = \{W_i : i \in S\}$. Note that for a non-empty recursively enumerable class $\mathcal{L}$, there exists a recursive function h such that $\mathcal{L} = \{W_{h(i)} : i \in \mathbb{N}\}$. A class $\mathcal{L}$ is said to be one-one recursively enumerable iff $\mathcal{L}$ is finite or there exists a recursive function h such that $\mathcal{L} = \{W_{h(i)} : i \in \mathbb{N}\}$ and, for all different i, j, $W_{h(i)} \neq W_{h(j)}$.

We now introduce the basic definitions of inductive inference, that is, of Gold-style computational learning theory.

Definition 2. A *sequence* σ is a mapping from an initial segment of $\mathbb{N}$ into $\mathbb{N}\cup\{\#\}$. The content of a finite sequence σ is the set of natural numbers occurring in σ and is denoted by $\mathrm{content}(\sigma)$. The length of a sequence σ is the number of elements in the domain of σ and is denoted by $|\sigma|$. For a subset L of $\mathbb{N}$, $Seg(L)$ denotes the set of sequences σ with $\mathrm{content}(\sigma) \subseteq L$. An infinite sequence T is a mapping from $\mathbb{N}$ to $\mathbb{N}\cup\{\#\}$. Furthermore, $\mathrm{content}(T)$ denotes the set of natural numbers in the range of T. T is a *text for* L iff $L = \mathrm{content}(T)$.

Concatenation of two sequences σ and τ is denoted by $\sigma\tau$. If $x \in (\mathbb{N} \cup \{\#\})$, then σx means $\sigma\tau$ where τ is the sequence consisting of exactly one element which is x. $\sigma \subseteq \tau$ means that σ is an initial segment of τ and $\sigma \subset \tau$ means that σ is a proper initial segment of τ.

Intuitively, a *text* for a language L is an infinite stream or sequential presentation of *all* the elements of the language L in any order and with the #'s representing pauses in the presentation of the data. For example, the only text for the empty language is an infinite sequence of #'s. We let T, with possible subscripts and superscripts, range over texts. $T[n]$ denotes the finite initial segment of T with length n, that is $T[n]$ is $T(0)T(1)\dots T(n-1)$. $\sigma \subset T$ denotes the fact that σ is an initial segment of T. Observe that in this case we have $\sigma = T[|\sigma|]$.

Note that one can effectively produce a text for a language L, from its grammar in a given numbering. Canonical text for W_j (W_j^ψ) denotes such an effective text.

A learner is an algorithmic mapping from finite sequences to $\mathbb{N} \cup \{?\}$. Output of ? denotes the fact that the learner does not wish to issue a conjecture on the input. The elements of $\mathbb{N}$ in the output of a learner are interpreted as a grammar in some predetermined numbering (also called hypotheses space). M, with possible superscripts and subscripts, is intended to range over language learning machines. We say that $M(T)\downarrow$ iff there exists an i such that, for all but finitely many n, $M(T[n]) = i$. In this case we say that $M(T)\downarrow = i$; in the case that there is no such i we say that $M(T)\uparrow$.

We now give the formal definitions of explanatory (**TxtEx**) learning, finite (**TxtFin**) learning and behaviourally correct (**TxtBc**) learning.

Definition 3. [10, 14, 25] Suppose ψ is a numbering and let $\mathbf{I}$ be a variable ranging over the criteria **TxtEx**, **TxtFin** and **TxtBc** which are defined now.

(a) M $\mathbf{TxtEx}_\psi$*-identifies a text* T just in case $(\exists i : W_i^\psi = \mathrm{content}(T))\ (\forall^\infty n)[M(T[n]) = i]$.

(b) M $\mathbf{TxtFin}_\psi$*-identifies a text* T just in case $(\exists i : W_i^\psi = \mathrm{content}(T))\ (\exists n)[(\forall m \geq n)$ $[M(T[m]) = i]$ and $(\forall m < n)[M(T[m]) = ?]]$.

(c) M $\mathbf{TxtBc}_\psi$*-identifies a text* T just in case $(\forall^\infty n)[W^\psi_{M(T[n])} = \mathrm{content}(T)]$.

(d) M $\mathbf{I}_\psi$*-identifies an r.e. language* L (written: $L \in \mathbf{I}_\psi(M)$) just in case M $\mathbf{I}_\psi$-identifies each text for L.

(e) M $\mathbf{I}_\psi$*-identifies a class* $\mathcal{L}$ *of r.e. languages* (written: $\mathcal{L} \subseteq \mathbf{I}_\psi(M)$) just in case M $\mathbf{I}_\psi$-identifies each language from $\mathcal{L}$.

(f) $\mathbf{I}_\psi = \{\mathcal{L} \subseteq \mathcal{E} : (\exists M)[\mathcal{L} \subseteq \mathbf{I}_\psi(M)]\}$ and $\mathbf{I} = \bigcup_\psi \mathbf{I}_\psi$.

Note that parts (d), (e) and (f) are not specific to $\mathbf{I} \in \{\mathbf{TxtEx}, \mathbf{TxtFin}, \mathbf{TxtBc}\}$ but also done for other learning criteria introduced later. Furthermore, as φ is acceptable numbering, it holds for all numberings ψ that $\mathbf{TxtEx}_\psi \subseteq \mathbf{TxtEx}_\varphi$, $\mathbf{TxtFin}_\psi \subseteq \mathbf{TxtFin}_\varphi$ and $\mathbf{TxtBc}_\psi \subseteq \mathbf{TxtBc}_\varphi$. Thus, $\mathbf{I} = \mathbf{I}_\varphi$ for $\mathbf{I} \in \{\mathbf{TxtEx}, \mathbf{TxtBc}, \mathbf{TxtFin}\}$. For this reason, we often use the notation $\mathbf{I}$-identification for $\mathbf{I}_\varphi$-identification.

Blum and Blum [5] introduced the notion of locking sequences and Fulk [13] generalized this notion to stabilizing sequences. We use these notions often in our proofs.

Definition 4. (a) [13] We say that σ is a $\mathbf{TxtEx}$-*stabilizing sequence* for a learner M on a set L iff $\sigma \in Seg(L)$ and $M(\sigma\tau) = M(\sigma)$ for all $\tau \in Seg(L)$.

(b) [5] σ is called a $\mathbf{TxtEx}_\psi$-*locking sequence* for M on L iff σ is a stabilizing sequence for M on L and $W^\psi_{M(\sigma)} = L$.

Lemma 5. [5] *Suppose M $\mathbf{TxtEx}_\psi$-identifies L. Then,*

(a) *there exists a $\mathbf{TxtEx}_\psi$-locking sequence for M on L;*

(b) *for every $\sigma \in Seg(L)$, there exists a $\tau \in Seg(L)$ such that $\sigma\tau$ is a $\mathbf{TxtEx}_\psi$-locking sequence for M on L;*

(c) *every $\mathbf{TxtEx}$-stabilizing sequence σ for M on L is also a $\mathbf{TxtEx}_\psi$-locking sequence for M on L.*

Note that the definitions for stabilizing and locking sequence, as well as Lemma 5, can be generalized to other learning criteria such as $\mathbf{TxtBc}$. We often omit the term like "$\mathbf{TxtEx}_\psi$" from $\mathbf{TxtEx}_\psi$-locking (stabilizing) sequence, when it is clear from context.

We assume some fixed one-one ordering of all the finite sequences, $\sigma_0, \sigma_1, \ldots$; thus, one can talk about the least stabilizing sequence and so on.

Definition 6. (a) [5] M is *order independent* iff for all texts T, if $M(T)\downarrow = i$, then for all T' such that $\text{content}(T') = \text{content}(T)$, $M(T')\downarrow = i$.

(b) [13, 27] M is *rearrangement independent* iff for all σ and τ such that $\text{content}(\sigma) = \text{content}(\tau)$ and $|\sigma| = |\tau|$, $M(\sigma) = M(\tau)$.

Given any learner M, one can construct a learner M' such that $\mathbf{TxtEx}(M) \subseteq \mathbf{TxtEx}(M')$ and M' is rearrangement and order independent [5, 13].

In this paper we are mainly interested in learnability in Friedberg numberings and Ke-numberings. To this end, for any learning criterion $\mathbf{I}$, we let $\mathbf{FrI}$ denote the union of all $\mathbf{I}_\psi$, where ψ is a Friedberg numbering and let $\mathbf{KeI}$ denote the union of all $\mathbf{I}_\psi$, where ψ is a Ke-numbering.

3 Ke-Numberings and Friedberg Numberings

In this section, some basic learnability properties are established for Ke-numberings and Friedberg numberings. The next result shows that there are quite natural examples of Ke-numberings:

Proposition 7. *If ψ is a universal numbering such that every infinite r.e. language has only one ψ-grammar, then ψ is a Ke-numbering.*

Proof. Given two different indices i, j, search with help of the oracle K until an x is found such that one of the following conditions hold:

- $x \in W_i^\psi \triangle W_j^\psi$;
- $(\forall y \in W_i^\psi \cup W_j^\psi)[y \leq x]$.

