Variations on U-shaped learning

Lorenzo Carlucci [a,1] Sanjay Jain [b,2] Efim Kinber [c]
Frank Stephan [d,3]

[a] *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, 00153, Siena, Italy*

[b] *School of Computing, National University of Singapore, Singapore 117543, Republic of Singapore*

[c] *Department of Computer Science, Sacred Heart University, Fairfield, CT 06432-1000, U.S.A.*

[d] *School of Computing and Department of Mathematics, National University of Singapore, 3 Science Drive 2, Singapore 117543, Republic of Singapore*

Abstract The paper deals with the following problem: is returning to wrong conjectures necessary to achieve full power of algorithmic learning? Returning to wrong conjectures complements the paradigm of *U-shaped learning* [3,7,9,23,27] 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 both cases, we show that, surprisingly, returning to wrong conjectures is necessary to achieve full learning power. In contrast inverted U's (i.e. return to old wrong conjecture with a correct conjecture in-between) and return to old "overinclusive" conjectures containing non-elements of the target language are dispensable strategies. 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 returning to old wrong conjectures, using an inverted-U-shaped strategy, returning to old overinclusive and returning to old overgeneralizing conjectures are all necessary for full learning power. We also show that, surprisingly, learners consistent with the input seen so far can be made *decisive* [3,24] – they do not have to return to any old conjectures – wrong or right.

Email addresses: `carlucci5@unisi.it` (Lorenzo Carlucci), `sanjay@comp.nus.edu.sg` (Sanjay Jain), `kinbere@sacredheart.edu` (Efim Kinber), `fstephan@comp.nus.edu.sg` (Frank Stephan).
[1] Supported in part by NSF grant number NSF CCR-0208616.
[2] Supported in part by NUS grant number R252–000–127–112.
[3] Supported in part by NUS grant number R252–000–212–112.

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 [7,23,27], understanding of temperature [27,28] and face recognition [8]. The ability of models of human learning to accommodate U-shaped learning progressively became one of the important criteria of their adequacy; see [23,25] and the recent [29]. Renewed interest in U-shaped learning is also witnessed by the fact that the Journal of Cognition and Development dedicated its first issue in the year 2004 to this phenomenon.

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 [3] 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 [16] 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 [9], 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 [10], forms a hierarchy of more and more powerful learning criteria between **TxtEx** and **TxtBc** identification. It was shown in [9] 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 "non monotonic" learning behaviour, when a learner chooses a hypothesis, then abandons it, then returns to it once again [4]. 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. Examples of this kind of apparently inefficient behaviour have been documented by developmental psy-

[4] These two meanings of U-shaped behaviour are explicitly distinguished at the beginning of [27], the main reference for the study of U-shaped behaviour.

chologists in the context of infants' face recognition. For example, it has been shown that children exhibit an "inverted-U-shaped" (wrong-correct-wrong) learning curve for recognition of inverted faces and an "N-shaped" (wrong-correct-wrong-correct) learning curve for recognition of upright faces [13,14]. In this paper we study the following general 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 is "overinclusive" in the sense of containing elements not belonging to the target concept.

We also study the following less restrictive natural variant of model (a).

(a') a model in which a learner cannot return to a previously abandoned wrong conjecture if a correct conjecture has been made in-between.

Model (a') is inspired by the concrete cases of inverted-U-shaped and N-shaped behaviour documented in the psychological literature [13,14]. It also represents the exact inverse of the original non U-shaped model studied in [3,9].

Finally we study the following natural variant of (b).

(b') a model in which a learner cannot return to a previously abandoned conjecture that "overgeneralizes" in the sense of being a proper superset of the target language.

The latter model is motivated by the fact that overgeneralization is one of the major concerns in the study of learning behaviour [23].

We compare the new 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, inverted U-shapes are redundant and it is not necessary to return to old overinclusive conjectures or to old overgeneralizing conjectures. On the other hand, for vacillatory identification, returning to wrong conjectures, inverted U-shapes and returning to overinclusive conjectures are all necessary in a very strong sense: forbidding this kind of U-shapes collapses the whole **TxtFex**-hierarchy to simple **TxtEx**-learning. In contrast, if return to previously conjectured proper supersets is forbidden, no such collapse occurs, but, still, returning to such overgeneralizing conjectures is necessary for full learning power at each level of the vacillatory hierarchy. We compare more thoroughly these conclusions with results from [9] 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 21); decisive learners are known [3] to be generally weaker than general **TxtEx**-learners. On the other hand, any **TxtEx**-learner can be replaced by a learner not returning to overinclusive conjectures (Theorem 22). From this result we also obtain that any **TxtEx**-learner can be replaced by a learner not returning to overgeneralizing conjectures and by a learner not showing an inverted-U behaviour as well (Corollaries 23 and 25 respectively).

In Section 5 we consider our four 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. Vacillatory learning constitutes a hierarchy of more and more powerful learning criteria strictly between **TxtEx** and **TxtBc**. The more vacillation is allowed, the more learning power is possible, as shown in [10]. 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 overinclusive conjectures and for vacillatory learners that do not show an inverted-U-shaped behavior, they turn out to be doing no better than regular **TxtEx**-learners of this type. It was shown in [9] that the same collapse of the vacillatory hierarchy occurs when return to correct conjectures is forbidden. Thus, forbidding return to wrong conjectures, forbidding return to overinclusive conjectures and forbidding inverted-U-shapes each nullifies the extra power of finite vacillation with respect to convergence to a single correct conjecture. In contrast, we show that forbidding return to overgeneralizing conjectures restricts the power of vacillatory learners in a less severe sense: for each $n > 0$ there are classes of languages that are learnable with vacillation between at most $n+1$ correct conjectures in the limit by a learner not returning to old overgeneralizing conjectures but not learnable by any learner who is allowed to vacillate between at most n correct conjectures (Theorem 30). Also, we show that there are classes of languages that are learnable with vacillation between at most $n+1$ correct conjectures but such that any such learner must return to old overgeneralizing conjectures (Theorem 31). Hence forbidding return to old overgeneralizing conjectures restricts the learning power of each level of the vacillatory hierarchy, but does annihilate the extra power of vacillation over **TxtEx**-identification.

In Section 6 we explore our four 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 [3] 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. In contrast, we show that inverted U's are unnecessary in the context of **TxtBc**-learning (Theorem 44). The main result of this section is that, as in case of **TxtEx**-learnability, returning to old overinclusive conjectures can be circumvented: every **TxtBc**-learner can be replaced by one not returning to overinclusive conjectures (Theorem 48). As a corollary, we obtain that returning to proper supersets of the target language is unnecessary for full learning power in the **TxtBc** context (Corollary 49).

In Section 7 we discover a relationship between the strongest type of non U-shaped learners, that is decisive learners, and *consistent* learners [4,24], 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 [4,24]; moreover, sacrificing consistency, one can learn *pattern languages* faster than any consistent learner [21]. We show that consistent **TxtEx**-learners can be made consistent and decisive (Theorem 53). 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).

Some open questions are formulated in the final section.

2 Notation and Preliminaries

Any unexplained recursion theoretic notation is from [26]. 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 $\langle \cdot, \cdot \rangle$ stand for Cantor's computable, bijective mapping $\langle x, y \rangle = \frac{1}{2}(x+y)(x + y + 1) + x$ from $\mathcal{N} \times \mathcal{N}$ onto $\mathcal{N}$ [26]. Note that $\langle \cdot, \cdot \rangle$ is monotonically increasing in both of its arguments. We define $\pi_1(\langle x, y \rangle) = x$ and $\pi_2(\langle x, y \rangle) = y$.

By φ we denote a fixed *acceptable* programming system for the partial-recursive functions mapping $\mathcal{N}$ to $\mathcal{N}$ [22,26]. By φ_i we denote the partial-recursive function computed by the program with number i in the φ-system. The symbol $\mathcal{R}$ denotes the set of all recursive functions, that is total computable functions. By Φ we denote an arbitrary fixed Blum complexity measure [6,18] for the φ-system. A partial recursive function $\Phi(\cdot,\cdot)$ is said to be a Blum complexity measure for φ, iff the following two conditions are satisfied:

(a) for all i and x, $\Phi(i,x)\downarrow$ iff $\varphi_i(x)\downarrow$.

(b) the predicate: $P(i,x,t) \equiv \Phi(i,x) \leq t$ is decidable.

By convention we use Φ_i to denote the partial recursive function $x \rightarrow \Phi(i,x)$. Intuitively, $\Phi_i(x)$ may be thought as the number of steps it takes to compute $\varphi_i(x)$.

By W_i we denote domain(φ_i). That is, W_i is then the recursively enumerable (r.e.) subset of $\mathcal{N}$ accepted by the φ-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. The next definition introduces the concept of a *sequence* of data.

Definition 1 (a) A *sequence* σ is a mapping from an initial segment of $\mathcal{N}$ into $(\mathcal{N} \cup \{\#\})$. The empty sequence is denoted by λ.

(b) The *content* of a sequence σ, denoted content(σ), is the set of natural numbers in the range of σ.

(c) The *length* of σ, denoted by $|\sigma|$, is the number of elements in σ. So, $|\lambda| = 0$.

(d) For $n \leq |\sigma|$, the initial sequence of σ of length n is denoted by $\sigma[n]$. So, $\sigma[0]$ is λ.

Intuitively, the pause-symbol # represents a pause in the presentation of data. We let σ, τ and γ range over finite sequences. We denote the sequence formed by the concatenation of τ at the end of σ by $\sigma\tau$. Sometimes we abuse the notation and use σx to denote the concatenation of sequence σ 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(content(δ_i)) $\leq i$. We let ind(σ) denote i such that $\delta_i = \sigma$.

We let SEG(L) denote the set $\{\sigma \mid \text{content}(\sigma) \subseteq L\}$.

Definition 2 [17] (a) A *text* T for a language L 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.

(b) The *content* of a text T, denoted by content(T), is the set of natural numbers in the range of T; that is, the language which T is a text for.

(c) $T[n]$ denotes the finite initial sequence of T with length n.

Definition 3 [17] A *learning machine* (or just *learner*) 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 criteria involving 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 4 [11,17] (a) $\mathbf{M}$ $\mathbf{TxtEx}$-*identifies a text* T just in case $(\exists i \mid W_i = \text{content}(T))$ $(\forall^\infty n)[\mathbf{M}(T[n]) = i]$.

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

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

(d) $\mathbf{TxtEx} = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtEx}(\mathbf{M})]\}$.

Definition 5 [11] (a) $\mathbf{M}$ $\mathbf{TxtBc}$-*identifies a text* T just in case $(\forall^\infty n)[W_{\mathbf{M}(T[n])} = \text{content}(T)]$.

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

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

(d) $\mathbf{TxtBc} = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtBc}(\mathbf{M})]\}$.

Definition 6 [10] Suppose $a \in \mathcal{N} \cup \{*\}$.

