# Variations on U-shaped learning

Lorenzo Carlucci*[1] Sanjay Jain**[2], Efim Kinber[3], Frank Stephan***[4],

[1] Department of Computer and Information Sciences, University of Delaware, Newark, DE 19716-2586,USA and Dipartimento di Matematica, Università di Siena, Pian dei Mantellini 44, Siena, Italy. Email: carlucci5@unisi.it

[2] School of Computing, National University of Singapore, Singapore 117543. Email: sanjay@comp.nus.edu.sg

[3] Department of Computer Science, Sacred Heart University, Fairfield, CT 06432-1000, U.S.A. Email: kinbere@sacredheart.edu

[4] School of Computing and Department of Mathematics, National University of Singapore, Singapore 117543. Email: fstephan@comp.nus.edu.sg

**Abstract.** The paper deals with the following problem: is returning to wrong conjectures necessary to achieve full power of learning? Returning to wrong conjectures complements the paradigm of *U-shaped learning* [2, 6, 8, 20, 24] when a learner returns to old *correct* conjectures. We explore our problem for classical models of learning in the limit: **TxtEx**-learning – when a learner stabilizes on a correct conjecture, and **TxtBc**-learning – when a learner stabilizes on a sequence of grammars representing the target concept. In all cases, we show that, surprisingly, returning to wrong conjectures is sometimes necessary to achieve full power of learning. On the other hand it is not necessary to return to old "overgeneralizing" conjectures containing elements not belonging to the target language. We also consider our problem in the context of so-called *vacillatory* learning when a learner stabilizes to a finite number of correct grammars. In this case we show that both returning to old wrong conjectures and returning to old "overgeneralizing" conjectures is necessary for full learning power. We also show that, surprisingly, learners consistent with the input seen so far can be made *decisive* [2, 21] – they do not have to return to any old conjectures – wrong or right.

## 1 Introduction

*U-shaped learning* is a well-known pattern of learning behaviour in which the learner first learns the correct behaviour, then abandons it, and finally returns to the correct behaviour once again. The phenomenon of U-shaped learning has been observed by cognitive and developmental psychologists in many different cases of child development – such as language learning [6, 20, 24], understanding of temperature [24, 25] and face recognition [7]. The ability of models of human

* Supported in part by NSF grant number NSF CCR-0208616.
** Supported in part by NUS grant number R252–000–127–112.
*** Supported in part by NUS grant number R252–000–212–112.

learning to accommodate U-shaped learning progressively became one of the important criteria of their adequacy; see [20, 22] and the recent [26].

Cognitive and developmental psychology deals primarily with the problem of designing models of learning that adequately accommodate U-shaped behaviour. Baliga, Case, Merkle, Stephan and Wiehagen [2] who initiated study of U-shaped learning in the context of Gold-style algorithmic learning, asked a different question: is U-shaped behaviour really *necessary* for full learning power? In particular, they showed that U-shaped behaviour is avoidable for so-called **TxtEx**-learning (explanatory learning) – when the learner stabilizes in the limit on a single correct conjecture. This result contrasts with the result by Fulk, Jain and Osherson [13] who demonstrated that U-shaped learning is necessary for the full power of so-called **TxtBc**-learners (behaviourally correct learners) that stabilize on a (possibly infinite) sequence of different grammars representing the target language. In a sequel paper [8], Carlucci, Case, Jain and Stephan investigated U-shaped behaviour with respect to the model of vacillatory (or **TxtFex**) learning, where the learner is required to stabilize on a finite number of correct conjectures. Vacillatory learning, introduced by Case [9], forms a hierarchy of more and more powerful learning criteria between **TxtEx** and **TxtBc** identification. It was shown in [8] that forbidding U-shaped behaviour for vacillatory learners makes the whole hierarchy collapse to simple **TxtEx**-learning, i.e. nullifies the extra power of allowing vacillation between a finite number of conjectures.

U-shaped learning can be viewed as a special case of a more general pattern of learning behaviour, when a learner chooses a hypothesis, then abandons it, then returns to it once again[5]. If a learner returns to a correct conjecture that the learner has previously abandoned, it is, of course, dictated by the goal of correctly learning the target concept. On the other hand, when a learner returns to a previously abandoned *wrong* conjecture, this is not desirable if a learner wants to be efficient. In this paper, we study the following question: if and when returning to wrong conjectures is necessary for the full power of learnability? In particular, we consider

(a) a model in which a learner cannot return to a previously abandoned wrong conjecture;

(b) a model in which a learner cannot return to a previously abandoned conjecture that "overgeneralizes" – more precisely, contains elements not belonging to the target concept.[6] The latter model is motivated by the fact that over-

[5] These two meanings of U-shaped behaviour are explicitly distinguished at the beginning of [24], the main reference for the study of U-shaped behaviour.

[6] A more appropriate term for this could be "partial overgeneralization", since, strictly speaking, overgeneralization means the situation when the language generated by a conjectured grammar is a proper superset of the target language [1], rather than just containing elements not in the target language. Still, we opted to use just the word "overgeneralization" to emphasize the "over the board" aspects of such type of conjectures. Note that (by Theorem 8 and 22), using the usual definition of "overgeneralization" from [1] for **NOEx** and **NOBc**, does not change these classes. However, the class **NOFex** might change.

generalization is one of the major concerns in the study of learning behaviour [20].

We compare both models with regular types of learning in the limit and provide a full answer to the question when and how returning to wrong conjectures is necessary. The results that we obtained lead us to the following general conclusions. If we take **TxtEx** or **TxtBc** identification as a model of learning behaviour, then returning to previously abandoned wrong conjectures is necessary to achieve full power of learnability; however, it is not necessary to return to old "overgeneralizing" conjectures. On the other hand, for vacillatory identification, both returning to wrong conjectures and returning to "overgeneralizing" conjectures is necessary in a very strong sense: forbidding this kind of U-shapes collapses the whole **TxtFex**-hierarchy to simple **TxtEx**-learning. We compare more thoroughly these conclusions with results from [8] on returning to correct conjectures.

The paper has the following structure. Section 2 contains necessary notation and basic definitions. Section 3 contains definitions of all variants of previously known models of non U-shaped learning, as well as the models introduced in the present paper. In Section 4 we explore our variants of non U-shaped learning in the context of **TxtEx**-learning – when learners stabilize on one correct grammar for the target language. Firstly, we show, that, surprisingly, returning to wrong conjectures may be necessary for the full power of **TxtEx**-learning. To prove this result, we establish that learners not returning to wrong conjectures are as powerful as so-called *decisive* learners – the ones that never return to old conjectures (Theorem 7); decisive learners are known [2] to be generally weaker than general **TxtEx**-learners. On the other hand, any **TxtEx**-learner can be replaced by a learner not returning to "overgeneralizing" conjectures (Theorem 8).