The search terminates as either the two sets are different or both are finite and equal. Having determined x,

$$W_i = W_j \Leftrightarrow W_i \cap \{0, 1, \ldots, x\} = W_j \cap \{0, 1, \ldots, x\}.$$

The above can be checked using the oracle K. $\square$

Remark 8. Note that the Friedberg numberings and Ke-numberings in this paper are numberings of sets, not of functions. Although they cover all r.e. sets, they do not cover all partial-recursive functions. The learnability results can be translated: Given a numbering ψ covering all r.e. sets and a Friedberg numbering μ covering all partial-recursive functions, let $e_0, e_1, e_2, \ldots$ be a recursive one-one enumeration of $\{e : \exists x\, [\mu_e(x)\downarrow > 0]\}$ and define

$$\nu_d(x) = \begin{cases} 0 & \text{if } x \in W_e^\psi \text{ and } d = 2e; \\ \mu_{e_k}(x) & \text{if } \mu_{e_k}(x)\downarrow \text{ and } d = 2k+1; \\ \uparrow & \text{otherwise.} \end{cases}$$

It is easy to see that (a) ν is a Ke-numbering (for functions) iff ψ is a Ke-numbering (for sets), (b) ν is a Friedberg numbering (for functions) iff ψ is a Friedberg numbering (for sets) and (c) all $\mathbf{I}_\psi$-learnable classes are also $\mathbf{I}_\nu$-learnable.

So considering numberings of all partial-recursive functions does not bring in really new phenomena except that one has to adapt the notion of Ke-numbering to a numbering where $\{\langle i, j\rangle : \nu_i = \nu_j\} \leq_T K$. The reason is that there is no numbering η of all partial-recursive functions such that $\{e : W_e^\eta = \mathbb{N}\} \leq_T K$ as otherwise there would be a numbering of all total-recursive functions.

For the ease of notation, we consider in this paper only numberings which are universal in the sense that they cover all possible domains of functions and not in the sense that they cover all partial-recursive functions.

Theorem 9. *Suppose ψ is a Ke-numbering. Then, there exists a Friedberg numbering η such that $\psi \leq^K \eta$ and $\eta \leq^K \psi$.*

Proof. We use a construction similar to that of Kummer [20, pages 29–30]. In our construction the role of $\mathcal{I}$ corresponds to the role of J_2 in Section 2.1 of Kummer's thesis; the role of $\mathcal{E} - \mathcal{I}$ corresponds to J_1. A journal version of Kummer's proof is available as [21]. Let ψ be a Ke-numbering. There is a recursive $\{0, 1\}$-valued function F such that

- $F(i, 0) = 0$ for all i;
- $(\forall^\infty t)\, [F(i, t) = 1]$ iff $(\forall j < i)\, [W_j^\psi \neq W_i^\psi]$ and $(\exists x)\, [x + 1 \in W_i^\psi \wedge x \notin W_i^\psi]$;

Now let

$$W_0^\eta = \mathbb{N};$$
$$W_{\langle i,t\rangle+1}^\eta = \begin{cases} W_i^\psi & \text{if } F(i,t)=0 \text{ and } F(i,s)=1 \text{ for all } s>t; \\ \{x : x < \langle i,t-1\rangle\} & \text{if } F(i,t)=1; \\ \{x : x < \langle i,s-1\rangle\} & \text{if } s \text{ is the least number with} \\ & s>t \text{ and } F(i,t)=F(i,s)=0. \end{cases}$$

Intuitively, for i being the minimal ψ-grammar for an r.e. language not in $\{\mathbb{N}\}\cup\mathcal{I}$, $\langle i,t\rangle+1$ is the (only) η-grammar for W_i^ψ, where t is the unique number such that $F(i,t)=0$ and $F(i,s)=1$ for all $s>t$. All the other η-grammars are for languages in $\{\mathbb{N}\}\cup\mathcal{I}$, where one makes sure that there is exactly one η-grammar for each of these languages.

It is easy to verify that η is a Friedberg numbering. Moreover, $W_j^\psi = W_r^\eta$ can be checked using oracle K as follows. As ψ is a Ke-numbering, one can find using the oracle K the minimal i with $W_j^\psi = W_i^\psi$. Then $W_i^\psi = W_r^\eta$ iff one of the following four conditions holds:

- $W_i^\psi = \mathbb{N}$ and $r=0$;
- $r=\langle k,t\rangle+1$, $F(k,t)=0$, $k=i$ and for all $s>t$, $F(i,t)=1$;
- $r=\langle k,t\rangle+1$, $F(k,t)=1$ and $W_i^\psi = \{x : x < \langle i,t-1\rangle\}$;
- $r=\langle k,t\rangle+1$, $F(k,t)=0$, $s=\min(\{u>t : F(k,u)=0\})$ exists and $W_i^\psi = \{x : x < \langle k,s-1\rangle\}$.

The k and t in the last three conditions are computed from r, thus these variables are not quantified. Hence each of the above conditions can be determined K-recursively. It also follows that one can find, using oracle K, for any given j the corresponding r with $W_r^\eta = W_j^\psi$ and for any given r the minimal i with $W_i^\psi = W_r^\eta$. Thus, the theorem follows. $\square$

Note that for Friedberg numberings, the grammar equivalence problem is recursive. Furthermore, as there is only one index per language, every learner which converges semantically to a language is already converging syntactically to the language; hence **FrTxtBc** = **FrTxtEx**. Theorem 9 implies that **KeTxtEx** = **FrTxtEx** as indices can be translated in the limit from a given Ke-numbering to a chosen Friedberg numbering. Theorem 21 below shows that **TxtEx** = **KeTxtEx**; note that the proof is delayed to that place as the theorem actually shows a bit more than just **TxtEx** = **KeTxtEx**. These two results together give the following as our first result. Here note that, for function learning, Freivalds, Kinber and Wiehagen [12] showed that every explanatorily learnable class of recursive functions is learnable in some Friedberg numbering.

Theorem 10. **TxtEx** $\subseteq$ **FrTxtEx**.

Note that Proposition 27 below shows that no single Friedberg numbering is enough to learn all the **TxtEx**-learnable classes.

4 Finite Learning

Freivalds, Kinber and Wiehagen [12] showed that in the context of learning recursive functions, every finitely learnable class of recursive functions can be learnt in some Friedberg numbering. In contrast, our next result shows that for **TxtFin**, requiring learning in some Friedberg numbering is restrictive. Note that the following result holds, even if one considers learnability of only infinite languages (which can be proved by easy cylinderification of the languages in the class considered in the following proof).

Theorem 11. **TxtFin** $\not\subseteq$ **FrTxtFin**.

Proof. Let $\mathcal{L} = \{L : (\forall x \in L)[W_x = L]\}$. Clearly, $\mathcal{L} \in$ **TxtFin**. Suppose by way of contradiction that M **TxtFin**-identifies $\mathcal{L}$ in Friedberg numbering ψ. Without loss of generality assume that M does not output more than one conjecture on any text. Then, by Smullyan's double recursion theorem [26], there exist distinct e_1, e_2 such that W_{e_1}, W_{e_2} may be defined as follows.

Let $W_{e_1} = \{e_1, e_2\}$ and $W_{e_2} = \{e_1, e_2\}$, if there exist τ_1, τ_2 such that content$(\tau_i) \subseteq \{e_i\}$, $M(\tau_1)\downarrow \neq ?, M(\tau_2)\downarrow \neq ?$ and $M(\tau_1)\downarrow \neq M(\tau_2)\downarrow$; otherwise, let $W_{e_1} = \{e_1\}$ and $W_{e_2} = \{e_2\}$. It is easy to verify that $W_{e_i} \in \mathcal{L}$. Furthermore, if for some p, M outputs either ? or p, on all sequences in $Seg(\{e_1\}) \cup Seg(\{e_2\})$, then clearly $W_{e_1} \neq W_{e_2}$ and thus M does not $\mathbf{TxtFin}_\psi$-identify $\mathcal{L}$. On the other hand, if there exist τ_1, τ_2 such that $\tau_i \in Seg(\{e_i\})$, $M(\tau_1)\downarrow \neq ?, M(\tau_2)\downarrow \neq ?$ and $M(\tau_1)\downarrow \neq M(\tau_2)\downarrow$, then $W_{e_1} = W_{e_2}$ and M does not $\mathbf{TxtFin}_\psi$-identify $\mathcal{L}$ (as ψ is a Friedberg numbering). In either case, M does not $\mathbf{TxtFin}_\psi$-identify $\mathcal{L}$. □

A learner is prudent [24] if it only outputs grammars (in a given numbering used as hypotheses space) for the languages it learns (according to a given criterion). We denote prudent learning by attaching "**Prudent**" to the name of the criteria. One can strengthen the above proof to show that **PrudentTxtFin** $\not\subseteq$ **FrTxtFin**. This can be done by using $\mathcal{L} = \{W_{e_1(M)}, W_{e_2(M)} : M$ is a learning machine$\}$, where $e_1(M)$ and $e_2(M)$ denote the values of e_1 and e_2 as in the proof above, obtained effectively from the learner M.

Remark 12. In contrast to Theorem 11, one can show that several natural classes are finitely learnable in Friedberg numberings. The main idea is to use the even indices to provide a one-one numbering of a natural class of sets and to use the odd indices to make a Friedberg numbering of all remaining r.e. sets. Hence, for every $n \in \mathbb{N}$, $\{S : card(S) = n\} \in$ **FrTxtFin**. Furthermore, $\{\{\langle i, j\rangle : j \in \mathbb{N}\} : i \in \mathbb{N}\} \in$ **FrTxtFin**. Another natural class in **FrTxtFin** is $\{S : (\exists i)\,[S \subseteq \{\langle i, j\rangle : j \in \mathbb{N}\}$ and $card(S) = f(i)]\}$ for some recursive function f where only non-empty sets S are considered.

Our next result gives a characterization of **FrTxtFin**-learning for uniformly recursively enumerable classes.

Theorem 13. *A recursively enumerable class is in* **FrTxtFin** *iff it is one-one recursively enumerable and in* **TxtFin**.