(a) $\mathbf{M}$ $\mathbf{TxtFex}_a$-*identifies a text* T just in case there exists a set D, $\text{card}(D) \leq a$ and $(\forall i \in D)[W_i = \text{content}(T)]$, such that $(\forall^\infty n)[W_{\mathbf{M}(T[n])} \in D]$.

(b) $\mathbf{M}$ $\mathbf{TxtFex}_a$-*identifies an r.e. language* L (written: $L \in \mathbf{TxtFex}_a(\mathbf{M})$) just in case $\mathbf{M}$ $\mathbf{TxtFex}_a$-identifies each text for L.

(c) $\mathbf{M}$ $\mathbf{TxtFex}_a$-*identifies a class* $\mathcal{L} \subseteq \mathcal{E}$ (written: $\mathcal{L} \subseteq \mathbf{TxtFex}_a(\mathbf{M})$) just in case $\mathbf{M}$ $\mathbf{TxtFex}_a$-identifies each language from $\mathcal{L}$.

(d) $\mathbf{TxtFex}_a = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtFex}_a(\mathbf{M})]\}$.

It is known that $\mathbf{TxtEx} \subset \mathbf{TxtFex}_2 \subset \mathbf{TxtFex}_3 \subset \ldots \subset \mathbf{TxtFex}_* \subset \mathbf{TxtBc}$ (see [10–12]).

Definition 7 (a) [15] σ is said to be a *stabilizing sequence* for $\mathbf{M}$ on L iff $\text{content}(\sigma) \subseteq L$, and for all $\tau \supseteq \sigma$ such that $\text{content}(\tau) \subseteq L$, $\mathbf{M}(\sigma) = \mathbf{M}(\tau)$.

(b) [5] σ is said to be a **TxtEx**-*locking sequence* for $\mathbf{M}$ on L iff σ is a stabilizing sequence for $\mathbf{M}$ on L, and $W_{\mathbf{M}(\sigma)} = L$.

(c) (Based on [5]) σ is said to be a **TxtBc**-*locking sequence* for $\mathbf{M}$ on L iff content(σ) $\subseteq L$, and for all $\tau \supseteq \sigma$ such that content(τ) $\subseteq L$, $W_{\mathbf{M}(\sigma)} = L$.

Lemma 8 [5] *If* $\mathbf{M}$ **TxtEx**-*identifies* L*, then there exists a* **TxtEx**-*locking sequence for* $\mathbf{M}$ *on* L*.*

Similar result as above can be shown for **TxtBc** (and $\mathbf{TxtFex}_a$) criteria of learning.

$\mathcal{L}$ is said to be an *indexed family* of languages iff there exists an indexing $L_0, L_1, \ldots$ of languages in $\mathcal{L}$ such that the question $x \in L_i$ is uniformly decidable, that is, there exists a recursive function f such that $f(i,x) = 1$ if $x \in L_i$ and $f(i,x) = 0$ otherwise.

We let INIT $= \{L \mid (\exists i)[L = \{x \mid x \leq i\}]\}$. Let $INIT_k$ denote the set $\{x \mid x \leq k\}$.

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

Firstly, we define the strongest type of non U-shaped behaviour – when a learner is not allowed to return to *any* old conjectures.

Definition 9 [24] (a) We say that $\mathbf{M}$ 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) We say that $\mathbf{M}$ is decisive on L if $\mathbf{M}$ is decisive on each text for L.

(c) We say that $\mathbf{M}$ is decisive on $\mathcal{L}$ if $\mathbf{M}$ is decisive on each $L \in \mathcal{L}$.

Now we define non U-shaped learning.

Definition 10 [2] (a) We say that $\mathbf{M}$ 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 $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.

(b) We say that $\mathbf{M}$ is non U-shaped on L if $\mathbf{M}$ is non U-shaped on each text for L.

(c) We say that $\mathbf{M}$ is non U-shaped on $\mathcal{L}$ if $\mathbf{M}$ is non U-shaped on each $L \in \mathcal{L}$.

Now we define our four models of non U-shaped learning when a learner is not allowed to return to previously used wrong conjectures ("Wr" in the next definition stands for "wrong").

Definition 11 (a) We say that $\mathbf{M}$ is *decisive on wrong conjectures* (abbreviated *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 $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.

(b) We say that $\mathbf{M}$ is Wr-decisive on L if $\mathbf{M}$ is Wr-decisive on each text for L.

(c) We say that $\mathbf{M}$ is Wr-decisive on $\mathcal{L}$ if $\mathbf{M}$ is Wr-decisive on each $L \in \mathcal{L}$.

Now we define our model of non inverted-U-shaped learning.

Definition 12 (a) We say that $\mathbf{M}$ is *non inverted-U-shaped* on text T, if there do not exist any m, n, t such that $m < n < t$, $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[t])} \neq W_{\mathbf{M}(T[n])} = \text{content}(T)$.

(b) We say that $\mathbf{M}$ is non inverted-U-shaped on L if $\mathbf{M}$ is non inverted-U-shaped on each text for L.

(c) We say that $\mathbf{M}$ is non inverted-U-shaped on $\mathcal{L}$ if $\mathbf{M}$ is non inverted-U-shaped on each $L \in \mathcal{L}$.

We now define our model of learning forbidding return to conjectures containing elements outside the target language ("OI" in "OI-decisive" below stands for "overinclusive").

Definition 13 (a) We say that $\mathbf{M}$ is *decisive on overinclusive conjectures* (abbreviated *OI-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 $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.

(b) We say that $\mathbf{M}$ is OI-decisive on L if $\mathbf{M}$ is OI-decisive on each text for L.

(c) We say that $\mathbf{M}$ is OI-decisive on $\mathcal{L}$ if $\mathbf{M}$ is OI-decisive on each $L \in \mathcal{L}$.

We now introduce our model in which return to proper supersets is forbidden ("OG" in "OG-decisive" below stands for "overgeneralizing").

Definition 14 (a) We say that $\mathbf{M}$ is *decisive on overgeneralizing conjectures* (abbreviated *OG-decisive*) on text T, if there do not exist m, n, t such that $m < n < t$, $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[t])} \supset \text{content}(T)$ and $W_{\mathbf{M}(T[m])} \neq W_{\mathbf{M}(T[n])}$.

(b) We say that $\mathbf{M}$ is OG-decisive on L if $\mathbf{M}$ is OG-decisive on each text for L.

(c) We say that $\mathbf{M}$ is OG-decisive on $\mathcal{L}$ if $\mathbf{M}$ is OG-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, non inverted-U-shaped, OI-decisive and OG-decisive criteria, where only the class version seems sensible.

Definition 15 (a) [24] We say that $\mathbf{M}$ $\mathbf{DecEx}$-identifies L (written: $L \in \mathbf{DecEx}(\mathbf{M})$), iff $\mathbf{M}$ $\mathbf{TxtEx}$-identifies L, and $\mathbf{M}$ is decisive on L.

We say that $\mathbf{M}$ $\mathbf{DecEx}$-identifies $\mathcal{L}$, iff $\mathbf{M}$ $\mathbf{DecEx}$-identifies each $L \in \mathcal{L}$.

$\mathbf{DecEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{DecEx}(\mathbf{M})]\}$.

(b) [2] We say that $\mathbf{M}$ $\mathbf{NUShEx}$-identifies L (written: $L \in \mathbf{NUShEx}(\mathbf{M})$), iff $\mathbf{M}$ $\mathbf{TxtEx}$-identifies L, and $\mathbf{M}$ is non U-shaped on L.

We say that $\mathbf{M}$ $\mathbf{NUShEx}$-identifies $\mathcal{L}$, iff $\mathbf{M}$ $\mathbf{NUShEx}$-identifies each $L \in \mathcal{L}$.

$\mathbf{NUShEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NUShEx}(\mathbf{M})]\}$.

(c) We say that $\mathbf{M}$ $\mathbf{WrDEx}$-identifies L (written: $L \in \mathbf{WrDEx}(\mathbf{M})$), iff $\mathbf{M}$ $\mathbf{TxtEx}$-identifies L, and $\mathbf{M}$ is Wr-decisive on L.

We say that $\mathbf{M}$ **WrDEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **WrDEx**-identifies each $L \in \mathcal{L}$.

$\mathbf{WrDEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{WrDEx}(\mathbf{M})]\}$.

(d) We say that $\mathbf{M}$ **NInvUEx**-identifies L (written: $L \in \mathbf{NInvUEx}(\mathbf{M})$), iff $\mathbf{M}$ **TxtEx**-identifies L, and $\mathbf{M}$ is non inverted-U-shaped on L.

We say that $\mathbf{M}$ **NInvUEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **NInvUEx**-identifies each $L \in \mathcal{L}$.

$\mathbf{NInvUEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NInvUEx}(\mathbf{M})]\}$.

(e) We say that $\mathbf{M}$ **OIDEx**-identifies L (written: $L \in \mathbf{OIDEx}(\mathbf{M})$), iff $\mathbf{M}$ **TxtEx**-identifies L, and $\mathbf{M}$ is OI-decisive on L.

We say that $\mathbf{M}$ **OIDEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **OIDEx**-identifies each $L \in \mathcal{L}$.

$\mathbf{OIDEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{OIDEx}(\mathbf{M})]\}$.

(f) We say that $\mathbf{M}$ **OGDEx**-identifies L (written: $L \in \mathbf{OGDEx}(\mathbf{M})$), iff $\mathbf{M}$ **TxtEx**-identifies L, and $\mathbf{M}$ is OG-decisive on L.

We say that $\mathbf{M}$ **OGDEx**-identifies $\mathcal{L}$, iff $\mathbf{M}$ **OGDEx**-identifies each $L \in \mathcal{L}$.

$\mathbf{OGDEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{OGDEx}(\mathbf{M})]\}$.

One can similarly define **DecJ**, **WrDJ**, **NInvUJ**, **OIDJ**, **OGDJ** and **NUShJ**, for $\mathbf{J} \in \{\mathbf{Fex}_a, \mathbf{Bc}\}$.

The following result is easy to verify.

Proposition 16 *Suppose* $a \in \mathcal{N} \cup \{*\}$ *and* $\mathbf{J} \in \{\mathbf{Ex}, \mathbf{Fex}_a, \mathbf{Bc}\}$.

(a) $\mathbf{DecJ} \subseteq \mathbf{WrDJ} \subseteq \mathbf{OIDJ} \subseteq \mathbf{OGDJ} \subseteq \mathbf{J}$.

(b) $\mathbf{DecJ} \subseteq \mathbf{WrDJ} \subseteq \mathbf{NInvUJ} \subseteq \mathbf{J}$.

(c) $\mathbf{DecJ} \subseteq \mathbf{NUShJ} \subseteq \mathbf{NInvUJ} \subseteq \mathbf{J}$.

For our proofs, we will be using the following results from [3].