In Section 5 we consider our two variants of non U-shaped learning in the context of *vacillatory* learning – when a learner stabilizes to a finite set of grammars describing the target language. We extend a result of Section 4 to show that vacillatory learners without returning to wrong conjectures do no better than just decisive **TxtEx**-learners. As for vacillatory learners not returning to "overgeneralizing" conjectures, they turn out to be doing no better than regular **TxtEx**-learners of this type. It was shown in [8] that the same collapse of the vacillatory hierarchy occurs when return to correct conjectures is forbidden. Thus, forbidding any of the three known variants of U-shaped behaviour nullifies the extra power of finite vacillation with respect to convergence to a single correct conjecture.

In Section 6 we explore our two variants of non U-shaped learning in the context of **TxtBc**-learnability – when learners stabilize on (potentially infinite) sequences of grammars correctly describing the target language. First, we show that there exist **TxtEx**-learnable classes of languages that cannot be learned without returning to wrong conjectures even by **TxtBc**-learners. From this Theorem and results from [2] it follows that **TxtBc**-learners not returning to correct conjectures sometimes do better than those never returning to wrong conjectures. On the other hand, we then show that, interestingly, **TxtBc**-learners not returning to wrong conjectures can sometimes do better than those never returning to right conjectures. Therefore these two forms of non U-shaped behaviour (avoiding to return to wrong conjectures and avoiding to return to correct conjectures) are of incomparable strength in the context of **TxtBc**-learning. The main result of this section is that, as in case of **TxtEx**-learnability, returning to old "overgeneralizing" conjectures can be circumvented: every **TxtBc**-learner can be replaced by one not returning to "overgeneralizing" conjectures (Theorem 22).

In Section 7 we discover a relationship between the strongest type of non U-shaped learners, that is decisive learners, and *consistent* learners [3, 21], whose conjectures are required to be consistent with the input data seen so far. Consistent learnability is known to be weaker than general **TxtEx**-learnability [3, 21]; moreover, sacrificing consistency, one can learn *pattern languages* faster than any consistent learner [18]. We show that consistent **TxtEx**-learners can be made consistent and decisive (Theorem 25). The result is surprising, since not returning to already used conjectures and being consistent with the input seen so far does not seem to be related – at least on the surface. On the other hand, some decisive learners cannot be made consistent (even if we sacrifice decisiveness).

## 2 Notation and Preliminaries

Any unexplained recursion theoretic notation is from [23]. The symbol $\mathcal{N}$ denotes the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$. The symbols $\emptyset$, $\subseteq$, $\subset$, $\supseteq$, and $\supset$ denote empty set, subset, proper subset, superset, and proper superset, respectively. Cardinality of a set $S$ is denoted by $\mathrm{card}(S)$. $\mathrm{card}(S) \leq *$ denotes that $S$ is finite. The maximum and minimum of a set are denoted by $\max(\cdot), \min(\cdot)$, respectively, where $\max(\emptyset) = 0$ and $\min(\emptyset) = \infty$. We let $\langle x, y \rangle = \frac{1}{2}(x + y)(x + y + 1) + y$, a standard pairing function.

By $\varphi$ we denote a fixed *acceptable* programming system for the partial computable functions mapping $\mathcal{N}$ to $\mathcal{N}$ [19, 23]. By $\varphi_i$ we denote the partial computable function computed by the program with number $i$ in the $\varphi$-system. The symbol $\mathcal{R}$ denotes the set of all recursive functions, that is total computable functions. By $\Phi$ we denote an arbitrary fixed Blum complexity measure [5, 15] for the $\varphi$-system.

By $W_i$ we denote $\mathrm{domain}(\varphi_i)$. That is, $W_i$ is then the recursively enumerable (r.e.) subset of $\mathcal{N}$ accepted by the $\varphi$-program $i$. Note that all acceptable numberings are isomorphic and that one therefore could also define $W_i$ to be the set generated by the $i$-th grammar. The symbol $\mathcal{E}$ will denote the set of all r.e. languages. The symbol $L$ ranges over $\mathcal{E}$. By $\overline{L}$, we denote the complement of $L$, that is $\mathcal{N} - L$. The symbol $\mathcal{L}$ ranges over subsets of $\mathcal{E}$. By $W_{i,s}$ we denote the set $\{x < s \mid \Phi_i(x) < s\}$.

We now present concepts from language learning theory. A *sequence* $\sigma$ is a mapping from an initial segment of $\mathcal{N}$ into $(\mathcal{N} \cup \{\#\})$. The empty sequence is denoted by $\Lambda$. The *content* of a sequence $\sigma$, denoted $\mathrm{content}(\sigma)$, is the set of natural numbers in the range of $\sigma$. The *length* of $\sigma$, denoted by $|\sigma|$, is the number of elements in $\sigma$. So, $|\Lambda| = 0$. For $n \leq |\sigma|$, the initial sequence of $\sigma$ of length $n$ is denoted by $\sigma[n]$. So, $\sigma[0]$ is $\Lambda$.

Intuitively, #'s represent pauses in the presentation of data. We let $\sigma$, $\tau$ and $\gamma$ range over finite sequences. We denote the sequence formed by the concatenation of $\tau$ at the end of $\sigma$ by $\sigma\tau$. Sometimes we abuse the notation and use $\sigma x$ to denote the concatenation of sequence $\sigma$ and the sequence of length 1 which contains the element $x$. SEQ denotes the set of all finite sequences. We let $\delta_0, \delta_1, \ldots$ denote a standard recursive 1–1 listing of all the finite sequences. We assume that $\max(\mathrm{content}(\delta_i)) \leq i$. We let $\mathrm{ind}(\sigma)$ denote $i$ such that $\delta_i = \sigma$.

A *text* $T$ for a language $L$ [14] is a mapping from $\mathcal{N}$ into $(\mathcal{N} \cup \{\#\})$ such that $L$ is the set of natural numbers in the range of $T$. $T(i)$ represents the $(i+1)$-th element in the text. The *content* of a text $T$, denoted by $\mathrm{content}(T)$, is the set of natural numbers in the range of $T$; that is, the language which $T$ is a text for. $T[n]$ denotes the finite initial sequence of $T$ with length $n$.

A *language learning machine from texts* [14] is an algorithmic device which computes a mapping from SEQ into $\mathcal{N}$. We note that, without loss of generality, for all criteria of learning discussed in this paper, except for consistent learning discussed in Section 7, a learner $\mathbf{M}$ may be assumed to be total.

We let $\mathbf{M}$ range over learning machines. $\mathbf{M}(T[n])$ is interpreted as the grammar (index for an accepting program) conjectured by the learning machine $\mathbf{M}$ on the initial sequence $T[n]$. We say that $\mathbf{M}$ converges on $T$ to $i$, (written: $\mathbf{M}(T)\downarrow = i$) iff $(\forall^\infty n)[\mathbf{M}(T[n]) = i]$.