Proof. Suppose $\mathcal{L}$ is r.e. and $\mathcal{L} \in \mathbf{FrTxtFin}$. Let M and Friedberg numbering ψ be such that $\mathcal{L} \subseteq \mathbf{TxtFin}_\psi(M)$. If $\mathcal{L}$ is finite, then the theorem immediately follows. So assume $\mathcal{L}$ is infinite. Let red be a recursive function such that $W_i^\psi = W_{red(i)}$, for all i. Let

$$S = \{red(i) : (\exists L \in \mathcal{L})(\exists \sigma \in Seg(L))[M(\sigma) = i]\}.$$

Let $h(j)$ denote the $(j+1)$-th element in some one-one enumeration of S. It is easy to verify that h witnesses that $\mathcal{L}$ is one-one recursively enumerable.

Now suppose $\mathcal{L}$ is one-one recursively enumerable and $\mathcal{L} \in \mathbf{TxtFin}$ as witnessed by M. Without loss of generality assume $\mathcal{L}$ is infinite. Let h be such that $\mathcal{L} = \{W_{h(i)} : i \in \mathbb{N}\}$ and, for all different i, j, $W_{h(i)} \neq W_{h(j)}$. Without loss of generality assume that M only outputs conjectures of form $h(j)$ on any input (whether from or outside the class $\mathcal{L}$).

Before defining the numbering ψ, we need to introduce an auxiliary function F which converges to 1 on minimal indices of non-members of $\mathcal{L} \cup \mathcal{I} \cup \{\mathbb{N}\}$ and outputs infinitely many zeroes on other inputs. More precisely, there is a $\{0,1\}$-valued recursive function F satisfying the following requirements:

- $F(i, 0) = 0$ for all i;
- $(\forall^\infty t)\,[F(i,t) = 1]$ iff $(\forall j < i)\,[W_j \neq W_i]$ and $(\exists x)\,[x+1 \in W_i \wedge x \notin W_i]$ and either $(\forall \sigma \in Seg(W_i))\,[M(\sigma) = ?]$ or $(\exists \sigma \in Seg(W_i))\,[M(\sigma) \neq ? \wedge W_{M(\sigma)} \neq W_i]$.

It is easy to verify that the second condition is a Σ_2 condition. Hence such a function F exists. Now the numbering ψ is defined as follows.

- $W_{3e}^\psi = W_{h(e)}$.
- $W_{3\langle i,t\rangle+1}^\psi = W_i$, if $F(i,t) = 0$ and for all $s > t$, $F(i,s) = 1$. Otherwise, $W_{3\langle i,t\rangle+1}^\psi$ will be spoiled and becomes some set from $\mathcal{I}$ not assigned to any other value.
- W_{3e+2}^ψ is either $\mathbb{N}$ or a member of $\mathcal{I}$.

We assume that the W_{3e+1}^ψ which are spoiled and W_{3e+2}^ψ together enumerate $\mathcal{I} \cup \{\mathbb{N}\}$ in one-one fashion (except for the unique element of $\mathcal{I} \cup \{\mathbb{N}\}$, if any, which belongs to $\mathcal{L}$).

It is now easy to verify that ψ is a Friedberg numbering and one can $\mathbf{TxtFin}_\psi$-identify $\mathcal{L}$ by outputting $3e$, whenever M outputs $h(e)$. $\square$

The above does not give a characterization of **FrTxtFin**, as the following theorem shows that there does exist a class in **FrTxtFin** which is not contained in any **TxtFin**-learnable recursively enumerable class.

Theorem 14. *There exists a class $\mathcal{L} \in$ **FrTxtFin** which is not contained in any r.e. class in* **TxtFin**.

Proof. Let $\mathcal{H}_e = \{W_i : i \in W_e\}$ denote the e-th recursively enumerable class. Let

$$L_e = \begin{cases} \{\langle e,1\rangle\} & \text{if there exists a } j \in W_e \text{ such that} \\ & \{\langle e,0\rangle, \langle e,1\rangle\} \subseteq W_j; \\ \{\langle e,0\rangle, \langle e,1\rangle\} & \text{otherwise.} \end{cases}$$

Let $\mathcal{L} = \{L_e : e \in \mathbb{N}\}$. On one hand one can show that $\mathcal{L}$ is not contained in any r.e. class in **TxtFin**: If $L_e = \{\langle e, 1\rangle\}$, then $\mathcal{H}_e$ contains a proper superset of L_e and is either not learnable or does not contain L_e; if $L_e = \{\langle e, 0\rangle, \langle e, 1\rangle\}$, then $\mathcal{H}_e$ does not contain L_e by the condition to choose L_e. Hence in each case, either $\mathcal{H}_e$ is not **TxtFin**-learnable or does not contain L_e.

On the other hand, it is easy to construct a Friedberg numbering ψ where the ψ-grammars for sets containing at most two elements can be effectively found from the set. Now consider the learner which outputs a ψ-grammar for $\{\langle e, 0\rangle, \langle e, 1\rangle\}$, if it sees $\langle e, 0\rangle$ in the input. The learner outputs a ψ-grammar for $\{\langle e, 1\rangle\}$, if it sees $\langle e, 1\rangle$ in the input and it can verify in time within the length of the input that $\{\langle e, 0\rangle, \langle e, 1\rangle\} \subseteq W_j$, for some $j \in W_e$. It is easy to verify that the above learner $\mathbf{TxtFin}_\psi$-identifies $\mathcal{L}$. □

In contrast to this, finite learning is preserved when all Ke-numberings are permitted as hypotheses spaces.

Theorem 15. $\mathbf{TxtFin} \subseteq \mathbf{KeTxtFin}$.

Proof. Suppose a **TxtFin**-learner M is given. Without loss of generality assume that if M outputs a conjecture on some text for L, then it outputs a conjecture on all texts for L.

Before defining the numbering ψ, we need to introduce an auxiliary function F which converges to 1 on minimal indices of non-members of $\mathbf{TxtFin}(M)$ and outputs infinitely many zeroes on other inputs. More precisely, there is a $\{0, 1\}$-valued recursive function F satisfying the following requirements:

- $F(i, 0) = 0$ for all i;
- $(\forall^\infty t)\,[F(i, t) = 1]$ iff $(\forall j < i)\,[W_j \neq W_i]$ and either $(\forall \sigma \in \mathit{Seg}(W_i))\,[M(\sigma) = ?]$ or $(\exists \sigma \in \mathit{Seg}(W_i))\,[M(\sigma) \neq ? \wedge W_{M(\sigma)} \neq W_i]$.

It is easy to verify that the second condition is a Σ_2 condition. Hence such a function F exists. Now the numbering ψ is defined as follows.

Let $W^\psi_{2\langle i,t\rangle} = W_i$, if $F(i, t) = 0$ and $F(i, t') = 1$ for all $t' > t$. $W^\psi_{2\langle i,t\rangle}$ is a finite subset of W_i otherwise.

For defining W^ψ_{2i+1}, let $R_s(i, j)$ be true iff $i \leq s$ and there exists a σ such that $|\sigma| \leq s$, content$(\sigma) \subseteq W_{i,s}$ and $M(\sigma) = j$. Let R^*_s be transitive closure of R_s.

Furthermore, let $W^\psi_{2i+1} = \bigcup_{s \in S_i}[\bigcup_{j : R^*_s(i,j)} W_{j,s}]$, where $S_i = \{s : R_s(i, i)$ and $(\exists t > s)\,(\forall j, j')$ $[(R^*_s(i, j) \wedge R^*_s(i, j')) \Rightarrow W_{j,s} \subseteq W_{j',t}]\}$.

Now, $2\langle i, j\rangle$ and k are equivalent ψ-grammars iff $2\langle i, j\rangle = k$ or both $W^\psi_{2\langle i,j\rangle}$ and W^ψ_k are finite and equal.

Furthermore, $2i+1$ and $2j+1$, where $i \neq j$, are equivalent ψ-grammars iff for some s, $R^*_s(i, j)$ and $R^*_s(j, i)$ holds and $s \in S_i \cap S_j$ or both W^ψ_{2i+1} and W^ψ_{2j+1} are finite and equal. Thus ψ is a Ke-numbering.

Also, one can **TxtFin**-identify $\mathbf{TxtFin}(M)$ in the numbering ψ by outputting $2M(\sigma) + 1$, on any input σ. □

5 Explanatory Learning with Additional Constraints

A learner is said to be *confident* [24] if it converges on all input texts, irrespective of whether the text is for a language in the class to be learnt or not. We denote confident learning by attaching "**Conf**" to the name of the criteria. The following theorem shows that confident learning in some Friedberg numbering can be achieved for every confident learnable class.

Theorem 16. **ConfTxtEx** = **ConfFrTxtEx**.

Proof. It suffices to show **ConfTxtEx** $\subseteq$ **ConfFrTxtEx**. Suppose M is a confident **TxtEx**-learner for $\mathcal{L}$. Without loss of generality assume that M is order independent.

Let $\mathcal{L}' = \{W_j$: there exists a least stabilizing sequence σ for M on W_j and it satisfies $M(\sigma) = j\}$. Note that $\mathcal{L} \subseteq \mathcal{L}'$ and M **TxtEx**-identifies $\mathcal{L}'$. By Theorem 10 there exists a Friedberg numbering η and a learner M' which $\mathbf{TxtEx}_\eta$-identifies $\mathcal{L}'$.