Lemma 17 [3] *Suppose* $\mathbf{M}$ **TxtEx**-*identifies* $\mathcal{L}$, *and* g *is a recursive function such that*

(I) $W_{g(i)} \neq W_{g(j)}$, *for* $i \neq j$,

(II) *for all finite sets* S, *there exist infinitely many* i *such that* $S \subseteq W_{g(i)}$,

(III) $W_{\mathbf{M}(\sigma)} \notin \{W_{g(i)} \mid i \in \mathcal{N}\}$, *for all* σ.

Then, $\mathcal{L} \in \mathbf{DecEx}$.

Proposition 18 [3] *Suppose* $\mathcal{L} \in \mathbf{TxtEx}$ *and* $\mathcal{N} \in \mathcal{L}$. *Then,* $\mathcal{L} \in \mathbf{DecEx}$.

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.

Lemma 19 *Suppose there exists a finite set* A *such that* $\mathcal{L}$ *does not contain any extension of* A. *Then* $\mathcal{L} \in \mathbf{TxtEx} \Rightarrow \mathcal{L} \in \mathbf{DecEx}$.

Proof. Suppose $\mathcal{L} \subseteq \mathbf{TxtEx}(\mathbf{M})$. Without loss of generality suppose $\mathbf{M}$ never outputs an extension of A: This can be achieved by converting any grammar i to $f(i)$ where

$$W_{f(i)} = \bigcup_s X_s \text{ and } X_s = \begin{cases} W_{i,s}, & \text{if } A \not\subseteq W_{i,s}; \\ \emptyset, & \text{otherwise.} \end{cases}$$

Clearly, $W_{f(i)}$ does not contain A. Furthermore, $W_i = W_{f(i)}$, if $A \not\subseteq W_i$.

Now without loss of generality assume $A = \{0, 1, 2, \ldots, m\}$, for some m. Let $W_{g(i)} = A \cup \{m+1, \ldots, m+1+i\} \cup \{m+1+i+2\}$, so that $W_{g(i)}$ is an extension of A and is not an initial segment of $\mathcal{N}$. $W_{g(i)}$ are pairwise distinct and every finite set is extended by infinitely many of $W_{g(i)}$'s. Furthermore, $\{W_{g(i)} \mid i \in \mathcal{N}\}$ is disjoint from $\{W_{\mathbf{M}(\sigma)} \mid \sigma \in \text{SEQ}\}$. Therefore by Lemma 17, $\mathcal{L} \in \mathbf{DecEx}$. ∎

Lemma 20 *Suppose every finite set has at least two extensions in $\mathcal{L}$. Suppose $a \in \mathcal{N} \cup \{*\}$ and $\mathbf{J} \in \{\mathbf{Ex}, \mathbf{Fex}_a, \mathbf{Bc}\}$. Then, $\mathcal{L} \subseteq \mathbf{DecJ}(\mathbf{M})$ iff $\mathcal{L} \subseteq \mathbf{WrDJ}(\mathbf{M})$.*

Proof. Suppose by way of contradiction that $\mathcal{L} \subseteq \mathbf{WrDJ}(\mathbf{M})$, $\mathcal{L} \not\subseteq \mathbf{DecJ}(\mathbf{M})$. 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(τ_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 τ_3. Then T witnesses that $\mathbf{M}$ does not $\mathbf{WrDJ}$-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 21 $\mathbf{DecEx} = \mathbf{WrDEx}$.

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

Case 1: $\mathcal{L}$ contains at least two extensions of every finite set. Then by Lemma 20, $\mathcal{L}$ is in $\mathbf{DecEx}$.

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 18, 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 19, we have that $\mathcal{L} \in \mathbf{DecEx}$. ∎

As $\mathbf{DecEx} \subset \mathbf{TxtEx}$ [3], we conclude that some families of languages in $\mathbf{TxtEx}$ 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 are not overinclusive), then all classes of languages in $\mathbf{TxtEx}$ become learnable, as the following result shows.

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

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

If $\mathcal{N} \in \mathcal{L}$, then $\mathcal{L} \in \mathbf{DecEx}$ (Proposition 18). So assume $\mathcal{N} \notin \mathcal{L}$. Let $\mathbf{M}$ be a machine such that, (I) $\mathbf{M}$ $\mathbf{TxtEx}$-identifies $\mathcal{L} \cup \mathrm{INIT}$, (II) $\mathbf{M}$ is prudent [5] and (III) all texts for $L \in \mathcal{L} \cup \mathrm{INIT}$, start with a $\mathbf{TxtEx}$-locking sequence for $\mathbf{M}$ on L. Note that Fulk [15] shows that this can be assumed without loss of generality. Also note that, for all k, if σ is a stabilizing sequence for $\mathbf{M}$ on $INIT_k$, then $\mathrm{content}(\sigma) = INIT_k$.

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

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

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

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 $W_{\mathbf{M}(T[m]),n} \supseteq \mathrm{content}(T[m])$ (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 on its domain.

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 $\mathrm{content}(T[n]) \subseteq W_{gram(T[n])}$, and $T[n]$ is valid, we immediately have that $T[m]$ is not a stabilizing sequence for $\mathbf{M}$ on $W_{\mathbf{M}(T[m])} = W_{\mathbf{M}(T[n])} \supseteq \mathrm{content}(T[n])$. Thus, $gram(T[m])$ follows the third clause in the definition of $gram$. Since, $\langle \mathrm{ind(T[m])}, \cdot\rangle \neq \langle \mathrm{ind(T[n])}, \cdot\rangle$, 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 $\mathrm{content}(T[n]) = W_{gram(T[n])}$ (since σ being stabilizing sequence for $\mathbf{M}$ on $INIT_k$ implies that $\mathrm{content}(\sigma) = INIT_k$). Thus, $W_{gram(T[m])} = W_{gram(T[n])} = \mathrm{content}(T[n]) \subseteq L$.

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

[5] $\mathbf{M}$ is said to be prudent [24] iff $\mathbf{M}$ $\mathbf{TxtEx}$-identifies every language W_i, such that i is in the range of $\mathbf{M}$.

By definition of **OGDEx** we have the following Corollary.

Corollary 23 **TxtEx** $\subseteq$ **OGDEx**.

Recall the following result about non U-shaped learners from [3].

Theorem 24 [3] (a) **TxtEx** $\not\subseteq$ **DecBc**.

(b) **TxtEx** $=$ **NUShEx**.

Clearly, **NUShEx** $\subseteq$ **NInvUEx** $\subseteq$ **TxtEx**. Thus from Theorem 24 we have the following.

Corollary 25 **NInvUEx** $=$ **TxtEx**.

Theorems 24 and 21 imply that forbidding return to abandoned wrong conjectures is more restrictive than forbidding return to abandoned correct conjectures in the context of **TxtEx**-learning. From Theorem 22, Corollaries 23 and 25, the latter requirement of forbidding return to abandoned correct conjectures is equivalent to forbidding inverted U's and to forbidding return to abandoned overinclusive or to overgeneralizing conjectures. We summarize these observations in the following immediate corollary.

Corollary 26 (a) **WrDEx** $\subset$ **NUShEx**.

(b) **NUShEx** $=$ **OIDEx** $=$ **OGDEx** $=$ **NInvUEx**.

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 [9].

Theorem 27 (a) $\mathbf{WrDFex}_* \subseteq \mathbf{WrDEx}$.

(b) $\mathbf{NInvUFex}_* \subseteq \mathbf{TxtEx}$.

(c) $\mathbf{OIDFex}_* \subseteq \mathbf{TxtEx}$.

Proof. (a) Suppose $\mathbf{M}$ $\mathbf{WrDFex}_*$-identifies $\mathcal{L}$, $L \in \mathcal{L}$ and T is a text for the language L.

Let us define a symmetric relation E_a as follows: $E_a(i,j)$ holds iff there exist n_1, n_2, n_3, n_4 such that $n_1 < n_2 < n_3 < n_4 < a$, $\mathbf{M}(T[n_1]) = \mathbf{M}(T[n_3])$, $\mathbf{M}(T[n_2]) = \mathbf{M}(T[n_4])$ and $\{\mathbf{M}(T[n_1]), \mathbf{M}(T[n_2])\} = \{i,j\}$ where $a \in \{0,1,2,\ldots,*\}$. That is, $E_a(i,j)$ holds if the learner alternates at least three times between these two hypotheses with possibly other hypotheses conjectured in between and this alternation occurs on an initial segment of length up to a; this last restriction on the length of the initial segment is void for $a = *$.

Note that $E_a(i,j)$ implies $W_i = W_j$: assuming by way of contradiction that $W_i \neq W_j$, the learner would return to the abandoned hypotheses W_i and W_j;

by definition of $\mathbf{WrDFex}_*$ both could not be wrong and thus they would have to be equal, contrary to the assumption.

By taking reflexive and transitive closure of E_a, we get an equivalence relation $\tilde{E}_a$. Note that still $W_i = W_j$ whenever $\tilde{E}_a(i,j)$.

It is easy to verify that for all grammars i, j which are output infinitely often by $\mathbf{M}$ on T, there is an n such that for all $a \geq n$ the predicate $\tilde{E}_a(i,j)$ holds.

Now a new learner $\mathbf{M}'$ is built by defining $\mathbf{M}'(T[n])$ to be a canonical index for the union of those W_e for which e satisfies $\tilde{E}_n(e, \mathbf{M}(T[n]))$.

First note that $\mathbf{M}'$ is well-defined since there are only finitely many such e with $\tilde{E}_n(e, \mathbf{M}(T[n]))$. Each such e has to be of the form $\mathbf{M}(T[m])$ for some $m \leq n$.

Second, $W_{\mathbf{M}'(T[n])} = W_{\mathbf{M}(T[n])}$ for all n. To see this note that $e = \mathbf{M}(T[n])$ satisfies $\tilde{E}_n(e, \mathbf{M}(T[n]))$ and thus the union is over a nonempty class of sets. Furthermore, all the sets in this class are equal since $\tilde{E}_n(e, \mathbf{M}(T[n]))$ implies $W_e = W_{\mathbf{M}(T[n])}$. So the union of the sets W_e with $\tilde{E}_n(e, \mathbf{M}(T[n]))$ is the set $W_{\mathbf{M}(T[n])}$.

Third, as $\mathbf{M}$ $\mathbf{WrDBc}$-identifies L from T so does $\mathbf{M}'$.

Fourth, it is easy to verify that all grammars i which are output infinitely by $\mathbf{M}$ belong to the same $\tilde{E}_*$-equivalence class D. Thus almost all hypotheses of $\mathbf{M}$ are members of this finite set D. There is an m such that $E_n(i,j)$ for all $i, j \in D$ and $n \geq m$. Thus $\mathbf{M}'(T[n])$ is the same index for the union of the W_e with $e \in D$ whenever $n \geq m$ and $\mathbf{M}(T[n]) \in D$. As this holds for almost all m, $\mathbf{M}'$ even $\mathbf{WrDEx}$-identifies L from T.