There are several criteria for a learning machine to be successful on a language. Below we define some of them.

**Definition 1.** (a) [10, 14] $\mathbf{M}$ **TxtEx**-*identifies a text* $T$ just in case $(\exists i \mid W_i = \mathrm{content}(T))$ $(\forall^\infty n)[\mathbf{M}(T[n]) = i]$.
(b) [10] $\mathbf{M}$ **TxtBc**-*identifies a text* $T$ just in case $(\forall^\infty n)[W_{\mathbf{M}(T[n])} = \mathrm{content}(T)]$.
(c) [9] $\mathbf{M}$ $\mathbf{TxtFex}_a$-*identifies a text* $T$ just in case there exists a set $D$ such that $\mathrm{card}(D) \leq a$, $(\forall i \in D)[W_i = \mathrm{content}(T)]$ and $(\forall^\infty n)[W_{\mathbf{M}(T[n])} \in D]$.

Furthermore, for $\mathbf{I} \in \{\mathbf{TxtEx}, \mathbf{TxtBc}, \mathbf{TxtFex}_a\}$: $\mathbf{M}$ $\mathbf{I}$-*identifies an r.e. language* $L$ (written: $L \in \mathbf{I}(\mathbf{M})$) just in case $\mathbf{M}$ $\mathbf{I}$-identifies each text for $L$; $\mathbf{M}$ $\mathbf{I}$-*identifies a class* $\mathcal{L}$ *of r.e. languages* (written: $\mathcal{L} \subseteq \mathbf{I}(\mathbf{M})$) just in case $\mathbf{M}$ $\mathbf{I}$-identifies each language from $\mathcal{L}$; $\mathbf{I} = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{I}(\mathbf{M})]\}$.

[9–11] show that, $\mathbf{TxtEx} \subset \mathbf{TxtFex}_2 \subset \mathbf{TxtFex}_3 \subset \ldots \subset \mathbf{TxtFex}_* \subset \mathbf{TxtBc}$.

**Definition 2.** (a) [12] $\sigma$ is said to be a *stabilizing sequence* for $\mathbf{M}$ on $L$ iff $\mathrm{content}(\sigma) \subseteq L$, and for all $\tau \supseteq \sigma$ such that $\mathrm{content}(\tau) \subseteq L$, $\mathbf{M}(\sigma) = \mathbf{M}(\tau)$.

(b) [4] $\sigma$ is said to be a **TxtEx**-*locking sequence* for $\mathbf{M}$ on $L$ iff $\sigma$ is a stabilizing sequence for $\mathbf{M}$ on $L$, and $W_{\mathbf{M}(\sigma)} = L$.

(c) (Based on [4]) $\sigma$ is said to be a **TxtBc**-*locking sequence* for $\mathbf{M}$ on $L$ iff $\mathrm{content}(\sigma) \subseteq L$, and for all $\tau \supseteq \sigma$ such that $\mathrm{content}(\tau) \subseteq L$, $W_{\mathbf{M}(\sigma)} = L$.

If $\mathbf{M}$ **TxtEx**-identifies $L$, then there exists a **TxtEx**-locking sequence for $\mathbf{M}$ on $L$ [4]. Similar result holds for **TxtBc** and $\mathbf{TxtFex}_a$ criteria of learning.

Let $\mathrm{INIT} = \{L \mid (\exists i)[L = \{x \mid x \leq i\}]\}$. Let $INIT_k = \{x \mid x \leq k\}$.

## 3 Decisive, Non U-Shaped and related criteria of learning

Part (a) below gives the strongest type of non U-shaped behaviour – when a learner is not allowed to return to *any* old conjectures. Part (b) gives the definition of non U-shaped learning. Parts (c) and (d) give our two models of non U-shaped learning when a learner is not allowed to return to previously used wrong conjectures. 'NO' in part (d) stands for non-overgeneralizing.

**Definition 3.** (a) [21] $\mathbf{M}$ is *decisive* on text $T$, if there do not exist any $m, n, t$ such that $m < n < t$, $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[t])}$ and $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.
(b) [2] $\mathbf{M}$ is *non U-shaped* on text $T$, if there do not exist any $m, n, t$ such that $m < n < t$, $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[t])} = \text{content}(T)$ and $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.
(c) $\mathbf{M}$ is *Wr-decisive* on text $T$, if there do not exist any $m, n, t$ such that $m < n < t$, $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[t])} \neq \text{content}(T)$ and $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.
(d) $\mathbf{M}$ is *NO-decisive* on text $T$, if there do not exist $m, n, t$ such that $m < n < t$, $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[t])} \not\subseteq \text{content}(T)$ and $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.

Furthermore, $\mathbf{M}$ is decisive (non U-shaped, Wr-decisive, NO-decisive) on $L$ if $\mathbf{M}$ is decisive (non U-shaped, Wr-decisive, NO-decisive) on each text for $L$.
$\mathbf{M}$ is decisive (non U-shaped, Wr-decisive, NO-decisive) on $\mathcal{L}$ if $\mathbf{M}$ is decisive (non U-shaped, Wr-decisive, NO-decisive) on each $L \in \mathcal{L}$.

We now define the learning criteria formed by placing the various constraints described above on the learner. Note that the definition used for decisive learning is class version of decisive, that is decisiveness is required to hold only for texts for the languages in the class. We do this to make it consistent with the definitions of non $U$-shaped, WR-decisive and NO-decisive criteria, where only the class version seems sensible.

**Definition 4.** (a) [21] $\mathbf{M}$ **DecEx**-identifies $L$ (written: $L \in \mathbf{DecEx}(\mathbf{M})$), iff $\mathbf{M}$ **TxtEx**-identifies $L$, and $\mathbf{M}$ is decisive on $L$. $\mathbf{M}$ **DecEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **DecEx**-identifies each $L \in \mathcal{L}$. $\mathbf{DecEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{DecEx}(\mathbf{M})]\}$.
(b) [2] $\mathbf{M}$ **NUShEx**-identifies $L$ (written: $L \in \mathbf{NUShEx}(\mathbf{M})$), iff $\mathbf{M}$ **TxtEx**-identifies $L$, and $\mathbf{M}$ is non U-shaped on $L$. $\mathbf{M}$ **NUShEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **NUShEx**-identifies each $L \in \mathcal{L}$. $\mathbf{NUShEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NUShEx}(\mathbf{M})]\}$.
(c) $\mathbf{M}$ **WrEx**-identifies $L$ (written: $L \in \mathbf{WrEx}(\mathbf{M})$), iff $\mathbf{M}$ **TxtEx**-identifies $L$, and $\mathbf{M}$ is Wr-decisive on $L$. $\mathbf{M}$ **WrEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **WrEx**-identifies each $L \in \mathcal{L}$. $\mathbf{WrEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{WrEx}(\mathbf{M})]\}$.
(d) $\mathbf{M}$ **NOEx**-identifies $L$ (written: $L \in \mathbf{NOEx}(\mathbf{M})$), iff $\mathbf{M}$ **TxtEx**-identifies $L$, and $\mathbf{M}$ is NO-decisive on $L$. $\mathbf{M}$ **NOEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **NOEx**-identifies each $L \in \mathcal{L}$. $\mathbf{NOEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NOEx}(\mathbf{M})]\}$.