Define M'' as follows. $M''(T)$ searches for the least stabilizing sequence σ for M on content(T). Let $j = M(\sigma)$. M'' then searches for least stabilizing sequence τ for M on W_j. Note that both these searches stabilize as M is a confident learner. If $\sigma = \tau$, then $M''(T)$ converges to $M'(T')$, where T' is the canonical text for W_j. Otherwise $M''(T)$ converges to 0. It is easy to verify that M'' $\mathbf{TxtEx}_\eta$-identifies $\mathcal{L}'$ and M'' is confident. □

Even though every class which is Confidently learnable can be learnt in Friedberg numberings, there is still a subtle difference between learning in Friedberg numberings and acceptable numberings.

Remark 17. Let $\mathcal{L}_1 = \{L : L \neq \emptyset$ and $W_{\min(L)} = L\}$. Let $\mathcal{L}_2 = \{L : card(L) \geq 2$ and $W_{\min(L-\{\min(L)\})} = L\}$. It is easy to see that both $\mathcal{L}_1$ and $\mathcal{L}_2$ are in **ConfTxtEx**. However, $\mathcal{L}_1 \cup \mathcal{L}_2 \notin$ **TxtEx** as can be shown by using the idea of the proof of Case [9] that $\mathbf{TxtFEx}_2 \not\subseteq$ **TxtEx** (here $\mathbf{TxtFEx}_2$ learning allows a learner to eventually vacillate among up to 2 grammars for the language being learnt — we refer the reader to [9] for details). So **ConfTxtEx** is not closed under union for acceptable numberings.

However, confident learning is closed under union, if a Friedberg numbering or Ke-numbering is used.

Proposition 18. *Suppose ψ is a Ke-numbering and suppose that $\mathcal{L}_1, \mathcal{L}_2 \in \mathbf{ConfTxtEx}_\psi$. Then $\mathcal{L}_1 \cup \mathcal{L}_2 \in \mathbf{ConfTxtEx}_\psi$.*

Proof. Suppose M_1, M_2 witness that $\mathcal{L}_1, \mathcal{L}_2 \in \mathbf{ConfTxtEx}_\psi$, respectively. Furthermore, there exists a limit-recursive function F which computes a value $F(i,j)$ such that $F(i,j) \in W_i^\psi \triangle W_j^\psi$ whenever $W_i^\psi \neq W_j^\psi$. Note that F always converges, even if the two sets are equal; such an F exists because ψ is a Ke-numbering. Let $(F_n)_{n\in\mathbb{N}}$ be a recursive approximation to F. Define a new learner M on a text T as follows.

Let T and n be given. Let $x = F_n(M_1(T[n]), M_2(T[n]))$.
If $x \in W_{M_1(T[n]),n} \Leftrightarrow x \in \text{content}(T[n])$,
then $M(T[n]) = M_1(T[n])$,
else $M(T[n]) = M_2(T[n])$.

In the limit, M_1 converges on T to some index i and M_2 to some index j. Furthermore, $\lim_{n\to\infty} F_n(i,j)$ exists and is some value x. If $x \in W_i \Leftrightarrow x \in \text{content}(T)$, then M converges to i else M converges to j. In the case that $W_i = W_j$, it does not matter which choice M takes. In case $W_i \neq W_j$, then $x \in W_i \Leftrightarrow x \notin W_j$ and $M(T)$ converges to i (respectively, $M(T)$ converges to j) if $x \in \text{content}(T) \Leftrightarrow x \in W_i$ (respectively, $x \in \text{content}(T) \Leftrightarrow x \in W_j$). It follows that M confidently $\mathbf{TxtEx}_\psi$-identifies $\mathcal{L}_1 \cup \mathcal{L}_2$. $\square$

In contrast to confidence, several other properties do not preserve their full learning power when using Friedberg numberings instead of Gödel numberings as hypotheses spaces.

A learner is said to be *U-shaped on* L (see [3, 7, 8]), if on some text T for L, for some n, m, k with $n < m < k$, $M(T[n])$ and $M(T[k])$ are grammars for L (in the numbering being used as hypotheses space), but $M(T[m])$ is not a grammar for L. A learner is said to be *non U-shaped on* L if it is not U-shaped on L. A learner **NUShI**-identifies a class $\mathcal{L}$ if it **I**-identifies $\mathcal{L}$ and is non U-shaped on each $L \in \mathcal{L}$.

The following theorem shows that even simple classes such as $\mathcal{F}$, the class of all finite sets, fail to be **NUShTxtEx**-identified in Friedberg numberings.

Theorem 19. $\mathcal{F} \notin$ **NUShFrTxtEx**.

Proof. Suppose by way of contradiction that M witnesses $\mathcal{F} \in \mathbf{NUShTxtEx}_\eta$, where η is a Friedberg numbering. Thus, for all σ, if $M(\sigma) = i$ and $\text{content}(\sigma) \subset W^\eta_{M(\sigma)}$, then W^η_i is infinite (otherwise, M is U-shaped on some text for $W^\eta_{M(\sigma)}$, as there exists a τ extending σ such that $\text{content}(\tau) = \text{content}(\sigma)$ and $M(\tau)$ is an η-grammar for $\text{content}(\sigma)$ and furthermore, there exists a γ extending τ such that $\text{content}(\gamma) = W^\eta_{M(\sigma)}$ and $M(\gamma)$ is a η-grammar for $W^\eta_{M(\sigma)}$). It is then easy to verify that W^η_i is infinite iff (a) there exists a σ such that $M(\sigma) = i$ and $\text{content}(\sigma) \subset W^\eta_i$; or (b) for all σ such that $M(\sigma) = i$, $\text{content}(\sigma) \not\subseteq W^\eta_i$.

This gives a Δ_2 procedure for enumerating all infinite r.e. sets, a contradiction to well known result [26]. $\square$

Conservative learning [2, 33] requires that a learner does not abandon a hypothesis which is consistent with the input seen so far. Strong monotonicity [18] is a requirement that learners always output larger and larger hypothesis: for all texts T and m, n with $m < n$, $W^\psi_{M(T[m])} \subseteq W^\psi_{M(T[n])}$ (where ψ is the numbering used as hypotheses space). Monotonicity is the related requirement that for all sets L in the class to be learnt, for all texts T for L and all m, n with $m < n$, $W^\psi_{M(T[m])} \cap L \subseteq W^\psi_{M(T[n])} \cap L$. The following result can be proven by the same idea as the above; namely the class of all infinite sets would be uniformly recursively enumerable if $\mathcal{F}$ would be learnable under one of these criteria.

Theorem 20. *The class $\mathcal{F}$ is not conservatively, prudently, monotonically or strong monotonically learnable in Friedberg numberings.*

However, prudence is not restrictive for Ke-numberings.

Theorem 21. **TxtEx** $\subseteq$ **PrudentKeTxtEx**.

Proof. Suppose a **TxtEx**-learner M is given. Without loss of generality assume that either M **TxtEx**-identifies $\mathbb{N}$ or M **TxtEx**-identifies each member of $\mathcal{I}$, the class of all initial segments of $\mathbb{N}$ (see [13]).

Let $F(\cdot,\cdot)$ be a recursive function such that $\lim_{t\to\infty} F(i,t)$ converges to σ, if σ is the least stabilizing sequence for M on W_i; $\lim_{t\to\infty} F(i,t)$ does not converge, if there exists no such σ.

Let $G(\cdot,\cdot)$ be a recursive function such that $\lim_{t\to\infty} G(i,t)$ converges to 1 iff i is the least φ-grammar for W_i; $\lim_{t\to\infty} G(i,t)$ does not converge if i is not the least φ-grammar for W_i.

By standard arguments, F and G as above exist. Let $Y = \mathbb{N}$ if M **TxtEx**-identifies $\mathbb{N}$. Otherwise, $Y = \emptyset$. Thus, M **TxtEx**-identifies $Y \cup S$, for each $S \in \mathcal{I}$. We define the W^ψ indexing as follows.

$W^\psi_{2\langle j,m,t\rangle} = W_j$, if the following properties hold for all $s \in \mathbb{N}$:

- $M(\sigma_m) = j$;
- if $s = t-1$, then $F(j,s) \neq F(j,t)$;
- if $s \geq t$, then $F(j,s) = \sigma_m$.

Otherwise, $W^\psi_{2\langle j,m,t\rangle} = Y \cup \{x : x < s\}$ for the least s where one of the above properties fails.

Intuitively, the above properties checked if $M(\sigma_m) = j$, σ_m is the least stabilizing sequence for M on W_j and t is the convergence point for $F(j,\cdot)$.

Let $W^\psi_{2\langle j,m,t\rangle+1} = W_j$, if the following properties hold for all $s \in \mathbb{N}$:

- if $s = t-1$, then $G(j,s) = 0$;
- if $s \geq t$, then $G(j,s) = 1$;
- if $m = 0$, then there exists an $s' > s$ such that $F(j,s') \neq F(j,s)$;
- if $m = \langle v,w\rangle + 1 \wedge s = v-1$, then $F(j,s) \neq F(j,v)$;
- if $m = \langle v,w\rangle + 1 \wedge s > v$, then $F(j,s) = F(j,v)$;
- if $m = \langle v,w\rangle + 1$, then there is an $s' \geq s$ such that $[w = \min(W_{M(F(j,v)),s'} \bigtriangleup W_{j,s'})]$.

Otherwise, $W^\psi_{2\langle j,m,t\rangle+1} = W_{j,s}$, for the least s for which one of the above properties fails.

Intuitively, the first two properties above check if $G(j,\cdot)$ converges to 1, with t being the convergence point for $G(j,\cdot)$. The third property checks, for $m = 0$, whether $F(j,\cdot)$ diverges. The fourth to sixth properties check, for $m = \langle v,w\rangle + 1$, whether v is the convergence point for $F(j,\cdot)$ and $w = \min(W_{M(F(j,v))} \bigtriangleup W_j)$.