In particular, $\mathbf{M}'$ is a $\mathbf{WrDEx}$-identifier for $\mathcal{L}$.

(b) The proof is analogue to (a) with the following differences: The relation $\tilde{E}_a(i,j)$ does not imply that $W_i = W_j$ but just that $W_i = L \Leftrightarrow W_j = L$. So D contains only correct indices and $\mathbf{M}'$ is a $\mathbf{TxtEx}$-identifier for $\mathcal{L}$, but it is not guaranteed that the $\mathbf{NInvU}$-property is preserved. In particular the second and third item of the verification break down, but the fourth item can make use of the fact that all members of D are correct indices and thus the union of the sets with indices in D is the set L to be learned.

(c) This part is similar to part (b), except that in this case, we have that $E_a(i,j)$ implies $W_i \subseteq L \Leftrightarrow W_j \subseteq L$. This follows from the definition of $\mathbf{OIDFex}$-identification as either $W_i = W_j$ or both are subsets of the input language. Finally all indices which are output infinitely often are correct, so D contains at least one correct index and perhaps some additional indices of subsets of L. So the union of the sets with indices in D gives the set L. ∎

Since every explanatory learner is by definition also a vacillatory learner, the inclusion (a) of the previous Theorem is not proper. Furthermore, using Theorem 21 from the previous section, we actually get decisiveness on the right side of the equality. Furthermore, the second and third inclusion of the previous

Theorem can be improved by using the equalities $\mathbf{OIDEx} = \mathbf{NInvUEx} = \mathbf{TxtEx}$ (see Theorems 22 and 24).

Corollary 28 (a) $\mathbf{WrDFex}_* = \mathbf{DecEx}$.

(b) $\mathbf{OIDFex}_* = \mathbf{OIDEx}$.

(c) $\mathbf{NInvUFex}_* = \mathbf{NInvUEx}$.

From the above Corollary we can conclude that, as was the case for **TxtEx**-learning, **WrD** is more restrictive than **NUSh** while **NInvU** and **OID** are 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 return to abandoned wrong conjectures than by forbidding return to correct conjectures. Also, some results from [9] 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 [9].

Theorem 29 [9] $\mathbf{TxtFex}_2 \subseteq \mathbf{NUShBc}$; $\mathbf{TxtFex}_3 \not\subseteq \mathbf{NUShBc}$.

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 34) that there are **TxtEx**-learnable classes that *cannot* be **TxtBc**-learned by any **WrD**-learner [6]. 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.

We now show that forbidding return to old overgeneralizing conjectures still restricts the full learning power of vacillatory learning, but in a different and less severe way. First we show that, for each $n > 0$, there are classes that are **OGD**-learnable with vacillation between at most $n + 1$ correct conjectures but not learnable at all with vacillation between at most n conjectures. Thus the vacillatory hierarchy does not collapse when returning to overgeneralizing hypotheses is forbidden.

Theorem 30 $\mathbf{OGDFex}_{n+1} \not\subseteq \mathbf{TxtFex}_n$.

Proof. For each j let $X_j = \{\langle j, x\rangle \mid x \in \mathcal{N}\}$. Let $\mathcal{L} = \{L \mid (\exists j)[\emptyset \subset L \subseteq X_j$ and $\mathrm{card}(D_j) \leq n + 1$ and $(\exists p \in D_j)[L = W_p]$ and $(\forall k \in D_j)[L \not\subset W_k]]\}$.

Clearly, $\mathcal{L} \in \mathbf{OGDFex}_{n+1}$, as, on input σ with non-empty content, a learner can first obtain a j such that $L \subseteq X_j$, and then output the $p \in D_j$ which maximizes $|\tau_p|$, where τ_p is the maximal prefix of σ such that $\mathrm{content}(\tau_p) \subseteq$

[6] Observe that this result is *not* a trivial consequence of $\mathbf{TxtEx} \not\subset \mathbf{DecBc}$ from [3], since we show in the next section (Corollary 39) that $\mathbf{DecBc} \subset \mathbf{WrDBc}$.

$W_{p,|\sigma|}$; If content$(\sigma) = \emptyset$, then the learner outputs an hypothesis for $\emptyset$. This learner clearly $\mathbf{OGDFex}_{n+1}$-identifies $\mathcal{L}$.

The diagonalization proof is essentially based on the proof of $\mathbf{TxtFex}_{n+1} \not\subseteq \mathbf{TxtFex}_n$ from [10]. Suppose by way of contradiction that $\mathbf{M}$ $\mathbf{TxtFex}_n$-identifies $\mathcal{L}$. Then, by $(n+1)$-ary recursion theorem [26] there exist distinct $e_1, \ldots, e_{n+1}$ such that $W_{e_1}, \ldots, W_{e_{n+1}}$ may be defined as follows. Let j be such that $D_j = \{e_1, \ldots, e_{n+1}\}$. Initially let σ_0 be such that content$(\sigma_0) = \{\langle j, 0\rangle\}$, and enumerate $\langle j, 0\rangle$ in each of W_{e_i}, $1 \leq i \leq n+1$. Let $\mathbf{Last}_n(\sigma)$ denote the set of last n grammars output by $\mathbf{M}$ on input σ. That is $\mathbf{Last}_n(\sigma) = \{\mathbf{M}(\tau) \mid \tau \subseteq \sigma \ \wedge \ \text{card}(\{\mathbf{M}(\tau') \mid \tau \subseteq \tau' \subseteq \sigma\}) \leq n\}$. Go to stage 0.

Stage s
1. Dovetail steps 2 and 3 until, if ever, step 2 succeeds. If and when step 2 succeeds, stop step 3 and go to step 4.
2. Search for an extension τ of σ_s such content$(\tau) \subseteq X_j$ and $\mathbf{Last}_n(\tau) \neq \mathbf{Last}_n(\sigma_s)$.
3. **For** $s = 1$ to ∞ Do
   For $1 \leq k \leq n+1$, Enumerate $\langle j, \langle k, s\rangle\rangle$ into W_{e_k}.
   Endfor
4. Let $S = \bigcup_{1 \leq k \leq n+1}[W_{e_k}$ enumerated up to now].
   Let σ_{s+1} be an extension of τ such that content$(\sigma_{s+1}) =$ content$(\tau) \cup S$.
   For $1 \leq k \leq n+1$, enumerate content(σ_{s+1}) in W_{e_k}.
   Go to stage $s+1$.

End stage s.

We now consider the following cases.

Case 1: There exist infinitely many stages.

In this case, clearly, $W_{e_1} = W_{e_2} = \ldots = W_{e_{n+1}}$, and $\mathbf{M}$ does not converge to a set of n grammars on $\bigcup_{s \in \mathcal{N}} \sigma_s$, a text for $L = W_{e_1}$, which is a member of $\mathcal{L}$.

Case 2: Stage s starts but never ends.

In this case consider $L_k = W_{e_k}$, for $1 \leq k \leq n+1$. Note that $L_k \subseteq X_j$ and for $1 \leq k, k' \leq n+1$, $k \neq k'$: $L_k \not\subseteq L_{k'}$. Thus, each L_k belongs to $\mathcal{L}$. Furthermore for $1 \leq k \leq n+1$, for any text T for L_k which extends σ_s, all grammars which are output by $\mathbf{M}$ on T beyond σ_s, are from $\mathbf{Last}_n(\sigma_s)$ (otherwise step 2 would succeed as $L_k \subseteq X_j$). Thus, $\mathbf{M}$ fails to $\mathbf{TxtFex}_n$-identify at least one of L_k, $1 \leq k \leq n+1$ (since $\mathbf{Last}_n(\sigma_s)$ can contain grammars for at most n of $L_1, \ldots, L_{n+1}$).

It follows from above cases that $\mathbf{M}$ cannot $\mathbf{TxtFex}_n$-identify $\mathcal{L}$. ∎

Next we show that the learning power of each level of the vacillatory hierarchy is restricted when return to overgeneralizing conjectures is forbidden. More precisely, there are classes that are learnable with vacillation between two correct indices in the limit but such that no vacillatory learner can learn those classes without returning to overgeneralizing conjectures, no matter what amount of vacillation is allowed.

Theorem 31 $\mathbf{TxtFex}_2 \not\subseteq \mathbf{OGDFex}_*$.

Proof. Let $L_i = \{\langle i, x\rangle \mid x \in \mathcal{N}\}$. Let $S_i = \{\langle i, x\rangle \mid x \leq \mathrm{card}(W_i)\}$ (thus, if $\mathrm{card}(W_i) = \infty$, then $S_i = L_i$).

Let $\mathcal{L} = \{L_i \mid i \in \mathcal{N}\} \cup \{S_i \mid i \in \mathcal{N}\}$. It is easy to verify that $\mathcal{L} \in \mathbf{TxtFex}_2$.

Now suppose by way of contradiction that $\mathbf{M}$ $\mathbf{OGDFex}_*$-identifies $\mathcal{L}$. Then we show that $Inf = \{i \mid \mathrm{card}(W_i) = \infty\}$ is Σ_2, a contradiction to Π_2 completeness of Inf.

Since, $\mathbf{M}$ $\mathbf{OGDFex}_*$ identifies L_i, there exists a σ and finite set D such that (I) $\sigma \in \mathrm{SEG}(L_i)$, (II) $(\forall j, j' \in D)(\exists \tau_1, \tau_2, \tau_3 \subseteq \sigma)[\mathbf{M}(\tau_1) = j, \mathbf{M}(\tau_2) = j', \mathbf{M}(\tau_3) = j]$ and (III) $(\forall \tau \in \mathrm{SEG}(L_i))[\mathbf{M}(\sigma\tau) \in D]$.

Such σ can be obtained by just choosing a locking sequence σ for $\mathbf{M}$ on L_i, where each of the final grammars have alternated with each other. D can be taken to be the set of final grammars. Here note that D contains a grammar for L_i.

Let us denote by $Prop_i(\sigma, D)$, the combination of three properties (I), (II), (III).

Now if there exists a σ, D satisfying $Prop_i(\sigma, D)$ and $\mathrm{card}(W_i) \geq \max(\{x \mid \langle i, x\rangle \in \mathrm{content}(\sigma)\})$, then W_i is infinite (otherwise on σ, which is a member of $\mathrm{SEG}(S_i)$, $\mathbf{M}$ returns to a conjecture for L_i with grammar for S_i in between; to see this note that D contains a grammar for both S_i and L_i and by clause (II) above, $\mathbf{M}$ alternates between grammars for L_i and S_i on prefixes of σ).

On the other hand, if there does not exists a σ, D satisfying $Prop_i(\sigma, D)$ and $\mathrm{card}(W_i) \geq \max(\{x \mid \langle i, x\rangle \in \mathrm{content}(\sigma)\})$, then clearly, $\mathrm{card}(W_i)$ is finite (since σ, D satisfying $Prop_i(\sigma, D)$ would then show the finiteness of W_i).