One can similarly define **DecI**, **WrI**, **NOI** and **NUShI**, for $\mathbf{I} \in \{\mathbf{Fex}_*, \mathbf{Bc}\}$. It is easy to verify that for all $a \in \mathcal{N} \cup \{*\}$ and $\mathbf{I} \in \{\mathbf{Ex}, \mathbf{Fex}_a, \mathbf{Bc}\}$, (a) $\mathbf{DecI} \subseteq \mathbf{WrI} \subseteq \mathbf{NOI} \subseteq \mathbf{I}$; (b) $\mathbf{DecI} \subseteq \mathbf{NUShI} \subseteq \mathbf{I}$.

## 4 Explanatory Learning

Our first goal is to show that, in the context of **TxtEx**-learnability, learners not returning to wrong conjectures do no better than decisive learners. To prove this, we first establish two lemmas. We omit the proof of Lemma 5.

**Lemma 5.** *Suppose there exists a finite set $A$ such that $\mathcal{L}$ does not contain any extension of $A$. Then $\mathcal{L} \in \mathbf{TxtEx} \Rightarrow \mathcal{L} \in \mathbf{DecEx}$.*

**Lemma 6.** *Suppose every finite set has at least two extensions in $\mathcal{L}$. Suppose $a \in \mathcal{N} \cup \{*\}$ and $\mathbf{I} \in \{\mathbf{Ex}, \mathbf{Fex}_a, \mathbf{Bc}\}$. Then, $\mathcal{L} \subseteq \mathbf{DecI}(\mathbf{M})$ iff $\mathcal{L} \subseteq \mathbf{WrI}(\mathbf{M})$.*

**Proof.** Suppose by way of contradiction that $\mathcal{L} \subseteq \mathbf{WrI}(\mathbf{M})$, $\mathcal{L} \not\subseteq \mathbf{DecI}(\mathbf{M})$. Thus, $\mathbf{M}$ is not decisive. Let $\tau_1 \prec \tau_2 \prec \tau_3$ be such that $W_{\mathbf{M}(\tau_1)} = W_{\mathbf{M}(\tau_3)} \neq W_{\mathbf{M}(\tau_2)}$. Let $L$ be an extension of content($\tau_3$) such that $W_{\mathbf{M}(\tau_1)} \neq L$ and $L \in \mathcal{L}$. Such an $L$ exists by the hypotheses on $\mathcal{L}$. Let $T$ be a text for $L$ extending $\tau_3$. Then $T$ witnesses that $\mathbf{M}$ does not $\mathbf{WrI}$-identify $\mathcal{L}$ since $\mathbf{M}$ returns to the wrong conjecture $W_{\mathbf{M}(\tau_1)}$ on text $T$. A contradiction. Lemma follows. ∎

Now we can establish one of our main results.

**Theorem 7.** $\mathbf{DecEx} = \mathbf{WrEx}$.

**Proof.** Suppose $\mathcal{L} \in \mathbf{WrEx}$. We consider the following cases.

Case 1: $\mathcal{L}$ contains at least two extensions of every finite set. Then by Lemma 6, $\mathcal{L}$ is in **DecEx**.

Case 2: Not Case 1. Let $A'$ be a finite set such that $\mathcal{L}$ contains at most one extension of $A'$.

Case 2.1: $\mathcal{N} \in \mathcal{L}$. Then by Proposition 17 in [2], we have that $\mathcal{L} \in \mathbf{DecEx}$.

Case 2.2: $\mathcal{N} \notin \mathcal{L}$. If $\mathcal{L}$ contains no extension of $A'$, then let $A = A'$. If $\mathcal{L}$ contains $L \neq \mathcal{N}$, $L \supseteq A'$, then let $A = A' \cup \{w\}$, where $w \notin L$. Now, $\mathcal{L}$ does not contain any superset of $A$. Thus, by Lemma 5, we have that $\mathcal{L} \in \mathbf{DecEx}$. ∎

As $\mathbf{DecEx} \subset \mathbf{TxtEx}$ [2], we conclude that some families of languages in **TxtEx** cannot be learned without returning to wrong conjectures.

However, if we allow to return to subsets of the target language (that is, wrong conjectures that do not overgeneralize), then all classes of languages in **TxtEx** become learnable, as the following result shows.

**Theorem 8.** $\mathbf{TxtEx} \subseteq \mathbf{NOEx}$.

**Proof.** Suppose $\mathcal{L} \in \mathbf{TxtEx}$.

If $\mathcal{N} \in \mathcal{L}$, then $\mathcal{L} \in \mathbf{DecEx}$ as shown by Baliga, Case, Merkle and Stephan [2]. So assume $\mathcal{N} \notin \mathcal{L}$. Let $\mathbf{M}$ be a machine such that, (I) $\mathbf{M}$ **TxtEx**-identifies $\mathcal{L} \cup \mathrm{INIT}$, (II) $\mathbf{M}$ is prudent[7] and (III) all texts for $L \in \mathcal{L} \cup \mathrm{INIT}$, start with a **TxtEx**-locking sequence for $\mathbf{M}$ on $L$. Note that Fulk [12] shows that this can be

[7] $\mathbf{M}$ is said to be prudent [21] iff $\mathbf{M}$ **TxtEx**-identifies every language $W_i$, such that $i$ is in the range of $\mathbf{M}$.

assumed without loss of generality. Also note that, for all $k$, if $\sigma$ is a stabilizing sequence for $\mathbf{M}$ on $INIT_k$, then $\text{content}(\sigma) = INIT_k$.

For a segment $\sigma$, let $f(\sigma) = \min(\mathcal{N} - \text{content}(\sigma))$. Let $valid = \{T[m] \mid m = 0 \text{ or } \mathbf{M}(T[m-1]) \neq \mathbf{M}(T[m])\}$. Let $consseq = \{T[m] \mid \text{content}(T[m]) \subseteq W_{\mathbf{M}(T[m])}\}$. Let $gram$ be a recursive function such that

$$W_{gram(T[m])} = \begin{cases} \emptyset, \\ \quad \text{if content}(T[m]) \not\subseteq W_{\mathbf{M}(T[m])}; \\ W_{\mathbf{M}(T[m])}, \\ \quad \text{if } T[m] \text{ is a stabilizing sequence for } W_{\mathbf{M}(T[m])}; \\ INIT_{\langle \text{ind(T[m])}, w\rangle}, \\ \quad \text{otherwise, for some } w \geq f(T[m]). \end{cases}$$

It is easy to verify that for $T[m] \in consseq$, $\text{content}(T[m]) \subseteq W_{gram(T[m])}$.