Claim. (a) If M has a least stabilizing sequence on L which is also a locking sequence for M on L, then $2\langle j,m,t\rangle$ is a ψ-grammar for L, where $M(\sigma_m) = j$, and σ_m is the least stabilizing sequence for M on L and t is the convergence point for $F(j,\cdot)$.

(b) ψ is a universal numbering (though not acceptable).

(c) every infinite recursively enumerable language L, except possibly for $\mathbb{N}$, has exactly one ψ-grammar.

(d) $\mathbb{N}$ has exactly one ψ-grammar, except possibly for grammars of the form $2\langle j,m,t\rangle$ which eventually follow the otherwise-clause in the definition of W^ψ above.

(e) M has a least stabilizing sequence for each W^ψ_{2i} which is also a locking sequence for M on W^ψ_{2i}.

We now prove the claim and then continue with the main proof.

Part (a) follows from the definition of $W^{\psi}_{2\langle j,m,t\rangle}$.

For (b), suppose L is r.e., If M has a least stabilizing sequence on L, which is also a locking sequence for M on L, then part (a) gives a ψ-grammar for L.

Otherwise, let i be the least φ-grammar for L. Let t be the convergence point for $G(i,\cdot)$. If M does not have a least stabilizing sequence on L, then $2\langle i,0,t\rangle+1$ is the ψ-grammar for L. Otherwise, let v be the convergence point of $F(i,\cdot)$. Let $w=\min(W_{M(\sigma)}\triangle W_j)$, where $\sigma=F(i,v)$. Then, $2\langle i,\langle v,w\rangle+1,t\rangle+1$ is a ψ-grammar for L.

For (c) note that if M has a least stabilizing sequence on L, which is also a locking sequence for M on L, then the proof of part (a) gives the only ψ-grammar for L. Otherwise the proof of part (b) gives the only ψ-grammar for L.

Part (d) can be proved similarly to part (c).

Part (e) follows directly from the definition of $W^{\psi}_{2\langle j,m,t\rangle}$: either σ_m is the least stabilizing sequence for M on W_j with t being convergence point for $F(i,\cdot)$ and $M(\sigma_m)=j$ (thus, $W^{\psi}_{2\langle j,m,t\rangle}=W_j$) or $W^{\psi}_{2\langle j,m,t\rangle}=Y\cup S$ for some $S\in\mathcal{I}$. Hence, (e) holds.

This completes the proof of the claim. Note that the ψ-grammars $2\langle j,m,t\rangle$, which follow the otherwise-clause in the definition, are either all grammars for $\mathbb{N}$ or are all ψ-grammars for finite sets. Thus, essentially Proposition 7 can be used to show that ψ is Ke-numbering. Using part (a) and (e) of the claim, prudent learning of $\mathbf{TxtEx}(M)$ follows easily as, on input σ, a learner can search for the least t and m such that the following two conditions hold:

- $M(\sigma_m)=M(\sigma_m\tau)$ for all τ such that $|\tau|\leq|\sigma|$ and $\tau\in Seg(\text{content}(\sigma))$,
- for all t' such that $t\leq t'\leq|\sigma|$, $F(M(\sigma_m),t')=\sigma_m$.

If t and m are found, then the learner outputs $2\langle M(\sigma_m),m,t\rangle$, else the learner outputs 0. Note that learner only uses grammars of form $2i$. It is easy to verify that M learns all languages of form $W^{\psi}_{2\langle j,m,t\rangle}$ (which, by part (a) of the above claim, includes all languages **TxtEx**-identified by M). Thus, M is a prudent learner. $\square$

Similar proofs can be used to show that non U-shaped learning and conservativeness are not restrictive for Ke-numberings.

Theorem 22. $\mathbf{TxtEx}\subseteq\mathbf{NUShKeTxtEx}$.

Proof-Sketch. The proof for this result is similar to the proof of Theorem 21. For this theorem, in the otherwise-clause of definition of $W^{\psi}_{2\langle j,m,t\rangle}$, we make $W^{\psi}_{2\langle j,m,t\rangle}$ to be outside the class being learnt (thus Y will be $\mathbb{N}$ if M does not **TxtEx**-identify $\mathbb{N}$; otherwise Y will be $\{x : x\leq\max(\text{content}(\tau))\}$, where τ is some fixed stabilizing sequence for M on $\mathbb{N}$). Other parts of the construction are as before. For identification, on input text T, at any stage n, one searches for the least sequence $\sigma_m\in Seg(\text{content}(T[n]))$ which satisfies

$$(\forall\tau\in Seg(\text{content}(T[n])) : |\tau|\leq n)[M(\sigma_m\tau)=M(\sigma_m)].$$

Then, the learner computes $j = M(\sigma_m)$ and the least $t \leq n$, such that $F(j, t') = \sigma_m$ for all t' with $t \leq t' \leq n$. If such m, t are not found, then the learner does not change its previous hypothesis and goes to stage $n + 1$. If such j, m, t are found, then the learner outputs $2\langle j, m, t\rangle$. The learner now goes to stage $n + 1$ only if it discovers that t is not the convergence point for $F(j, \cdot)$ or σ_m is not a stabilizing sequence for M on content(T). We omit the details. $\square$

Theorem 23. *Every class which can be conservatively* **TxtEx** *learnt can be conservatively learnt in some Ke-numbering.*

Proof-Sketch. This proof is also similar to the proof of Theorem 21. Here we do not assume that M identifies $\mathbb{N}$ or each member of $\mathcal{I}$ (as this cannot be assumed without loss of generality for conservative learning). However, that is fine as the Y is not needed in the modified construction here.

For this theorem, in the otherwise-clause of definition of $W^{\psi}_{2\langle j,m,t\rangle}$, we make $W^{\psi}_{2\langle j,m,t\rangle}$ to be $W_{j,s}$ for some s. Other parts of the construction are as before. For identification, on input text T, at any stage n, one searches for the least sequence $\sigma_m \in Seg(\text{content}(T[n]))$ which satisfies

$$(\forall \tau \in Seg(\text{content}(T[n])) : |\tau| \leq n)[M(\sigma_m \tau) = M(\sigma_m)].$$

Then, one computes $j = M(\sigma_m)$ and the least $t \leq n$, such that $F(j, t') = \sigma_m$ for all t' with $t \leq t' \leq n$. If such m, t are not found, then the learner does not change its previous hypothesis and goes to stage $n+1$. If such m, t are found, then the learner outputs $2\langle j, m, t\rangle$. Note that, by conservativeness of M, if M learns the input language, then the input language cannot be proper subset of W_j and hence $W^{\psi}_{2\langle j,m,t\rangle}$. The learner now goes to stage $n + 1$ only if it discovers that (a) t is not the convergence point for $F(j, \cdot)$ and $W^{\psi}_{2\langle j,m,t\rangle}$ does not contain the input language (note that if t is not the convergence point for $F(j, \cdot)$, then $W^{\psi}_{2\langle j,m,t\rangle}$ would be made finite by otherwise-clause eventually; thus one can eventually discover if $W^{\psi}_{2\langle j,m,t\rangle}$ does not contain the input language) or (b) σ_m is not a stabilizing sequence for M on content(T) (in which case, by conservativeness of M, W_j and thus $W^{\psi}_{2\langle j,m,t\rangle}$ does not contain the input segment, as seen at the time when it is discovered that σ_m is not a stabilizing sequence for M). We omit the details. $\square$

Remark 24. An iterative learner [28, 29] does not remember its history, but bases its conjecture on just the latest input and its previous conjecture. The proof of Theorem 19 can be easily modified to show that $\mathcal{F}$ cannot be iteratively learnt in any Friedberg numbering. It is open at present whether every iteratively **TxtEx**-learnable class can be learnt iteratively in some Ke-numbering.

A learner is said to be *consistent* [1, 4, 32] if for all σ, content$(\sigma) \subseteq W^{\psi}_{M(\sigma)}$, where ψ is the numbering used for hypotheses space. There have been three different versions of consistency studied in the literature. The notion considered here is often referred to as **TCons** (see [32]) where the "T" indicates that the learner has to be consistent on all total functions. **RCons** (see [19]) refers to consistent learning when the learners are total, but may not be consistent on inputs outside the class. In **Cons** learning (see [4]) the requirement is further relaxed to allow the learners to be partial: the learner may be defined and consistent only on inputs from the class being learnt. Theorem 26 can be extended to **Cons**, too. We do not yet know if the result extends to **RCons**.

Remark 25. For every $n \in \mathbb{N}$, there exists a Friedberg numbering η and a prudent, strongly monotonic and consistent learner M which $\mathbf{TxtEx}_\eta$-identifies $\{S : card(S) \leq n\}$.

Theorem 26. *Every consistently learnable class can be learnt consistently in some Friedberg numbering.*

Proof. Suppose M consistently $\mathbf{TxtEx}$-identifies $\mathcal{L}$ in the acceptable numbering φ. Without loss of generality assume that either M $\mathbf{TxtEx}$-identifies $\mathbb{N}$ or M $\mathbf{TxtEx}$-identifies all members of $\mathcal{I}$. Let F, G and ψ be as defined in the Proof of Theorem 21. Then, $W^\psi_{2\langle j,m,t\rangle} = W_j$, if $M(\sigma_m) = j$, $F(j,\cdot)$ converges to σ_m and t is the convergence point of $F(j,\cdot)$.