The check whether

$$(\exists \sigma, D)[Prop_i(\sigma) \text{ and } \mathrm{card}(W_i) \geq \max(\{x \mid \langle i, x\rangle \in \mathrm{content}(\sigma)\})]$$

is a Σ_2 property. This gives us the desired contradiction. ∎

Corollary 32 *For all $a \geq 2$,* $\mathbf{OGDFex}_a \subset \mathbf{TxtFex}_a$.

Corollary 33 *For all $a \in \mathcal{N}^+$,* $\mathbf{OGDFex}_{a+1}$ *and* $\mathbf{TxtFex}_a$ *are incomparable.*

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 that is not allowed to return to wrong conjectures.

The reason is that the class $\mathcal{L}$ in **TxtEx** − **DecBc** from [3] contains two distinct extensions of every finite set and thus the next theorem follows from Lemma 20.

Theorem 34 **TxtEx** $\not\subseteq$ **WrDBc**.

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 **TxtEx** = **NUShEx** $\subseteq$ **NUShBc**, we have the following.

Corollary 35 **NUShBc** $\not\subseteq$ **WrDBc**.

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

Theorem 36 **WrDBc** $\not\subseteq$ **NUShBc**.

Proof. The proof uses the same class as in the proof of **TxtBc** $\neq$ **DecBc** from [16]. The proof that this class witnesses the theorem is a minor modification of the proof of Fulk, Jain and Osherson [16]. We give the details for completeness.

Let $L_j = \{\langle j, x\rangle \mid x \in \mathcal{N}\}$. Let $\sigma_{j,k} = (\langle j, 0\rangle, ..., \langle j, k\rangle)$, $L^1_{j,k} = \{\langle j, i\rangle \mid i \leq k\}$, and $L^2_{j,k} = W_{\mathbf{M}_j(\sigma_{j,k})}$.

Let $P(j, k)$ be the property that $L^1_{j,k} \subset L^2_{j,k} \subseteq L_j$.

If $(\exists k)[P(j, k)]$, then let k_j be the least k such that $P(j, k)$ holds, and then let $\mathcal{S}_j = \{L^1_{j,k_j}, L^2_{j,k_j}\}$; otherwise, let $\mathcal{S}_j = \{L_j\}$. Let $\mathcal{L} = \bigcup_j \mathcal{S}_j$.

We will show that $\mathcal{L} \in$ **WrDBc** − **NUShBc**.

The proof of $\mathcal{L} \in$ **WrDBc** is based on utilization of the fact that, if $(\exists k)[L^1_{j,k} \subset L^2_{j,k} \subseteq L_j]$, then the least such k can be found in the limit.

Claim 37 $\mathcal{L} \in$ **WrDBc**.

Proof. Note that $L \in \mathcal{S}_j \Rightarrow L \subseteq L_j$.

Let $Cand^n_j = \{k \leq n \mid L^1_{j,k} \subset W^n_{\mathbf{M}_j(\sigma_{j,k})} \subseteq L_j\}$.

Consider **M** which behaves as follows:

M on input $T[n]$
- **If** content$(T[n]) = \emptyset$, then output a grammar for $\emptyset$.
- **Else**, let j be such that content$(T[n]) \subseteq L_j$.
- (* If there is no such j, then the input language is not in the class $\mathcal{L}$.*)
  - **If** $Cand^n_j = \emptyset$
  - **Then** let $\mathbf{M}(T[n])$ be a grammar for L_j.
  - **Else**
    - Let $k^n_j = \min(Cand^n_j)$;
    - **If** content$(T[n]) \subseteq L^1_{j,k^n_j}$
    - **Then** let $\mathbf{M}(T[n]) = f(j, k^n_j, n)$, where f is as defined below.
    - **Else** let $\mathbf{M}(T[n]) = g(j, k^n_j, n)$, where g is as defined below.

Endif
Endif
Endif
End **M**

In the above, f and g are such that:

$$W_{f(j,k,n)} = \begin{cases} L^1_{j,k}, & \text{if } (\forall m > n)[\min(Cand^m_j) = \min(Cand^n_j)]; \\ L_j, & \text{otherwise.} \end{cases}$$

$$W_{g(j,k,n)} = \begin{cases} L^2_{j,k} \cap L_j, & \text{if } (\forall m > n)[\min(Cand^m_j) = \min(Cand^n_j)]; \\ L_j, & \text{otherwise.} \end{cases}$$

We claim that **M** **WrDBc**-identifies $\mathcal{L}$. Let T be a text for $L \in \mathcal{S}_j$. Now consider the following cases.

Case 1: $(\forall k)[\neg P(j,k)]$.

In this case $L = L_j$. Furthermore, for all n, either $Cand^n_j = \emptyset$ or there exists an $m > n$ such that $\min(Cand^n_j) \neq \min(Cand^m_j)$. Thus, if **M** outputs $f(j,k,n)$ or $g(j,k,n)$, then $[W_{f(j,k,n)} = W_{g(j,k,n)} = L_j]$. Thus **M** on $T[n]$ always outputs a grammar for L_j, except for the case when $\text{content}(T[n]) = \emptyset$. Thus, **M** **WrDBc**-*identifies* L.

Case 2: $P(j,k)$ holds for some k.

Let k_j be minimal such that $P(j,k)$ holds. Let m be minimal such that for all $n > m$, $\min(Cand^n_j) = \min(Cand^m_j) = k_j$.

Now, for $n < m$, either $Cand^n_j = \emptyset$ or for some $n' > n$, $\min(Cand^n_j) \neq \min(Cand^{n'}_j)$. Thus, if $\text{content}(T[n]) \neq \emptyset$, then by definition of **M** and $f(j,k,n)$ and $g(j,k,n)$, the grammar output by $\mathbf{M}(T[n])$ is for L_j.

For $n \geq m$, such that $\text{content}(T[n]) \neq \emptyset$, $\mathbf{M}(T[n])$, outputs $f(j,k_j,n)$ or $g(j,k_j,n)$, based on whether $\text{content}(T[n]) \subseteq L^1_{j,k_j}$ or not.

Thus, if $L = L^1_{j,k_j}$, then the sequence of grammars output by **M** on T are initially for $\emptyset$ (while $\text{content}(T[n]) = \emptyset$), followed by grammars for L_j (while $n < m$ and $\text{content}(T[n]) \neq \emptyset$), and eventually for L^1_{j,k_j} (when $n \geq m$ and $\text{content}(T[n]) \neq \emptyset$). Thus, **M** **WrDBc**-identifies L^1_{j,k_j}.

On the other hand, if $L = L^2_{j,k_j}$, then the sequence of grammars output by **M** on T are initially for $\emptyset$ (while $\text{content}(T[n]) = \emptyset$), followed by grammars for L_j (while $n < m$ and $\text{content}(T[n]) \neq \emptyset$), followed by grammars for L^1_{j,k_j} (while $n \geq m$ and $\emptyset \subset \text{content}(T[n]) \subseteq L^1_{j,k_j}$), and then eventually grammars for L^2_{j,k_j} (when $n \geq m$ and $\text{content}(T[n]) \not\subseteq L^1_{j,k_j}$). Thus, again **M** **WrDBc**-identifies L. Note that L^2_{j,k_j} might be equal to L_j, and thus decisiveness does not hold.

Claim 38 $\mathcal{L} \notin$ **NUShBc**.

Proof. Suppose by way of contradiction that machine $\mathbf{M}_j$ **NUShBc**-identifies $\mathcal{L}$.

Now consider $\mathcal{S}_j$.

If $(\forall k)[\neg P(j,k)]$, then $L_j \in \mathcal{L}$ which is not **TxtBc**-identified by $\mathbf{M}_j$.

If $(\exists k)[P(j,k)]$, then let k_j be the least such k. Now $L^1_{j,k_j}, L^2_{j,k_j} \in \mathcal{L}$. Since on σ_{j,k_j} $\mathbf{M}_j$ outputs a grammar for $L^2_{j,k_j} \neq L^1_{j,k_j}$, there must be extension σ of σ_{j,k_j} such that content$(\sigma) = L^1_{j,k_j}$ and $W_{\mathbf{M}_j(\sigma)} = L^1_{j,k_j}$. Also there must be an extension σ' of σ, such that content$(\sigma') \subseteq L^2_{j,k_j}$ and $W_{\mathbf{M}_j(\sigma')} = L^2_{j,k_j}$ (since $\mathbf{M}_j$ identifies both L^1_{j,k_j}, L^2_{j,k_j}). But then $\mathbf{M}_j$ is U-shaped on L^2_{j,k_j}. This proves the claim.

The theorem follows from above claims. ∎

Observe that, in contrast to the case of **TxtEx** and **TxtFex**-learning, Theorem 36 implies that **WrDBc** does *not* coincide with **DecBc**. We have in fact the following corollary of Theorem 36.

Corollary 39 **DecBc** $\subset$ **WrDBc**.

We next show that, as was the case for **Ex**, inverted-U-shapes are redundant for full **Bc**-learning power. In fact we have **NInvUBc** = **TxtBc**. For this, we first establish Corollary 43 below based on work of [20].

Definition 40 [15] $\mathbf{M}$ is said to be *rearrangement independent* iff for all σ, τ such that content(σ) = content(τ) and $|\sigma| = |\tau|$, $\mathbf{M}(\sigma) = \mathbf{M}(\tau)$.

Definition 41 [20] A sequence σ is *normalized* if $x \in$ content$(\sigma) \Rightarrow x \leq |\sigma|$.

A text T is normalized if $T[n]$ is normalized for all n.

Theorem 42 [20] *Suppose* $\mathbf{M}$ *is given. Then we can effectively define* $\mathbf{M}'$ *such that:*

(a) *If* $L \in$ **TxtBc**$(\mathbf{M})$*, then for all normalized texts* T *for* L*, for all but finitely many* n*,* $\mathbf{M}'(T[n])$ *is a grammar for* L*.*

(b) $\mathbf{M}'$ *is rearrangement independent.*

Corollary 43 *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*.*

Proof. Suppose $\mathbf{M}$ **TxtBc**-identifies $\mathcal{L}$ on normalized texts and $\mathbf{M}$ is rearrangement independent (by Theorem 42 such $\mathbf{M}$ exists). Let $\mathbf{M}'$ be defined as follows. $\mathbf{M}'(\sigma) = \mathbf{M}(\tau)$, where $|\tau| = 2 * |\sigma| + \max(\text{content}(\sigma))$ and content(τ) = content(σ), and τ is normalized. Clearly, $\mathbf{M}'$ is rearrangement independent.