Define $\mathbf{M}'$ as follows. $\mathbf{M}'(T[n]) = gram(T[m])$, for the largest $m \leq n$, such that $T[m]$ is valid and $W_{\mathbf{M}(T[m]),n} \supseteq \text{content}(T[m])$ (there exists such an $m$, as $m = 0$ satisfies the constraints). Note that the mapping from $n$ to that $m$ for which $\mathbf{M}'(T[n]) = gram(T[m])$, is monotonically non-decreasing in $n$.

Now suppose $T$ is a text for $L \in \mathcal{L}$. We now show that if $W_{\mathbf{M}'(T[m'])} = W_{\mathbf{M}'(T[n'])} \neq W_{\mathbf{M}'(T[s'])}$, for $m' < s' < n'$, then $W_{\mathbf{M}'(T[m'])} \subseteq L$. So suppose $m', s', n'$ as above are given. Suppose $\mathbf{M}'(T[m']) = gram(T[m])$, $\mathbf{M}'(T[s']) = gram(T[s])$, and $\mathbf{M}'(T[n']) = gram(T[n])$. By monotonicity of $\mathbf{M}'$ mentioned above, $m' < s' < n'$ implies $m \leq s \leq n$. If $m = n$, then we are done, as $\mathbf{M}'(T[s'])$ would also be equal to $gram(T[m])$. So assume $m < n$. As $\text{content}(T[n]) \subseteq W_{gram(T[n])}$, and $T[n]$ is valid, we immediately have that $T[m]$ is not a stabilizing sequence for $\mathbf{M}$ on $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[n])} \supseteq \text{content}(T[n])$. Thus, $gram(T[m])$ follows the third clause in its definition. Since, $\langle \text{ind(T[m])}, \cdot\rangle \neq \langle \text{ind(T[n])}, \cdot\rangle$, for $m \neq n$, it follows that $gram(T[n])$ must follow the second clause, and thus $T[n]$ is a stabilizing sequence for $W_{\mathbf{M}(T[n])}$. As $W_{gram(T[m])}$ $(= W_{gram(T[n])})$ is in INIT, it follows that $\text{content}(T[n]) = W_{gram(T[n])}$ (since $\sigma$ being stabilizing sequence for $\mathbf{M}$ on $INIT_k$ implies that $\text{content}(\sigma) = INIT_k$). Thus, $W_{gram(T[m])} = W_{gram(T[n])} = \text{content}(T[n]) \subseteq L$.

It follows that $\mathbf{M}'$ **NOEx**-identifies $\mathcal{L}$. ∎

We now compare Theorems 7 and 8 with the following result about **NUSh**-learners from [2].

**Theorem 9.** [2] (a) **TxtEx** $\not\subseteq$ **DecBc**; (b) **TxtEx** $=$ **NUShEx**.

Thus, Theorem 7 implies that forbidding return to abandoned wrong conjectures is more restrictive than forbidding return to abandoned correct conjectures in the context of **TxtEx**-learning, while, from Theorem 8, the latter requirement is equivalent to forbidding to return to abandoned "overgeneralizing" conjectures. We summarize these observations in the following immediate corollary.

**Corollary 10.** **WrEx** $\subset$ **NUShEx** $=$ **NOEx**.

## 5 Vacillatory Learning

In this section we show that when returning to wrong conjectures is not allowed in vacillatory learning, then the vacillatory hierarchy $\mathbf{TxtFex}_1 \subset \mathbf{TxtFex}_2 \subset \ldots \subset \mathbf{TxtFex}_*$ collapses to $\mathbf{TxtFex}_1 = \mathbf{TxtEx}$, so that the extra learning power given by vacillation is lost. That the same collapse occurs when returning to correct abandoned conjectures is forbidden was shown in [8].

**Theorem 11.** (a) $\mathbf{WrFex}_* \subseteq \mathbf{TxtEx}$. (b) $\mathbf{NOFex}_* \subseteq \mathbf{TxtEx}$.

**Proof.** (a) Suppose $\mathbf{M}$ $\mathbf{WrFex}_*$-identifies $\mathcal{L}$.

Given a text $T$ for a language $L$ $(\in \mathcal{L})$, let us define an equivalence relation $E(i,j)$ as follows: If there exist $n_1, n_2, n_3, n_4$ such that $n_1 < n_2 < n_3 < n_4$, $\mathbf{M}(T[n_1]) = \mathbf{M}(T[n_3]) = i$ and $\mathbf{M}(T[n_2]) = \mathbf{M}(T[n_4]) = j$, then $E(i,j)$ (and $E(j,i)$) holds. Intuitively, $E(i,j)$ implies that $W_i$ is a grammar for $L$ iff $W_j$ is a grammar for $L$. This follows by definition of $\mathbf{WrFex}_*$ as either $W_i = W_j$, or both $W_i$ and $W_j$ are grammars for $L$. By taking reflexive and transitive closure of $E$, we get an equivalence relation.

It is easy to verify that all grammars which are output infinitely often by $\mathbf{M}$ on $T$ are equivalent (as they will pairwise satisfy $E(\cdot,\cdot)$).

Define $\mathbf{M}'$ as follows. $\mathbf{M}'(T[n])$ first builds an approximation to $E$ above based on $T[n]$, by setting $E(i,j)$ and $E(j,i)$ to true iff there exist $n_1, n_2, n_3, n_4$ such that $n_1 < n_2 < n_3 < n_4 \leq n$, $\mathbf{M}(T[n_1]) = \mathbf{M}(T[n_3]) = i$ and $\mathbf{M}(T[n_2]) = \mathbf{M}(T[n_4]) = j$. It then takes reflexive and transitive closure of $E$ so formed. $\mathbf{M}'(T[n])$, then outputs on $T[n]$ the union of languages enumerated by members of the equivalence class of $\mathbf{M}(T[n])$.

Now for all but finitely many $n$, as $\mathbf{M}$ outputs a grammar for $L$, $\mathbf{M}'(T[n])$ will be a grammar for $L$. Furthermore, there will be syntactic convergence as equivalence relation $E$ eventually stabilizes. Thus, $\mathbf{M}'$ $\mathbf{TxtEx}$-identifies $\mathcal{L}$.