Let η be a Friedberg numbering such that $\psi \leq^K \eta$ (such η exists by Theorem 9). Let H be a recursive function such that for all i, $\lim_{s\to\infty} H(i,s)\!\downarrow$ and is a η-grammar for W^ψ_i. Thus, either $H(2\langle j,m,t\rangle, s)$ is an η-grammar for W_j, or $M(\sigma_m) \neq j$ or $F(j,t) \neq \sigma_m$, or t is not the convergence point for $F(j,\cdot)$ or $H(2\langle j,m,t\rangle, s') \neq H(2\langle j,m,t\rangle, s)$, for some $s' \geq s$. We define M' as follows.

- $M'(\sigma)$ first determines $j = M(\sigma)$ and the least m such that $\sigma_m \in Seg(\text{content}(\sigma))$ and $M(\sigma_m) = M(\tau)$ holds for all $\tau \in Seg(\text{content}(\sigma))$ satisfying $|\tau| \leq |\sigma|$ and $\sigma_m \subseteq \tau$.
- If $M(\sigma_m) \neq j$ or $F(j,|\sigma|) \neq \sigma_m$, then $M'(\sigma)$ outputs an arbitrary η-grammar i such that $W^\eta_i \supseteq \text{content}(\sigma)$.
- Otherwise, M' computes least t such that $F(j,t') = \sigma_m$, for $t \leq t' \leq |\sigma_m|$. M' then waits until one of the following conditions hold:
  (a) $W^\eta_{H(2\langle j,m,t\rangle,|\sigma|)}$ enumerates content(σ);
  (b) a $t' \geq t$ is found such that $F(j,t') \neq \sigma_m$;
  (c) a $s' \geq |\sigma|$ is found such that $H(2\langle j,m,t\rangle, s') \neq H(2\langle j,m,t\rangle, |\sigma|)$.
- In case (a), M' outputs $H(2\langle j,m,t\rangle, |\sigma|)$.
- In case (b) or (c), M' outputs an arbitrary η-grammar i such that $W^\eta_i \supseteq \text{content}(\sigma)$.

It is easy to see that M' is defined on all inputs as either σ_m is not the least stabilizing sequence for M on W_j or t (as in the definition of M') is not the convergence point of $F(j,\cdot)$ or $H(2\langle j,m,t\rangle,|\sigma|) \neq \lim_{s'\to\infty} H(2\langle j,m,t\rangle, s')$, or $H(2\langle j,m,t\rangle,|\sigma|)$ is an η-grammar for W_j and thus $W^\eta_{H(2\langle j,m,t\rangle,s)}$ contains content(σ), as M is consistent.

Thus, it is easy to verify that M' is consistent (for numbering η as hypotheses space), and M' on any text T for $L \in \mathcal{L}$ converges to $\lim_{s'\to\infty} H(2\langle j,m,t\rangle, s')$, where σ_m is the least stabilizing sequence for M' on L, $M(\sigma_m) = j$ and t is the convergence point for $F(j,\cdot)$. It follows that M' $\mathbf{TxtEx}_\eta$-identifies $\mathcal{L}$. $\square$

6 Learning with Respect to a Fixed Friedberg Numbering

We now investigate how powerful it is to learn with respect to one fixed Friedberg numbering. While $\mathbf{TxtEx} = \mathbf{TxtEx}_\varphi$ for every acceptable numbering φ, there is no optimal Friedberg numbering in this sense. This result can also be shown using the result of [12] that for every Friedberg numbering η (for partial functions), one can find an explanatory learnable class of functions, which is not explanatory learnable using η as hypothesis space. Theorem 28 and Remark 29 below show that there is an adversary Friedberg numbering ψ such that $\mathbf{TxtEx}_\psi \subseteq \mathbf{TxtEx}_\eta$ for every universal numbering η. This is language learning counterpart of the result from [12] that, for function learning, there exists a Friedberg numbering in which only finite classes of recursive functions can be learnt.

Proposition 27. *Let η be a Ke-numbering and $\mathcal{L}_1, \mathcal{L}_2$ be as in Remark 17. Then either $\mathcal{L}_1 \notin \mathbf{TxtEx}_\eta$ or $\mathcal{L}_2 \notin \mathbf{TxtEx}_\eta$. In particular, $\mathbf{TxtEx} \neq \mathbf{TxtEx}_\eta$.*

Proof. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be as defined in Remark 17. Note that if $\mathcal{L}_i \in \mathbf{TxtEx}_\eta$, then $\mathcal{L}_i \in \mathbf{ConfTxtEx}_\eta$. To see this for $\mathcal{L}_1$, suppose M is a $\mathbf{TxtEx}_\eta$ learner for $\mathcal{L}_1$. Define M' as follows. On input text T, M' first finds $e = \min(\mathrm{content}(T))$ in the limit. Then, it determines, in the limit, if $e = \min(W_e)$. If not, then $M'(T)$ converges to 0. Otherwise, $M'(T)$ converges to $M(T')$, where T' is canonical text for W_e. It is easy to verify that M' is confident and $\mathbf{TxtEx}_\eta$-identifies $\mathcal{L}_1$.

Thus, if both $\mathcal{L}_1, \mathcal{L}_2$ belong to $\mathbf{TxtEx}_\eta$, then by Proposition 18, $\mathcal{L}_1 \cup \mathcal{L}_2 \in \mathbf{ConfTxtEx}_\eta$, a contradiction to Remark 17. $\Box$

Theorem 28. *There exists a Friedberg numbering ψ such that every class in $\mathbf{TxtEx}_\psi$ contains only finitely many infinite languages.*

Proof. Let ϑ be a Friedberg numbering and $V_0, V_1, V_2, \ldots$ be a uniformly r.e. sequence of cofinite sets such that the function f mapping e to $\max(\overline{V_e})$ is total and satisfies $f(e) > \varphi_i^K(j)$ whenever $\varphi_i^K(j)$ is defined and $i, j \leq e$. Such a set V_e can be defined as follows. Let $g(i, j, s)$ be such that $\lim_{s\to\infty} g(i, j, s) = \varphi_i^K(j)$ (where the limit $\lim_{s\to\infty} g(i, j, s)$ does not exist, if $\varphi_i^K(j)$ is undefined). Now let $x \in V_e$ iff $x > 0$ and there are no $i, j \leq e$ such that $x = 1 + \max(\{i, j, s_{i,j}, g(i, j, s_{i,j})\})$, where $s_{i,j}$ is the convergence point of $g(i, j, \cdot)$, if any. Let $V_{i,x}$ denote V_i enumerated within x steps. Now define a numbering η such that

$$x \in W^\eta_{\langle i,j\rangle} \Leftrightarrow j \notin V_{i,x} \wedge j+1, j+2, \ldots, j+x \in V_i \wedge x \in W_i^\vartheta.$$

In other words, for each i and all $j \neq f(i)$, $W^\eta_{\langle i,j\rangle}$ is finite and $W^\eta_{\langle i,f(i)\rangle} = W_i^\vartheta$. As W^ϑ is a Friedberg numbering, one can conclude that in the numbering η, every infinite set has exactly one index. Finite sets may have several indices. Thus, η is a Ke-numbering by Proposition 7. Here note that, for infinite W_i^ϑ, only η-grammar for W_i^ϑ is $\langle i, f(i)\rangle$.

Then by Theorem 9 there is a Friedberg numbering ψ and a K-recursive function g such that, for all k, $W_k^\eta = W^\psi_{g(k)}$. Here note that ψ-grammar for W_i^ϑ is $g(\langle i, f(i)\rangle)$.

Now consider any class $\mathcal{L}$ in $\mathbf{TxtEx}_\psi$ and a witness M for this. One can define a partial K-recursive function h such that $h(i)$ is the index to which M converges to on the canonical text of W_i^ϑ; $h(i)$ is undefined if M does not converge on this canonical text. There is a partial-recursive function φ_e^K such that $\varphi_e^K(i)$ is the component j of the first pair $\langle k, j\rangle$ with $g(\langle k, j\rangle) = h(i)$

whenever $h(i)$ is defined. Now if $i > e$ and W_i^ϑ is infinite, then $\varphi_e^K(i)$ is either undefined or less than $f(i)$, hence $h(i) \neq g(\langle i, f(i)\rangle)$, the only ψ-grammar for W_i^ϑ. As a consequence, $\mathcal{L}$ contains only finitely many infinite sets. $\square$

Remark 29. If $\mathcal{L}$ is a **TxtEx**-learnable class containing only finitely many infinite languages, then $\mathcal{L}$ is in $\mathbf{TxtEx}_\eta$ for every universal numbering η.

Recall that $\mathcal{L}$ is inclusion free if there are no $L, H \in \mathcal{L}$ with $L \subset H$. Note that every finite inclusion-free class $\mathcal{L}$ is finitely learnable with respect to every universal numbering; the next result shows that for some numberings also the converse is true.

Proposition 30. *There is a Friedberg numbering ψ such that a class $\mathcal{L}$ is in $\mathbf{TxtFin}_\psi$ iff $\mathcal{L}$ is finite and inclusion-free.*

Proof. Let μ be a one-one numbering of all r.e. sets L with $card(\mathbb{N} - L) \neq 1$. Note that there exists such a numbering.

Let S be a simple set such that there is a (non-recursive) enumeration $a_0, a_1, a_2, \ldots$ of the elements of $\mathbb{N} - S$ such that

- for all n there is an m with $a_n = \langle n, m\rangle$ and
- for all n and $e < n$, if $\varphi_e(n)\downarrow$, then $a_n > \varphi_e(n)$.