Now consider any text T for $L \in \mathcal{L}$. Furthermore, let α be a **TxtBc**-locking sequence (on normalized texts) for $\mathbf{M}$ on L. Let n be such that content$(\alpha) \subseteq$ content$(T[n])$, and $|\alpha| \leq n$. Consider any σ such that content$(\sigma) \subseteq L$. Thus, we have that

$$\mathbf{M}'(T[n]\sigma) = \mathbf{M}(\alpha \#^r T[n]\sigma),$$

where $r = |\sigma|+n-|\alpha|+\max(\text{content}(T[n]\sigma))$. Thus, $\mathbf{M}'(T[n]\sigma)$ is a grammar for L. Hence, $T[n]$ is a **TxtBc**-locking sequence for $\mathbf{M}'$ on L and $\mathbf{M}'$ **TxtBc**-identifies L on T. ▮

Theorem 44 **TxtBc** $\subseteq$ **NInvUBc**.

Proof. Suppose **M** **TxtBc**-identifies $\mathcal{L}$. Without loss of generality (by Corollary 43) assume that every text for $L \in \mathcal{L}$ starts with a **TxtBc**-locking sequence for **M** on L. Using s-m-n theorem, let f be a recursive function such that

$W_{f(\sigma)} = \bigcup_{s\in\mathcal{N}} A_\sigma^s$, where A_σ^s is defined as follows.

$A_\sigma^0 = \text{content}(\sigma)$.

$A_\sigma^{s+1} = A_\sigma^s \cup \bigcup_{\tau\in\{\tau':|\tau'|\leq s, \text{content}(\tau')\subseteq A_\sigma^s\}} W_{\mathbf{M}(\tau),s}$.

Intuitively, $W_{f(\sigma)}$ is the smallest set S such that S contains $\text{content}(\sigma)$ and $W_{\mathbf{M}(\tau)}$ for every τ satisfying: $\sigma \subseteq \tau$, and $\text{content}(\tau) \subseteq S$.

Let $\mathbf{M}'(\sigma) = f(\sigma)$.

Now, it is easy to verify that if σ is **TxtBc**-locking sequence for **M** on L, then $W_{f(\sigma\tau)} = L$, for any τ such that $\text{content}(\tau) \subseteq L$. Thus, using the property that every text T for L starts with a **TxtBc**-locking sequence for **M** on L, we have that $\mathbf{M}'$ **TxtBc**-identifies $\mathcal{L}$.

The following claim follows from closure property of $W_{f(\sigma)}$.

Claim 45 *If $\sigma \subseteq \tau$, and $\text{content}(\tau) \subseteq W_{f(\sigma)}$, then $W_{f(\tau)} \subseteq W_{f(\sigma)}$.*

Now suppose T is a text for $L \in \mathcal{L}$, and $\sigma \subseteq \tau \subseteq \gamma \subseteq T$, are such that $W_{f(\sigma)} = W_{f(\gamma)} \neq W_{f(\tau)}$. Then, we have

(I) $\text{content}(\tau) \subseteq \text{content}(\gamma) \subseteq W_{f(\gamma)} = W_{f(\sigma)}$ and thus, by Claim 45 $W_{f(\tau)} \subseteq W_{f(\sigma)}$.

(II) If $\text{content}(\gamma) \subseteq W_{f(\tau)}$ then by Claim 45, $W_{f(\gamma)} \subseteq W_{f(\tau)}$, and thus using (I) we would have $W_{f(\tau)} = W_{f(\sigma)}$. A contradiction. Thus, $\text{content}(\gamma) \not\subseteq W_{f(\tau)}$.

It immediately follows from (II) that $W_{f(\tau)}$ is not a grammar for L.

It follows from above analysis that $\mathbf{M}'$ **NInvUBc**-identifies $\mathcal{L}$. ▮

Our next goal is to show that any **TxtBc**-learner can be transformed into one that does not return to overinclusive conjectures. For this, we first establish Lemmas 46 and 47.

Lemma 46 *Suppose* **M** *is given. Then there exists an r.e. set $P(\sigma)$ such that*

- *A grammar for $P(\sigma)$ can be effectively obtained from σ;*
- *If σ is a* **TxtBc**-*locking sequence for* **M** *on $W_{\mathbf{M}(\sigma)}$, then $P(\sigma)$ contains only grammars for $W_{\mathbf{M}(\sigma)}$;*
- *If σ is not a* **TxtBc**-*locking sequence for* **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 $\mathrm{content}(\sigma) \not\subseteq W_{\mathbf{M}(\sigma)}$, then let $P(\sigma) = \emptyset$.

Otherwise let $P(\sigma) = \{\mathbf{M}(\tau) \mid \sigma \subseteq \tau, \mathrm{content}(\tau) \subseteq W_{\mathbf{M}(\sigma)}\}$.

Now if σ is a **TxtBc**-locking sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$, then $P(\sigma)$ consists only of grammars for L. On the other hand if σ is not a **TxtBc**-locking sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$, then either $P(\sigma)$ is empty or it contains grammars for at least two distinct languages. ∎

Lemma 47 *Suppose* $\mathbf{M}$ *is given. Then, there exists a recursive function* g *such that:*

(a) *If* σ *is a* **TxtBc***-locking sequence for* $\mathbf{M}$ *on* $W_{\mathbf{M}(\sigma)}$*, then* $W_{g(\sigma)} = W_{\mathbf{M}(\sigma)}$*.*

(b) *If* σ *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

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

Let f be a recursive function such that, $W_{f(X)} = \bigcup_{s \in \mathcal{N}} CommonElem(X, s)$. Here we assume that $W_{f(\emptyset)} = \emptyset$.

Intuitively, $CommonTime(X, s)$ finds the largest time t such that all elements enumerated up to time t by some grammars in X are included in all languages enumerated by grammars in X up to time s. $CommonElem(X, s)$ then gives the set of the elements enumerated by all grammars in X up to 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 $CommonTime(X, s) \geq CommonTime(X', s)$;
(III) If $W_p \neq W_{p'}$ then for all s, $CommonTime(\{p, p'\}, s)$ is bounded by the least t such that $W_{p,t} \cup W_{p',t} \not\subseteq W_p \cap W_{p'}$.

Suppose $X_0 \subseteq X_1 \subseteq \ldots$ is given. Let Y be the set of all y such that there is an $s \geq y$, such that $y \in W_{f(X_s)}$. Note that (II) and (III) above imply that if $\{p, p'\} \subseteq \bigcup_{i \in \mathcal{N}} X_i$ and $W_p \neq W_{p'}$, 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 46 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 overinclusive conjectures.

Theorem 48 **TxtBc** $\subseteq$ **OIDBc**.

Proof. Suppose $\mathbf{M}$ **TxtBc**-identifies $\mathcal{L}$. Without loss of generality (Corollary 43) assume that for any text T for $L \in \mathcal{L}$, there exists a $\sigma \subseteq T$, such that

σ is a **TxtBc**-locking sequence for $\mathbf{M}$ on L. Intuitively, the proof employs two tricks. First trick (as given by g in Lemma 47) is to make sure that the infinite languages output by the learner are only on σ'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 47.

Let $q_0, q_1, \ldots$ denote an increasing sequence of primes.

Now define $\mathbf{M}''$ as follows. $\mathbf{M}''(\sigma) = h(\sigma)$, where $W_{h(\sigma)}$ is defined as follows. We assume without loss of generality that for all i and s, $W_{i,s+1} - W_{i,s}$ contains at most one element.

$W_{h(\sigma)}$
1. Enumerate content(σ)
2. Loop
   Search for s such that $W_{h(\sigma)}$ enumerated up to now is a proper subset of $W_{g(\sigma),s}$, and card($W_{g(\sigma),s}$) is $(q_{\mathrm{ind}(\sigma)})^k$ 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(σ) or has cardinality a power of $q_{\mathrm{ind}(\sigma)}$.

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

It follows from above cases that $\mathbf{M}''$ does not return to overinclusive hypotheses. 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 $\mathbf{M}$ on L (such an n exists by assumption on $\mathbf{M}$). 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 enumerate anything beyond first step, and thus equals L. ∎

Corollary 49 **TxtBc** $\subseteq$ **OGDBc**.

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 [4,24]. In this section, we establish a new interesting boundary on consistent **TxtEx**-learnability – in Theorem 53 we show that consistent **TxtEx**-learners can be made decisive (still being consistent) – contrast this result with Theorem 24.

Definition 50 [4,24] $\mathbf{M}$ is said to be *consistent* on T iff, for all n, $\mathbf{M}(T[n])\downarrow$ and $\text{content}(T[n]) \subseteq W_{\mathbf{M}(T[n])}$.

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

Definition 51 (a) [4,24] $\mathbf{M}$ **ConsTxtEx**-*identifies* L iff $\mathbf{M}$ is consistent on L, and $\mathbf{M}$ **TxtEx**-identifies L.

(b.1) [4] $\mathbf{M}$ **ConsTxtEx**-*identifies* $\mathcal{L}$ iff $\mathbf{M}$ **ConsTxtEx**-identifies each $L \in \mathcal{L}$.

(b.2) $\mathbf{ConsTxtEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathbf{M}\ \mathbf{ConsTxtEx}\text{-identifies } \mathcal{L}]\}$.

Note that for $\mathbf{M}$ to **ConsTxtEx**-identify a text T, it must be defined on each initial segment of T. One can similarly define combination of consistency with decisive (called **ConsDecEx**) and other related criteria such as **ConsNUShEx**, **ConsOIDEx**, **ConsWrDEx**, etc.

There are two other versions of consistency considered in the literature, namely $\mathcal{R}$**Cons** [19] where the learner must be total but might be inconsistent on data not belonging to the class to be learned and $\mathcal{T}$**Cons** [30] 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**.

Definition 52 We say that σ is *self-stabilizing* for $\mathbf{M}$ if σ is a **TxtEx**-locking sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$.

Theorem 53 $\mathbf{ConsTxtEx} \subseteq \mathbf{ConsDecEx}$.

Proof. Suppose $\mathbf{M}$ **ConsTxtEx**-identifies $\mathcal{L}$. An easy modification of the proof of Lemma 19 of current paper (along with corresponding modification of Lemma 17) can be used to show that, if there exists a finite set A such that no extension of A is in $\mathcal{L}$ then $\mathcal{L} \in \mathbf{ConsDecEx}$.

On the other hand, if $\mathcal{L}$ contains an extension of every finite set A, then $\mathbf{M}$ is total, and consistent on all inputs. Now for each σ define:

$$F_\sigma(x) = \begin{cases} 1, & \text{if } \mathbf{M}(\sigma x) = \mathbf{M}(\sigma); \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, F_σ is total for each σ. Furthermore, if σ is self-stabilizing for $\mathbf{M}$, then $F_\sigma^{-1}(1) = W_{\mathbf{M}(\sigma)}$. Thus, $\mathcal{L} \subseteq \{F_\sigma^{-1}(1) \mid \sigma \in \text{SEQ}\}$.

Let $G(2x) = 0$ and $G(2\langle i, x\rangle + 1) = 1 - F_{\delta_i}(2\langle i, x\rangle + 1)$. Thus, $G \neq^* F_\sigma$, for all σ, and G is 0 on all even inputs.