(b) Similar to part (a), except that in this case, the meaning of equivalence relation is $E(i,j)$ implies $W_i \subseteq L \Leftrightarrow W_j \subseteq L$. This follows from the definition of $\mathbf{NOFex}$-identification as either $W_i = W_j$ or both are subsets of input language. ∎

As $\mathbf{TxtEx} = \mathbf{NOEx}$, we get the following result.

**Corollary 12.** $\mathbf{NOFex}_* = \mathbf{NOEx}$.

The following corollary extends Theorem 7 from the previous section.

**Corollary 13.** $\mathbf{WrFex}_* = \mathbf{DecEx}$.

From the above Corollaries we can conclude that, as was the case for **TxtEx**-learning, **Wr** is more restrictive than **NUSh** while **NO** is equivalent to **NUSh**. A closer look reveals a finer picture. We have shown that more learning power is lost, in the vacillatory case, by forbidding to return to abandoned wrong conjectures than by forbidding to return to correct conjectures. Also, some results from [8] seem to suggest that the necessity of returning to wrong conjectures is even *deeper* than the necessity of returning to correct conjectures, from the $\mathbf{TxtFex}_3$ level of the **TxtFex** hierarchy up, in the following sense. Recall the following result from [8].

**Theorem 14.** [8] *$\mathbf{TxtFex}_2 \subseteq \mathbf{NUShBc}$; $\mathbf{TxtFex}_3 \not\subseteq \mathbf{NUShBc}$.*

Thus, returning to correct conjectures is avoidable for the $\mathbf{TxtFex}_2$ level of the vacillatory hierarchy by shifting to the more liberal criterion of **TxtBc** identification, while there are classes learnable in the $\mathbf{TxtFex}_b$ sense for every $b > 2$ that cannot be learned by a **NUSh**-learner even in the **TxtBc** sense. In the next section we prove (Theorem 15) that there are **TxtEx**-learnable classes that *cannot* be **TxtBc**-learned by any **Wr**-learner. Thus, the necessity of returning to wrong abandoned conjectures is *not* avoidable by allowing infinitely many correct grammars in the limit, not even for the $\mathbf{TxtFex}_2$ level of the vacillatory hierarchy, while the necessity of returning to correct abandoned conjectures is so avoidable for this level of the vacillatory hierarchy.

## 6 Behaviourally Correct Learning

Our first result shows that, in the context of **TxtBc**-learnability, similarly to **TxtEx**-learnability, disallowing to return to wrong conjectures significantly limits the power of a learner: even **TxtEx**-learners can sometimes learn more than any **TxtBc**-learner if returning to wrong conjectures is not allowed. The reason is that the class $\mathcal{L}$ in $\mathbf{TxtEx} - \mathbf{DecBc}$ from [2] contains two distinct extensions of every finite set and thus the next theorem follows from Lemma 6.

**Theorem 15.** *$\mathbf{TxtEx} \not\subseteq \mathbf{WrBc}$.*

Now we compare non U-shaped learning (when a learner cannot abandon a correct conjecture) with learning by disallowing to return to wrong conjectures. From the previous Theorem and from the fact that $\mathbf{TxtEx} = \mathbf{NUShEx} \subseteq \mathbf{NUShBc}$, we have the following.

**Corollary 16.** *$\mathbf{NUShBc} \not\subseteq \mathbf{WrBc}$.*

We now show that, interestingly, the converse is true: **Wr** learners can sometimes do better than **NUSh** learners in the **TxtBc** setting. So **Wr** and **NUSh** are incomparable restrictions in the context of **TxtBc**-identification.

**Theorem 17.** *$\mathbf{WrBc} \not\subseteq \mathbf{NUShBc}$.*

We omit the proof of above theorem. Observe that, in contrast to the case of **TxtEx** and **TxtFex**-learning, Theorem 17 implies that **WrBc** does not coincide with **DecBc**. We have in fact the following corollary of Theorem 17.

**Corollary 18.** *$\mathbf{DecBc} \subset \mathbf{WrBc}$.*

Our next goal is to show that, by contrast to Theorem 15, any **TxtBc**-learner can be transformed into one that does not return to "overgeneralizing" conjectures. First, we need to establish a number of preliminary facts.

**Theorem 19.** (Based on [17]) *Suppose $\mathcal{L} \in$ **TxtBc**. Then there exists a machine $\mathbf{M}'$ such that $\mathbf{M}'$ **TxtBc**-identifies $\mathcal{L}$, and every text $T$ for $L \in \mathcal{L}$ starts with a **TxtBc**-locking sequence for $\mathbf{M}'$ on $L$.*

**Lemma 20.** *Suppose $\mathbf{M}$ is given. Then there exists an r.e. set $P(\sigma)$ such that*

- *A grammar for $P(\sigma)$ can be effectively obtained from $\sigma$;*
- *If $\sigma$ is a **TxtBc**-locking sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$, then $P(\sigma)$ contains only grammars for $W_{\mathbf{M}(\sigma)}$;*
- *If $\sigma$ is not a **TxtBc**-locking sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$, then $P(\sigma)$ is either empty or contains grammars for at least two distinct languages.*

**Proof.** Define $P(\sigma)$ as follows. If $\text{content}(\sigma) \not\subseteq W_{\mathbf{M}(\sigma)}$, then let $P(\sigma) = \emptyset$, else let $P(\sigma) = \{\mathbf{M}(\tau) \mid \sigma \subseteq \tau, \text{content}(\tau) \subseteq W_{\mathbf{M}(\sigma)}\}$. Now if $\sigma$ is a **TxtBc**-locking sequence for $\mathbf{M}$ on $L$, then $P(\sigma)$ consists only of grammars for $L$. On the other hand if $\sigma$ is not a **TxtBc**-locking sequence for $\mathbf{M}$ on $L$, then either $P(\sigma)$ is empty or it contains grammars for at least two distinct languages. ∎

**Lemma 21.** *Given $\mathbf{M}$, there exists a recursive function $g$ such that:*
(a) *If $\sigma$ is a **TxtBc**-locking sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$, then $W_{g(\sigma)} = W_{\mathbf{M}(\sigma)}$.*
(b) *If $\sigma$ is not a **TxtBc**-locking sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$, then $W_{g(\sigma)}$ is finite.*

**Proof.** For a finite set $X$ and number $s$, let

- $CommonTime(X, s) = \max(\{t \le s \mid (\forall p, p' \in X) W_{p,t} \subseteq W_{p',s}\})$;
- $CommonElem(X, s) = \bigcap_{p \in X} W_{p, CommonTime(X,s)}$.