Let $e_0, e_1, e_2, \ldots$ denote a recursive one–one enumeration of S. Then, for $e = \langle n, m\rangle$, define W_e^ψ as follows:

$$x \in W_e^\psi \Leftrightarrow (e \neq e_x) \wedge (e \in S \vee x \in W_n^\mu).$$

It is easy to verify that ψ is a Friedberg numbering. Now consider any finite learner M. Note that M $\mathbf{TxtFin}_\psi$-learns at most finitely many sets in $\{L : card(\mathbb{N} - L) = 1\}$, as any finite set belongs to almost all members of $\{L : card(\mathbb{N} - L) = 1\}$. Now we argue that M $\mathbf{TxtFin}_\psi$-learns at most finitely many languages of form W_n^μ. Define φ_e such that $\varphi_e(n)$ is the only grammar (if any) output by M on canonical text for W_n^μ. Now, for all $n > e$, $\varphi_e(n) < a_n$, which is the only ψ-grammar for W_n^μ. Thus, M can $\mathbf{TxtFin}_\psi$-identify W_n^μ, only for $n \leq e$. It follows that M $\mathbf{TxtFin}_\psi$-identifies only finitely many sets. Also clearly, if $L \subset H$ then no class containing both L and H can be $\mathbf{TxtFin}_\psi$-identified. $\square$

7 Behaviourally Correct Learning and Its Variants

TxtFEx-learning [9] denotes **TxtBc**-learning with the additional constraint that the learner outputs only finitely many distinct conjectures on a text for an input language from the class to be learnt. As $\mathbf{TxtFEx} \not\subseteq \mathbf{TxtEx}$, the next result establishes that behaviourally correct learning in Ke-numberings is more powerful than explanatory learning in acceptable numberings.

Theorem 31. $\mathbf{TxtFEx} \subseteq \mathbf{KeTxtBc}$.

Proof. One defines the following numbering ψ recursively. $W_{\langle i,n\rangle}^\psi$ is enumerated according to the following two steps:

1. Enumerate more and more of W_i until a $j < i$ is found such that $W_{j,n} \subseteq W_i$ and $W_{i,n} \subseteq W_j$.
2. If and when such a j as above is found, wait until it is found that $W^{\psi}_{\langle i,n\rangle}$ enumerated until now is contained in $W^{\psi}_{\langle j,n\rangle}$. If this never happens, then no further number is enumerated in $W^{\psi}_{\langle i,n\rangle}$. Otherwise, $W^{\psi}_{\langle i,n\rangle}$ follows $W^{\psi}_{\langle j,n\rangle}$.

First it is proven that ψ is a universal numbering. More precisely, one shows that, for all j and for all but finitely many n, $W^{\psi}_{\langle j,n\rangle} = W_j$.

To see this, consider for given j the set $S = \{i \le j : W_i = W_j\}$ and let m be so large that, for all $i \in S$, for all $k \le j$ such that $k \notin S$, either $W_{k,m} \not\subseteq W_i$ or $W_{i,m} \not\subseteq W_k$. It is then easy to see, by induction on elements i of S, that, for all $n \ge m$, $W_{\langle i,n\rangle} = W_i = W_j$. So ψ is a universal numbering.

Next, for given M, $\mathbf{TxtFEx}(M) \subseteq \mathbf{TxtBc}_{\psi}$, is shown. This holds as one can convert $M(\sigma)$ to $\langle M(\sigma), |\sigma|\rangle$ to achieve $\mathbf{TxtBc}_{\psi}$-learning of $\mathbf{TxtFEx}(M)$.

It remains to show that grammar equivalence problem for ψ is K-recursive. Note that for each $\langle i,n\rangle$, one can find in the limit $p(i,n)$ such that for some $i_0 = i > i_1 > \ldots > i_r = p(i,n)$, for $w < r$, $W^{\psi}_{\langle i_w,n\rangle}$ eventually follows $W^{\psi}_{\langle i_{w+1},n\rangle}$ and $W^{\psi}_{\langle i_r,n\rangle}$ does not follow any other grammar in the construction above.

Thus, determining equivalence of $W^{\psi}_{\langle i,n\rangle}$ and $W^{\psi}_{\langle j,m\rangle}$ is same as determining equivalence of $W^{\psi}_{\langle p(i,n),n\rangle}$ and $W^{\psi}_{\langle p(j,m),m\rangle}$. Now, $W^{\psi}_{\langle p(i,n),n\rangle}$ and $W^{\psi}_{\langle p(j,m),m\rangle}$ are same iff $W^{\psi}_{\langle p(i,n),n\rangle}$ and $W^{\psi}_{\langle p(j,m),m\rangle}$ are both finite and same or $p(i,j) = p(i,m)$ and $W^{\psi}_{\langle p(i,n),n\rangle}$ and $W^{\psi}_{\langle p(j,m),m\rangle}$ never leave step 1 in the construction above. Thus, one can solve grammar equivalence problem for ψ using oracle K. $\square$

Note that $\mathbf{FrTxtBc} = \mathbf{FrTxtFEx} = \mathbf{FrTxtEx}$ and $\mathbf{KeTxtFEx} = \mathbf{KeTxtEx}$. These equivalences, together with Theorem 31, give the following proper inclusion for behaviourally correct learning; unfortunately it is still unknown whether $\mathbf{KeTxtBc} = \mathbf{TxtBc}$.

Corollary 32. $\mathbf{FrTxtBc} \subset \mathbf{KeTxtBc}$.

Note that $\mathbf{TxtFEx} \subseteq \mathbf{KeTxtBc}$ by Theorem 31. Furthermore, $\mathbf{TxtFEx} \not\subseteq \mathbf{NUShTxtBc}$ [7]. Thus one obtains the following corollary.

Corollary 33. $\mathbf{NUShKeTxtBc} \subset \mathbf{KeTxtBc}$.

Recall that for Friedberg numberings explanatory and behaviourally correct learning coincide. Hence Theorem 19 also shows that $\mathcal{F} \notin \mathbf{NUShFrTxtBc}$. Furthermore, Theorem 22 shows that $\mathcal{F}$ is in $\mathbf{NUShKeTxtEx}$ and $\mathbf{NUShKeTxtBc}$. This establishes the first proper inclusion in the chain $\mathbf{NUShFrTxtBc} \subset \mathbf{NUShKeTxtBc} \subset \mathbf{NUShTxtBc}$; the second proper inclusion is proven in the next theorem.

Theorem 34. $\mathbf{NUShKeTxtBc} \subset \mathbf{NUShTxtBc}$.

Proof. For all e, define auxiliary sets $A_e = \{e\} \cup \{e + x : x \in W_e\}$ and $B_e = \{x : x \ge e\}$. The class $\mathcal{L} = \{L : L \ne \emptyset$ and $A_{\min(L)} \subseteq L$ and $card(L - A_{\min(L)}) < \infty\}$ then witnesses that the two learning criteria are different.

A learner, which, on input σ, outputs a φ-grammar for $\text{content}(\sigma) \cup A_{\min(\text{content}(\sigma))}$, can be easily seen to **NUShTxtBc**-identify $\mathcal{L}$.

Now suppose by way of contradiction that M $\mathbf{NUShTxtBc}_\psi$-identifies $\mathcal{L}$, where ψ is a Ke-numbering. We claim that the following three properties hold.

(P1) If there exists $\sigma \in Seg(B_e)$ such that $W^\psi_{M(\sigma)} = B_e$ and $B_e = \text{content}(\sigma) \cup A_e$, then A_e is cofinite.

(P2) If there exists $\sigma \in Seg(B_e)$ such that $W^\psi_{M(\sigma)} = B_e$ and $B_e \neq \text{content}(\sigma) \cup A_e$, then A_e is coinfinite.

(P3) If there does not exist a $\sigma \in Seg(B_e)$ such that $W^\psi_{M(\sigma)} = B_e$, then A_e is coinfinite.

To see (P1) and (P3), note that if W_e is cofinite, then $B_e \in \mathcal{L}$. Thus, there exists a σ such that $W^\psi_{M(\sigma)} = B_e$ and $\text{content}(\sigma) \cup A_e = B_e$.

To see (P2), suppose $\sigma \in Seg(B_e)$, $W^\psi_{M(\sigma)} = B_e$ and $\text{content}(\sigma) \cup A_e \neq B_e$. Suppose by way of contradiction that A_e is cofinite. Then, there exists a τ extending σ such that $\tau \in Seg(A_e \cup \text{content}(\sigma))$ and $M(\tau)$ is a ψ-grammar for $A_e \cup \text{content}(\sigma)$. Furthermore, there exists a τ' extending τ such that $\tau' \in Seg(B_e)$ and $M(\tau')$ is a ψ-grammar for B_e. But this contradicts non U-shaped learning of B_e by M. Thus, A_e is coinfinite.

However, (P1), (P2) and (P3) give us a Σ_3 procedure for checking whether W_e is coinfinite, a contradiction to a well known result [23]. (Note that one can first find a ψ-grammar p_e for B_e, using oracle for K'; then using Ke-numbering property of ψ, one can check using oracle for K' whether there exists a σ such that $M(\sigma)$ and p_e are equivalent. If so, then one can search for such a σ and then check whether $\text{content}(\sigma) \cup A_e = B_e$, using oracle for K'). $\square$

8 Partial Identification

Osherson, Stob and Weinstein [24, Exercise 7.5A] introduced the notion of partial identification. Here the learner, on any text T for a set L to be learnt, has to output infinitely often an index e with $W^\psi_e = \text{content}(T)$, while all other indices are output only finitely often. One can easily see that $\mathcal{E}$, the class of all recursively enumerable sets, is partially identifiable in an acceptable numbering. The same holds for Ke-numberings.