Let $s_\sigma = \min(\{t \mid (\exists \tau)[\mathrm{content}(\tau) \subseteq W_{\mathbf{M}(\sigma),t} \ \wedge \ |\tau| \leq t \ \wedge \ \mathbf{M}(\sigma) \neq \mathbf{M}(\sigma\tau)]\})$. Thus, $s_\sigma = \infty$ iff σ is self-stabilizing for $\mathbf{M}$. Moreover, one can effectively determine if $s_\sigma \leq t$, for any given t. Now define $h(\sigma), g(\sigma)$ as follows.

$$\varphi_{h(\sigma)}(x) = \begin{cases} F_\sigma(x), & \text{if } x \leq s_\sigma; \\ G(x), & \text{if } x > s_\sigma \text{ and } x \text{ is odd}; \\ 1, & \text{if } x > s_\sigma \text{ and } x = 2 * \langle 2 * \mathrm{ind}(\sigma), 1 + s_\sigma \rangle; \\ 0, & \text{otherwise}; \end{cases}$$

$$\varphi_{g(\sigma)}(x) = \begin{cases} 1, & \text{if } x \in \mathrm{content}(\sigma); \\ G(x), & \text{if } x \notin \mathrm{content}(\sigma) \text{ and } x \text{ is odd}; \\ 1, & \text{if } x = 2 * \langle 2 * \mathrm{ind}(\sigma) + 1, \max(\mathrm{content}(\sigma)) \rangle; \\ 0, & \text{otherwise}. \end{cases}$$

Intuitively, the aim of $h(\sigma)$ is to follow F_σ, if σ is self-stabilizing. Otherwise, it computes a finite variant of G. Third clause in the definition of $h(\sigma)$ above is used to ensure that the $h(\sigma)$'s which compute finite variant of G are pairwise different. $g(\sigma)$ is used below only for maintaining consistency, in case one cannot find an appropriate program $h(\sigma)$. Again, third clause in the definition of $g(\sigma)$ is to ensure that different $g(\sigma)$'s and $h(\sigma)$'s which compute finite variant of G are pairwise different.

It can be easily verified using the definition of g and h above that

(1) If σ is self-stabilizing for $\mathbf{M}$ then, $\varphi_{h(\sigma)} = F_\sigma$.

(2) If σ is not self-stabilizing for $\mathbf{M}$ then, $\varphi_{h(\sigma)} =^* G$, and $\max(\{x \mid \varphi_{h(\sigma)}(2x) = 1\})$ is of form $\langle 2 * \mathrm{ind}(\sigma), \cdot \rangle$. Note that $2 * \mathrm{ind}(\sigma)$ in the pair makes function $\varphi_{h(\sigma)}$ different from $\varphi_{g(\sigma)}$ and $\varphi_{h(\sigma')}/\varphi_{g(\sigma')}$, for $\sigma \neq \sigma'$.

(3) $\varphi_{g(\sigma)} =^* G$, and $\max(\{x \mid \varphi_{g(\sigma)}(2x) = 1\})$ is of form $\langle 2 * \mathrm{ind}(\sigma) + 1, \cdot \rangle$. Note that $2 * \mathrm{ind}(\sigma) + 1$ in the pair makes function computed $\varphi_{g(\sigma)}$ different from $\varphi_{h(\sigma)}$ and $\varphi_{h(\sigma')}/\varphi_{g(\sigma')}$, for $\sigma \neq \sigma'$.

Now define $\mathbf{M}'$ on T as follows. Let

$$\sigma_{\langle i,j \rangle} = \begin{cases} \delta_i, & \text{if } \mathrm{content}(\delta_i) \subseteq \mathrm{content}(T[\langle i,j \rangle]); \\ \lambda, & \text{otherwise}. \end{cases}$$

Let $gram$ be a recursive function such that $W_{gram(i)} = \varphi_i^{-1}(1)$.

$$\mathbf{M}'(T[n]) = \begin{cases} gram(h(\sigma_{m_n})), & \text{for least } m_n \leq n, \text{ such that} \\ & s_{\sigma_{m_n}} \geq n \text{ and} \\ & \mathrm{content}(T[n]) \subseteq F^{-1}_{\sigma_{m_n}}(1) \text{ and} \\ & \mathrm{content}(T[n]) \subseteq \varphi^{-1}_{h(\sigma_{m_n})}(1); \\ gram(g(T[n])), & \text{otherwise, if there is no such } m_n \leq n. \end{cases}$$

It is easy to verify that $\mathbf{M}'$ is consistent. Moreover, $\mathbf{M}'$ **TxtEx**-identifies $\mathcal{L}$, since for any text T for $L \in \mathcal{L}$, for the least m such that σ_m is a **TxtEx**-locking sequence for $\mathbf{M}$ on L, $\mathbf{M}'$ stabilizes to $gram(h(\sigma_m))$. We now show that $\mathbf{M}'$

is decisive. Note that m_n (when defined) is increasing in n. Suppose by way of contradiction, $W_{\mathbf{M}'(T[n_1])} = W_{\mathbf{M}'(T[n_3])} \neq W_{\mathbf{M}'(T[n_2])}$, where $n_1 < n_2 < n_3$. Note that if $\mathbf{M}'(T[n_1]) = gram(g(T[n_1]))$ or $\mathbf{M}'(T[n_3]) = gram(g(T[n_3]))$, then $W_{\mathbf{M}'(T[n_1])} \neq W_{\mathbf{M}'(T[n_3])}$ (by definition of $g(\cdot)$, and properties (2) and (3) above). Thus, $W_{\mathbf{M}'(T[n_1])} = W_{gram(h(m_{n_1}))}$ and $W_{\mathbf{M}'(T[n_3])} = W_{gram(h(m_{n_3}))}$. If $m_{n_1} = m_{n_3}$, then by monotonicity we will also have $m_{n_2} = m_{n_1}$, and thus $W_{\mathbf{M}'(T[n_2])} = W_{\mathbf{M}'(T[n_1])}$. On the other hand, if $m_{n_1} \neq m_{n_3}$, then $\sigma_{m_{n_1}}$ and $\sigma_{m_{n_3}}$ must both be self-stablizing for $\mathbf{M}$ since otherwise $W_{gram(h(\sigma_{m_{n_1}}))} \neq W_{gram(h(\sigma_{m_{n_3}}))}$ (by (1), (2) above and the fact that $G \neq^* F_\sigma$ for any σ). But, then content$(T[n_3]) \not\subseteq F^{-1}_{\sigma_{m_{n_1}}}(1) = W_{h(\sigma_{m_{n_1}})}$ (by definition of h, and the fact that $\sigma_{m_{n_1}}$ is not a stabilizing sequence for $\mathbf{M}$ on content$(\sigma_{m_{n_3}})$), a contradiction to $W_{gram(h(\sigma_{m_{n_1}}))} = W_{gram(h(\sigma_{m_{n_3}}))} \supseteq$ content$(T[n_3])$. It follows that $\mathbf{M}'$ must be decisive. ∎

Theorem 54 **NUShBc** = **ConsNUShBc**.

Proof. Suppose **M** **NUShBc**-identifies $\mathcal{L}$. Let $E(\sigma) = \{\tau \subseteq \sigma \mid \text{content}(\tau) = \text{content}(\sigma)\}$ and define $\mathbf{M}'$ as follows.

$$W_{\mathbf{M}'(\sigma)} = \begin{cases} W_{\mathbf{M}(\sigma)}, & \text{if } (\forall \tau \in E(\sigma))[\text{content}(\sigma) \subseteq W_{\mathbf{M}(\tau)}]; \\ \text{content}(\sigma), & \text{otherwise.} \end{cases}$$

Clearly, $\mathbf{M}'$ is consistent.

We now show that $\mathbf{M}'$ **NUShBc**-identifies $\mathcal{L}$. To see this, consider any text T for $L \in \mathcal{L}$. Let n be least number such that $W_{\mathbf{M}(T[n])} = L$. It follows by non U-shaped nature of $\mathbf{M}$ that for all $m \geq n$, $L = W_{\mathbf{M}(T[m])}$. We now consider two cases.

Case 1: L is finite.

In this case, for all $m \geq n$, $W_{\mathbf{M}'(T[m])} = L$, (based on either clause of the definition of $\mathbf{M}'$). Thus, $\mathbf{M}'$ **TxtBc**-identifies T. Now suppose there exists an $m' < n$ such that $W_{\mathbf{M}'(T[m'])} = L$. Then, since $W_{\mathbf{M}(T[m'])} \neq L$, we have that $W_{\mathbf{M}'(T[m'])}(= L) = \text{content}(T[m'])$, and content$(T[m']) \not\subseteq W_{\mathbf{M}(T[m'])}$. It follows by definition of $\mathbf{M}'$ that for all $m'' \geq m'$, $W_{\mathbf{M}'(T[m''])} = \text{content}(T[m'']) = L$. It follows that $\mathbf{M}'$ is non-U-shaped on T.

Case 2: L is infinite.

In this case, let $n' \geq n$ be minimal such that, content$(T[n]) \neq$ content$(T[n'])$. Clearly, for all $m \geq n'$, $\mathbf{M}'(T[m])$ is a grammar for L. Thus, $\mathbf{M}'$ **TxtBc**-identifies T. Furthermore, for all $m \leq n-1$, $\mathbf{M}'(T[m])$ is not a grammar for L (using either clause of the definition of $\mathbf{M}'$). Furthermore, for all m such that $n \leq m < n'$, $\mathbf{M}'(T[m])$ will be a grammar for L iff for all s such that content$(T[s]) =$ content$(T[n])$, content$(T[s]) \subseteq W_{\mathbf{M}(T[s])}$ (by Definition of $\mathbf{M}'$, and using the fact that content$(T[m']) \subseteq L = W_{\mathbf{M}(T[m'])}$, for $n \leq m' < n'$). It follows that $\mathbf{M}'$ is non-U-shaped on T. ∎

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

Theorem 55 **DecEx** $\not\subseteq$ **ConsTxtEx**.

Proof. Without loss of generality assume $\varphi_0(0)\uparrow$ (and thus $\varphi_0(0) \neq 0$).

Let $\mathcal{C} = \{f \mid (\forall^\infty x)[f(x) = 0]\} \cup \{f \mid f$ is monotonically increasing, $\varphi_{f(0)} = f$, and for all $x, \Phi_{f(0)}(x) < f(x+1)\}$. For a function f, let $L_f = \{\langle x, f(x)\rangle \mid x \in \mathcal{N}\}$. Let $\mathcal{L} = \{L_f \mid f \in \mathcal{C}\}$. It is well known that $\mathcal{C} \notin$ **ConsEx** (for function version of consistency, see for example [4]), and hence $\mathcal{L} \notin$ **ConsTxtEx**.