Let $f$ be a recursive function with $W_{f(X)} = \bigcup_{s \in \mathcal{N}} CommonElem(X, s)$. Here we assume that $W_{f(\emptyset)} = \emptyset$. Intuitively, $CommonTime(X, s)$ finds the largest $s$ such that enumerations upto time $CommonTime(X, s)$ by grammars in $X$ are included in all languages enumerated by grammars in $X$. $CommonElem(X, s)$ then gives intersection of elements enumerated by grammars in $X$ upto time $CommonTime(X, s)$. Note that

(I) $\lim_{s \to \infty} CommonTime(X, s)$ is infinite iff all grammars in $X$ are for the same language;
(II) If $X \subseteq X'$, then $CommonTime(X, s) \ge CommonTime(X', s)$;
(III) If $W_p \ne W_{p'}$ then for all $s$, $CommonTime(\{p, p'\}, s)$ is bounded by the least $t$ such that $W_{p,t} \cup W_{p',t} \not\subseteq W_p \cap W_{p'}$.

Let $Y$ be the set of all $y$ such that there is an $s \ge y$, such that $y \in W_{f(X_s)}$. Note that (II) and (III) above imply that if $X_0 \subseteq X_1 \subseteq X_2 \subseteq \ldots$, $\{p, p'\} \subseteq \bigcup_{i \in \mathcal{N}} X_i$ and $W_p \ne W_{p'}$, then $Y$ is finite. On the other hand, if all $p, p' \in \bigcup_{i \in \mathcal{N}} X_i$ are grammars for the same language, then $Y = W_p$ for any $p \in \bigcup_{i \in \mathcal{N}} X_i$.

Let $P$ be as in Lemma 20 and let $P_s(\sigma)$ denote $P(\sigma)$ enumerated in $s$ steps.

Now let $g(\sigma)$ be such that $W_{g(\sigma)} = \bigcup_{s\in\mathcal{N}}[\{y \leq s \ \wedge \ y \in W_{f(P_s(\sigma))}\}]$. It is now easy to verify that Lemma holds. ∎

Now we can prove one of our main results: any **TxtBc**-learner can be replaced by the one not returning to "overgeneralizing" conjectures.

**Theorem 22.** **TxtBc** $\subseteq$ **NOBc**.

**Proof.** Suppose **M** **TxtBc**-identifies $\mathcal{L}$. Without loss of generality (Theorem 19) assume that for any text $T$ for $L \in \mathcal{L}$, there exists a $\sigma \subseteq T$, such that $\sigma$ is a **TxtBc**-locking sequence for **M** on $L$. Intuitively, the proof employs two tricks. First trick (as given by $g$ in Lemma 21) is to make sure that the infinite languages output by the learner are only on $\sigma$'s which are **TxtBc**-locking sequences for the language output. This automatically ensures no semantic mind changes occur between different grammars output for the same infinite language by the learner. The second trick makes sure that finite languages output by the learner, which go beyond what is seen in the input at the time of conjecture, are for pairwise different languages. We now proceed formally.

Let $g$ be as in Lemma 21. Let $q_0, q_1, \ldots$ be an increasing sequence of primes and $\mathbf{M}''(\sigma) = h(\sigma)$ where $W_{h(\sigma)}$ is defined as follows.

Begin $W_{h(\sigma)}$
  Enumerate content$(\sigma)$
  Loop
  Search for $s$ such that $W_{h(\sigma)}$ enumerated upto now is a proper subset of $W_{g(\sigma),s}$, and card$(W_{g(\sigma),s})$ is $(q_{\text{ind}(\sigma)})^k$ for some $k$.
  If and when such $s$ is found, enumerate $W_{g(\sigma),s}$.
  Forever
End

Thus, $W_{h(\sigma)}$ is $W_{g(\sigma)}$ if $W_{g(\sigma)}$ is infinite. Furthermore, if $W_{h(\sigma)}$ is finite, then it is either content$(\sigma)$ or has cardinality a power of $q_{\text{ind}(\sigma)}$.

It follows that if $W_{h(\sigma)} = W_{h(\tau)}$, for $\sigma \subset \tau$, then either $W_{h(\sigma)}$ is infinite and $\sigma$ is a **TxtBc**-locking sequence for **M** on $W_{g(\tau)} = W_{g(\sigma)} = W_{h(\sigma)}$, and thus, there is no semantic mind change by $\mathbf{M}''$ in between $\sigma$ and $\tau$, or $W_{h(\sigma)}$ is finite, and thus, it must be the case that $W_{h(\sigma)} = W_{h(\tau)} =$ content$(\tau)$ (otherwise, $q_{\text{ind}(\sigma)} \neq q_{\text{ind}(\tau)}$ would imply that $W_{h(\sigma)} \neq W_{h(\tau)}$).

It follows from above cases that $\mathbf{M}''$ does not return to "overgeneralizing" hypothesis. To see **TxtBc**-identification of $L \in \mathcal{L}$, let $T$ be a text for $L$. Let $T[n]$ be a **TxtBc**-locking sequence for **M** on $L$ (such an $n$ exists by Theorem 19). Thus, $g(T[n])$ is a grammar for $L$. If $L$ is finite, then without loss of generality we also assume that $n$ is large enough such that $L \subseteq$ content$(T[n])$. Now consider any $m \geq n$. It is easy to verify that if $L$ is infinite then $W_{h(T[m])} = W_{g(T[m])} = L$. On the other hand, if $L$ is finite, then again $W_{h(T[m])}$ does not go beyond first step, and thus equals $L$. ∎

## 7 Consistency

Consistency is a natural and important requirement for **TxtEx** and **TxtBc** types of learning. While for the latter, consistency requirement can be easily achieved, it is known to be restrictive for **TxtEx**-learnability [3, 21]. In this section, we establish a new interesting boundary on consistent **TxtEx**-learnability – in Theorem 25 we show that consistent **TxtEx**-learners can be made decisive (still being consistent) – contrast this result with Theorem 9(a).

**Definition 23.** [3, 21] **M** is said to be *consistent* on $T$ iff, for all $n$, $\mathbf{M}(T[n])\downarrow$ and content$(T[n]) \subseteq W_{\mathbf{M}(T[n])}$.

**M** is said to be *consistent* on $L$ iff, **M** is consistent on each text for $L$.

**Definition 24.** (a) [3, 21] **M** **ConsTxtEx**-*identifies* $L$ iff **M** is consistent on $L$, and **M** **TxtEx**-identifies $L$.
(b) [3] **M** **ConsTxtEx**-*identifies* $\mathcal{L}$ iff **M** **ConsTxtEx**-identifies each $L \in \mathcal{L}$.
**ConsTxtEx** $= \{\mathcal{L} \mid (\exists \mathbf{M})[\mathbf{M}$ **ConsTxtEx**-identifies $\mathcal{L}]\}$.

Note that for **M** to **ConsTxtEx**-identify a text $T$, it must be defined on each initial segment of $T$.[8] One can similarly define combination of consistency with decisive (called **ConsDecEx**) and other related criteria such as **ConsNUShEx**, **ConsNOEx**, **ConsWrEx**, etc. We omit proof of theorems in this section.