Theorem 35. *The class $\mathcal{E}$ can be partially identified using any given Ke-numbering as a hypotheses space.*

Proof. Given a Ke-numbering ψ, one can find out in the limit whether an index i is minimal for W^ψ_i. Hence a learner M partially identifying $\mathcal{E}$ can be built as follows. M, on a text T, outputs the index e at least n times iff there is a stage $s \geq n$ such that $W^\psi_{e,s} \cap \{0, 1, \ldots, n\} = \text{content}(T[s]) \cap \{0, 1, \ldots, n\}$ and e is believed to be a minimal ψ-index at stage s. It can be easily verified that the minimal correct index for $\text{content}(T)$ is output infinitely often, and other indices are output only finitely often. $\square$

Although $\mathcal{E}$ is partially identifiable relative to every Gödel numbering, every Friedberg numbering and every Ke-numbering, the next result shows that there are numberings relative to which only classes with finitely many infinite sets are partially identifiable. So Ke-numberings are well-suited for partial identification, compared to some other universal numberings.

Theorem 36. *There is a universal numbering η such that every class partially identifiable relative to η contains only finitely many infinite sets.*

Proof. Starting with a Friedberg-numbering ψ, one constructs a new numbering η as follows. Let $I_n = \{2^n - 1, 2^n, \ldots, 2^{n+1} - 2\}$.

Let C^K be the plain Kolmogorov complexity [22] relative to the oracle K. In the case that φ is a Kolmogorov numbering, one can define C^K by $C^K(x) = \min(\{n : (\exists y \in I_n)[\varphi_y^K(0) = x\})$. Let

$$A = \{m : (\exists n)\,[m \in I_n \wedge C^K(m) < n]\}$$

be the set of all C^K-compressible numbers. Note that A is a K-r.e. set and, for every n, $I_n \not\subseteq A$. Now define η such that, for every n and every $m \in I_n$: if $m \notin A$, then $W_m^\eta = W_n^\psi$, else W_m^η is a finite subset of W_n^ψ. Note that an infinite set W_n^ψ has exactly those η-indices m where $m \in I_n \wedge C^K(m) \geq n$.

Now suppose $\mathcal{L}$ is partially identified by a learner M. Let T_n be the canonical text for W_n^ψ, where W_n^ψ is infinite. Let

$$B = \{m : (\exists n)\,[m \in I_n \wedge M \text{ outputs } m \text{ on } T_n \text{ only finitely often}]\}.$$

If M partially identifies W_n^ψ, then there is an $m \in I_n$ such that $I_n - B = \{m\}$. Hence, there is a constant c such that $C^K(m) \leq C^K(n) + c$. So, for almost all n where M partially identifies W_n^ψ and W_n^ψ is infinite, there is a unique index $m \in I_n$ which is infinitely often output by M on T_n and which satisfies $m \in A$. Thus W_m^η is finite in contradiction to the assumption. It follows that $\mathcal{L}$ contains only finitely many infinite sets. □

Remark 37. Although for acceptable numberings and Ke-numberings the implication "$\mathcal{L}$ is behaviourally correct learnable $\Rightarrow$ $\mathcal{L}$ is partially identifiable" holds, this is not true for every universal numbering. Suppose $\mathcal{L}$ is a class with infinitely many languages which is learnable relative to a Friedberg numbering ψ. Let η be built from ψ as in the proof of Theorem 36. Then $\mathbf{TxtEx}_\psi \subseteq \mathbf{TxtFEx}_\eta$: Given a $\mathbf{TxtEx}_\psi$-learner M and considering any σ, the hypothesis $n = M(\sigma)$ is translated into an $m \in I_n$ which maximizes the cardinality of $W_{m,|\sigma|}^\eta$. One can show that, whenever M converges to n, then the new learner is eventually vacillating among those $m \in I_n$, which satisfy $W_m^\eta = W_n^\psi$. Hence $\mathcal{L} \in \mathbf{TxtFEx}_\eta$ and $\mathcal{L} \in \mathbf{TxtBc}_\eta$.

Furthermore, Theorem 36 could be slightly improved to show that some classes, with only one infinite set, are not partially identifiable with respect to some universal numbering η. However, one does not get a characterization (see also Theorem 28). Indeed, the criterion of being identifiable with respect to every universal numbering lies somewhere between the criterion from Theorem 28 and the one that a class has only finitely many infinite languages.

Acknowledgements. We thank the anonymous referees for helpful comments. We would like to thank Sergey Goncharov, Carl Jockusch, Bakhadyr Khoussainov, Martin Kummer, Steffen Lempp, Rolf Wiehagen, Jason Teutsch and Sandra Zilles for correspondence and comments.

References

1. Dana Angluin. Finding patterns common to a set of strings. *Journal of Computer and System Sciences*, 21:46–62, 1980.
2. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
3. Ganesh Baliga, John Case, Wolfgang Merkle, Frank Stephan and Rolf Wiehagen. When unlearning helps. Technical Report Technical Report TRA5/06, School of Computing, National University of Singapore, 2005.
4. Janis Bārzdiņš. Inductive inference of automata, functions and programs. In *International Congress of Mathematicians, Vancouver*, pages 771–776, 1974.
5. Lenore Blum and Manuel Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.
6. Manuel Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.
7. Lorenzo Carlucci, John Case, Sanjay Jain and Frank Stephan. U-shaped learning may be necessary. *Journal of Computer and System Sciences*, To appear.
8. Lorenzo Carlucci, Sanjay Jain, Efim Kinber and Frank Stephan. Variations on U-shaped learning. *Information and Computation*, 204(8):1264–1294, 2006.
9. John Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28(6):1941–1969, 1999.
10. John Case and Christopher Lynes. Machine inductive inference and language identification. In M. Nielsen and E. M. Schmidt, editors, *Proceedings of the 9th International Colloquium on Automata, Languages and Programming*, volume 140 of *Lecture Notes in Computer Science*, pages 107–115. Springer-Verlag, 1982.
11. Richard Friedberg. Three theorems on recursive enumeration. *Journal of Symbolic Logic*, 23(3):309–316, 1958.
12. 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.
13. Mark Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.
14. E. Mark Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
15. Sergey Goncharov. Nonequivalent constructivizations. In *Proceedings of the Mathematical Institute, Siberian Branch of Russian Academy of Sciences*. Nauka, Novosibirsk, 1982.
16. Sanjay Jain, Daniel Osherson, James Royer and Arun Sharma. *Systems that Learn: An Introduction to Learning Theory*. MIT Press, Cambridge, Mass., 2nd ed., 1999.
17. Sanjay Jain and Arun Sharma. Characterizing language learning in terms of computable numberings. *Annals of Pure and Applied Logic*, 84(1):51–72, 1997. Special issue on Asian Logic Conference, 1993.

18. Klaus-Peter Jantke. Monotonic and non-monotonic inductive inference. *New Generation Computing*, 8:349–360, 1991.
19. Klaus-Peter Jantke and Hans-Rainer Beick. Combining postulates of naturalness in inductive inference. *Journal of Information Processing and Cybernetics* (*EIK*), 17:465–484, 1981.
20. Martin Kummer. Beiträge zur Theorie der Numerierungen: Eindeutige Numerierungen. PhD Thesis, Karlsruhe, 1989.
21. Martun Kummer. An easy priority-free proof of a theorem of Friedberg. *Theoretical Computer Science*, 74:249–251, 1990.
22. Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer, Heidelberg, 1993.
23. Piergiorgio Odifreddi. *Classical Recursion Theory.* North-Holland, Amsterdam, 1989.
24. Daniel Osherson, Micheal Stob and Scott Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists.* MIT Press, 1986.
25. Daniel Osherson and Scott Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.
26. Hartley Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967. Reprinted by MIT Press in 1987.
27. Gisela Schäfer-Richter. Some results in the theory of effective program synthesis - learning by defective information. In W. Bibel and K. Jantke, editors, *Mathematical Methods of Specification and Synthesis of Software Systems, Wendisch-Rietz, GDR*, volume 215 of *Lecture Notes in Computer Science*, pages 219–225. Springer-Verlag, 1985.
28. Kenneth Wexler and Peter W. Culicover. *Formal Principles of Language Acquisition.* MIT Press, 1980.
29. Rolf Wiehagen. Limes-Erkennung rekursiver Funktionen durch spezielle Strategien. *Journal of Information Processing and Cybernetics* (*EIK*), 12:93–99, 1976.
30. Rolf Wiehagen. Characterization problems in the theory of inductive inference. In G. Ausiello and C. Böhm, editors, *Proceedings of the 5th International Colloquium on Automata, Languages and Programming*, volume 62 of *Lecture Notes in Computer Science*, pages 494–508. Springer-Verlag, 1978.
31. Rolf Wiehagen. A thesis in inductive inference. In J. Dix, K. Jantke and P. Schmitt, editors, *Nonmonotonic and Inductive Logic, 1st International Workshop*, volume 543 of *Lecture Notes in Artificial Intelligence*, pages 184–207. Springer-Verlag, 1990.
32. Rolf Wiehagen and Walter Liepe. Charakteristische Eigenschaften von erkennbaren Klassen rekursiver Funktionen. *Journal of Information Processing and Cybernetics* (*EIK*), 12:421–438, 1976.
33. Thomas Zeugmann and Steffen Lange. A guided tour across the boundaries of learning recursive languages. In K. Jantke and S. Lange, editors, *Algorithmic Learning for Knowledge-Based Systems*, volume 961 of *Lecture Notes in Artificial Intelligence*, pages 190–258. Springer-Verlag, 1995.