On the other hand, $\mathcal{L} \in$ **DecEx** can be shown as follows. Let p be a recursive function such that

$$\varphi_{p(i)}(x) = \begin{cases} i, & \text{if } x = 0 \text{ and } \varphi_i(0) = i; \\ \varphi_i(x), & \text{if } x > 0, \text{ and } (\forall y < x)[\max(\{\Phi_i(y), \varphi_i(y)\}) < \varphi_i(y+1)]; \\ \uparrow, & \text{otherwise.} \end{cases}$$

Now, **M** on input σ behaves as follows:

M(σ)
1. **If** $\langle 0, e\rangle \notin \mathrm{content}(\sigma)$ for any e then output a grammar for $\emptyset$.
2. **ElseIf** $\langle x, y_0\rangle, \langle x, y_1\rangle \in \mathrm{content}(\sigma)$, for some x, and $y_0 \neq y_1$, then output a grammar for $\mathcal{N}$.
4. **Else**
    5. Let $\langle 0, e\rangle \in \mathrm{content}(\sigma)$ (here such e is unique).
    6. **If** $\varphi_{e,|\sigma|}(0) \neq e$ or $[\mathrm{content}(\sigma) - \mathrm{content}(\sigma[\Phi_e(0)]) \subseteq \{\langle z, 0\rangle \mid z \in \mathcal{N}\}]$, then output a grammar for

$$L = \mathrm{content}(\sigma) \cup \{\langle x, 0\rangle \mid (\forall y > 0)[\langle x, y\rangle \notin \mathrm{content}(\sigma)]\}.$$

    7. **Else** let $m = \max(\{x \mid (\exists y)[\langle x, y\rangle \in \mathrm{content}(\sigma)]\})$.
    Let $m' = y$ such that $\langle m, y\rangle \in \mathrm{content}(\sigma)$.
        8. **If** $(\forall x \mid 0 < x < m)[\Phi_e(x) < m'$ and $\max(\{\Phi_e(x-1), \varphi_e(x-1)\}) < \varphi_e(x)]$ and $(\forall \langle x, y\rangle \mid x < m, \langle x, y\rangle \in \mathrm{content}(\sigma))[\varphi_e(x) = y]$, then output a grammar for $L_{\varphi_{p(e)}}$.
        9. **Else** Output a grammar for

$$L = \mathrm{content}(\sigma) \cup \{\langle x, 0\rangle \mid (\forall y > 0)[\langle x, y\rangle \notin \mathrm{content}(\sigma)]\}.$$

        10. **Endif**
    11. **Endif**
12. **Endif**

End

It is easy to verify that **M** **TxtEx**-identifies $\mathcal{L}$. To see decisiveness, note that grammar for $\emptyset$ appear only until some $\langle 0, x\rangle$ appears in the input. Furthermore, if a grammar for $\mathcal{N}$ is output on some σ, then for all $\tau \supseteq \sigma$, **M** outputs a grammar for $\mathcal{N}$. Once $\langle 0, e\rangle$ appears in input for some e, with $\varphi_e(0) = e$ (and the previous conjecture by **M** being known to be wrong, see step 6), $L_{\varphi_{p(e)}}$ is output as long as it is consistent with the input, and it seems that $\varphi_e \in \{f \mid f$ is monotonically increasing, $\varphi_{f(0)} = f$, and for all $x, \Phi_{f(0)}(x) < f(x+1)\}$. Thus, once $L_{\varphi_{p(e)}}$, is abandoned, it is never conjectured again. Furthermore, trivially, outputs in step 6 and 9 are monotonic in input. Thus, **M** is decisive. ∎

The proof of Theorem 44 also shows the following inclusion.

Theorem 56 **TxtBc** $\subseteq$ **ConsNInvUBc**.

The proof of Theorem 48 also shows the following inclusion.

Theorem 57 **TxtBc** $\subseteq$ **ConsOIDBc**.

The proof of Theorem 27 also works for the case when we are considering consistent identification.

Theorem 58 (a) **ConsWrDFex**$_*$ $\subseteq$ **ConsTxtEx**.

(b) **ConsOIDFex**$_*$ $\subseteq$ **ConsTxtEx**.

Corollary 59 (a) **ConsWrDFex**$_*$ $\subseteq$ **ConsDecEx**.

(b) **ConsOIDFex**$_*$ $\subseteq$ **ConsDecEx**.

The proof of Theorem 36 also shows the following.

Theorem 60 **ConsWrDBc** $\not\subseteq$ **NUShBc**.

Also note that Proofs of Theorems 30 and 31 give us:

Theorem 61 (a) **ConsTxtFex**$_2$ $\not\subseteq$ **OGDFex**$_*$.

(b) **ConsOGDFex**$_{n+1}$ $\not\subseteq$ **TxtFex**$_n$.

8 Summary and Open Problems

We summarize our results on the impact of the **WrD**, **NInvU**, **OID** and **OGD** variants of non U-shaped behaviour and how they compare to previous results about the original notion **NUSh** from [3] and [9].

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 inverted-U-shaped learning, returning to abandoned overinclusive conjectures and returning to abandoned overgeneralizing conjectures are necessary only for the vacillatory case and avoidable otherwise. This can be compared to results in [3] and [9] 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 return to abandoned wrong conjectures is more restrictive than forbidding 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 inverted U's, forbidding return to wrong overgeneralizing conjectures, and forbidding return to overinclusive conjectures are equivalent to forbidding return to correct conjectures for **TxtEx**. For **TxtFex**-identification instead, forbidding return to overgeneralizing conjectures is less restrictive than, equivalently, forbidding return to correct or overinclusive conjectures, and forbidding inverted U's.

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 **WrD** learner. This and the above observations may suggest that freedom of returning to wrong abandoned conjectures is even more central for full learning power, than 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.

The above results are illustrated in Figure 1, where an arrow indicates proper inclusion, a double arrow indicates equality and the absence of (transitive chains of) arrows indicates incomparability, except that it is open whether, for $a \geq 3$, $\mathbf{OGDFex}_a \subseteq \mathbf{NUShBc}$.

The following three questions are open:

(a) **ConsWrDBc** = **WrDBc**?

(b) **ConsDecBc** = **DecBc**?

(c) For $a \geq 3$ or $a = *$, is $\mathbf{OGDFex}_a \subseteq \mathbf{NUShBc}$?.

9 Acknowledgements

We would like to thank Rolf Wiehagen for useful discussions. The referees of COLT 2005 provided 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 and F. Stephan. Unlearning helps. In Ugo Montanari, José D. P. Rolim and Emo Welzl, editors, *Automata, Languages and Programming, 27th International Colloquium*, volume 1853 of *Lecture Notes in Computer Science*, pages 844–855. Springer-Verlag, 2000.

[3] 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 as [2].

$$
\begin{tikzcd}
 & \mathbf{Bc} \arrow[r, leftrightarrow] & \mathbf{OGDBc} \arrow[r, leftrightarrow] & \mathbf{NInvUBc} \arrow[r, leftrightarrow] & \mathbf{OIDBc} \\
 & \mathbf{OGDFex}_* \arrow[r] & \mathbf{Fex}_* \arrow[u] & \mathbf{NUShBc} \arrow[u] & \mathbf{WrDBc} \arrow[u] \\
 & \vdots & \vdots & & \\
 & \mathbf{OGDFex}_3 \arrow[r] & \mathbf{Fex}_3 & & \mathbf{DecBc} \arrow[uu] \arrow[uul] \\
 & \mathbf{OGDFex}_2 \arrow[r] \arrow[u] & \mathbf{Fex}_2 \arrow[u] \arrow[uuur] & & \\
 & \mathbf{NUShEx} \arrow[r, leftrightarrow] \arrow[u] \arrow[d, leftrightarrow] & \mathbf{NInvUEx} \arrow[d, leftrightarrow] & & \\
\mathbf{Ex} \arrow[r, leftrightarrow] & \mathbf{OGDEx} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \mathbf{OIDEx} \arrow[d, leftrightarrow] & & \\
 & \mathbf{NUShFex}_* \arrow[r, leftrightarrow] & \mathbf{NInvUFex}_* \arrow[r, leftrightarrow] & \mathbf{OIDFex}_* & \\
 & & \mathbf{WrDEx} \arrow[u] \arrow[r, leftrightarrow] & \mathbf{DecEx} \arrow[r, leftrightarrow] & \mathbf{WrDFex}_* \arrow[uuuuu]
\end{tikzcd}
$$

Figure 1. Summary of the results. An arrow indicates proper inclusion, a bidirectional arrow indicates equality. It is still open whether $\mathbf{OGDFex}_a \subseteq \mathbf{NUShBc}$, for $a \geq 3$.

[4] J. Bārzdiņš. Inductive inference of automata, functions and programs. In *Int. Math. Congress, Vancouver*, pages 771–776, 1974.

[5] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

[6] M. Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.

[7] 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.

[8] S. Carey. Face perception: Anomalies of development. In S. Strauss and R. Stavy, editors, *U-Shaped Behavioral Growth*, Developmental Psychology

Series. Academic Press, New York, 1982.

[9] 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. *Journal of Computer and System Sciences*, to appear.

[10] J. Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28(6):1941–1969, 1999.

[11] 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.

[12] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.

[13] C.A. Cashon and L.B. Cohen. Beyond U-shaped development in infants' processing of faces: An information-processing account. *Journal of Cognition and Development*, 5(1):59–80, 2004.

[14] C.H. Cashon and L.B. Cohen. The construction, deconstruction and reconstruction of infant face perception. In A. Slater and O. Pascalis, editors, *The development of face processing in infancy and early childhood*, pages 55–58. NOVA Science Publishers, New York, 2003.

[15] M. Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.

[16] M. Fulk, S. Jain and D. Osherson. Open problems in systems that learn. *Journal of Computer and System Sciences*, 49(3):589–604, 1994.

[17] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[18] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, 1979.

[19] K. P. Jantke and H.-R. Beick. Combining postulates of naturalness in inductive inference. *Journal of Information Processing and Cybernetics (EIK)*, 17:465–484, 1981.

[20] 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.

[21] S. Lange and R. Wiehagen. Polynomial time inference of arbitrary pattern languages. *New Generation Computing*, 8:361–370, 1991.

[22] M. Machtey and P. Young. *An Introduction to the General Theory of Algorithms*. North Holland, New York, 1978.

[23] 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.

[24] D. Osherson, M. Stob and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, 1986.

[25] 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.

[26] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, 1967. Reprinted by MIT Press in 1987.

[27] S. Strauss and R. Stavy. *U-Shaped Behavioral Growth*. Developmental Psychology Series. Academic Press, New York, 1982.

[28] S. Strauss, R. Stavy and N. Orpaz. The child's development of the concept of temperature. Manuscript, Tel-Aviv University, 1977.

[29] 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.

[30] R. Wiehagen and W. Liepe. Charakteristische Eigenschaften von erkennbaren Klassen rekursiver Funktionen. *Journal of Information Processing and Cybernetics (EIK)*, 12:421–438, 1976.