**Theorem 25.** **ConsTxtEx** $\subseteq$ **ConsDecEx**.

**Theorem 26.** **NUShBc** $=$ **ConsNUShBc**.

Next we show that decisive learning is stronger than consistent learning.

**Theorem 27.** **DecEx** $-$ **ConsTxtEx** $\neq \emptyset$

The proof of Theorem 22 also shows that **TxtBc** $\subseteq$ **ConsNOBc**. Thus, we have

**Theorem 28.** **TxtBc** $\subseteq$ **ConsNOBc**.

The proof of Theorem 11 also works for the case when we are considering consistent identification, so we have

**Theorem 29.** $\mathbf{ConsWrFex}_*$ $\subseteq$ **ConsTxtEx**; $\mathbf{ConsNOFex}_*$ $\subseteq$ **ConsTxtEx**.

**Corollary 30.** $\mathbf{ConsWrFex}_*$ $\subseteq$ **ConsDecEx**; $\mathbf{ConsNOFex}_*$ $\subseteq$ **ConsDecEx**.

The proof of Theorem 17 shows the following as well.

**Theorem 31.** **ConsWrBc** $\not\subseteq$ **NUShBc**.

The following are open: (a) **ConsWrBc** $=$ **WrBc**? (b) **ConsDecBc** $=$ **DecBc**?

[8] There are two other versions of consistency considered in the literature, namely $\mathcal{R}$**Cons** [16] where the learner must be total but might be inconsistent on data not belonging to the class to be learned and $\mathcal{T}$**Cons** [27] where the learner must be total and consistent on every input, whether it belongs to some language to be learnt or not. Our results also hold for $\mathcal{T}$**Cons**, however some of our results do not hold for $\mathcal{R}$**Cons**.

## 8 Conclusions

We summarize our results on the impact of the **Wr** and **NO** variants of non U-shaped behaviour and how they compare to previous results about the original notion **NUSh** from [2] and [8].

Returning to abandoned wrong conjectures turned out to be necessary for full learning power in all three of the models **TxtEx**, **TxtFex** and **TxtBc**, while returning to abandoned wrong "overgeneralizing" conjectures is necessary only for the vacillatory case and avoidable otherwise. This can be compared to results in [2] and [8] showing that returning to abandoned correct conjectures is avoidable in the **TxtEx** case while being necessary for vacillatory and behaviourally correct identification.

Also, we can conclude that forbidding to return to abandoned wrong conjectures is more restrictive than forbidding to return to correct conjectures in the **TxtEx** and in the **TxtFex** models, while the two restrictions are incomparable in the **TxtBc** case. On the other hand, forbidding to return to wrong "overgeneralizing" conjectures is equivalent to forbidding to return to correct conjectures for **TxtEx** and **TxtFex** identification.

Also, while, for the level $\mathbf{TxtFex}_2$ of the vacillatory hierarchy, the necessity of returning to correct conjectures is avoidable by shifting to the more liberal criterion of **TxtBc**-identification, the necessity of returning to wrong conjectures is *not* avoidable in this way: there are $\mathbf{TxtFex}_2$-learnable classes that cannot be **TxtBc**-learned by any **Wr** learner. This and the above observations seem to suggest that the freedom of returning to wrong abandoned conjectures is even more central for full learning power, than the freedom of returning to correct conjectures. We defer a deeper analysis of the possible significance of these results from the perspective of cognitive science to a more appropriate place.

We would like to thank Rolf Wiehagen for useful discussions, and referees of COLT'2005 for several helpful comments.

## References

1. D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
2. G. Baliga, J. Case, W. Merkle, F. Stephan, and R. Wiehagen. When unlearning helps. Manuscript, http://www.cis.udel.edu/˜case/papers/decisive.ps, 2005. Preliminary version of the paper appeared in ICALP, 2000.
3. J. Bārzdiņš. Inductive inference of automata, functions and programs. In *Int. Math. Congress, Vancouver*, pages 771–776, 1974.
4. L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.
5. M. Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.
6. M. Bowerman. Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Strauss and R. Stavy, editors, *U-Shaped Behavioral Growth*. Developmental Psychology Series. Academic Press, New York, 1982.

7. S. Carey. An analysis of a learning paradigm. In S. Strauss and R. Stavy, editors, *U-Shaped Behavioral Growth*. Developmental Psychology Series. Academic Press, New York, 1982.
8. L. Carlucci, J. Case, S. Jain, and F. Stephan. U-shaped learning may be necessary. Technical Report TRA11/04, School of Computing, National University of Singapore, Nov 2004.
9. J. Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28(6):1941–1969, 1999.
10. J. Case and C. 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. J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.
12. M. Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.
13. M. Fulk, S. Jain, and D. Osherson. Open problems in systems that learn. *Journal of Computer and System Sciences*, 49(3):589–604, 1994.
14. E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
15. J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.
16. K. Jantke and H. Beick. Combining postulates of naturalness in inductive inference. *Journal of Information Processing and Cybernetics (EIK)*, 17:465–484, 1981.
17. S. Kurtz and J. Royer. Prudence in language learning. In D. Haussler and L. Pitt, editors, *Proceedings of the Workshop on Computational Learning Theory*, pages 143–156. Morgan Kaufmann, 1988.
18. S. Lange and R. Wiehagen. Polynomial time inference of arbitrary pattern languages. *New Generation Computing*, 8:361–370, 1991.
19. M. Machtey and P. Young. *An Introduction to the General Theory of Algorithms*. North Holland, New York, 1978.
20. G. Marcus, S. Pinker, M. Ullman, M. Hollander, T. Rosen, and F. Xu. *Overregularization in Language Acquisition*. Monographs of the Society for Research in Child Development, vol. 57, no. 4. University of Chicago Press, 1992. Includes commentary by Harold Clahsen.
21. D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, 1986.
22. K. Plunkett and V. Marchman. U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, 38(1):43–102, 1991.
23. H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, 1967. Reprinted by MIT Press in 1987.
24. S. Strauss and R. Stavy. *U-Shaped Behavioral Growth*. Developmental Psychology Series. Academic Press, New York, 1982.
25. S. Strauss, R. Stavy, and N. Orpaz. The child's development of the concept of temperature. Manuscript, Tel-Aviv University, 1977.
26. N.A. Taatgen and J.R. Anderson. Why do children learn to say broke? a model of learning the past tense without feedback. *Cognition*, 86(2):123–155, 2002.
27. R. Wiehagen and W. Liepe. Charakteristische Eigenschaften von erkennbaren Klassen rekursiver Funktionen. *Journal of Information Processing and Cybernetics (EIK)*, 12:421–438, 1976.