Cautious Limit Learning

Vanja Doskoč VANJA.DOSKOC@HPI.DE
Timo Kötzing TIMO.KOETZING@HPI.DE
Hasso Plattner Institute
University of Potsdam, Germany

Editors: Aryeh Kontorovich and Gergely Neu

Abstract

We investigate language learning in the limit from text with various *cautious* learning restrictions. Learning is *cautious* if no hypothesis is a proper subset of a previous guess. While dealing with a seemingly natural learning behaviour, cautious learning does severely restrict explanatory (syntactic) learning power. To further understand why exactly this loss of learning power arises, Kötzing and Palenta (2016) introduced weakened versions of cautious learning and gave first partial results on their relation.

In this paper, we aim to understand the restriction of cautious learning more fully. To this end we compare the known variants in a number of different settings, namely full-information and (partially) set-driven learning, paired either with the syntactic convergence restriction (explanatory learning) or the semantic convergence restriction (behaviourally correct learning). To do so, we make use of normal forms presented in Kötzing et al. (2017), most notably strongly locking and consistent learning. While strongly locking learners have been exploited when dealing with a variety of syntactic learning restrictions, we show how they can be beneficial in the semantic case as well. Furthermore, we expand the normal forms to a broader range of learning restrictions, including an answer to the open question of whether cautious learners can be assumed to be consistent, as stated in Kötzing et al. (2017).

Keywords: language learning in the limit, inductive inference, behaviourally correct learning, explanatory learning, cautious learning, normal forms

1. Introduction

Introduced by Gold (1967), in Computational Learning Theory we analyse the problem of algorithmically learning a description of a formal language when successively presented all and only the elements of that very language. For example, a learner h may be presented more and more odd numbers divisible by three. After each new input, h outputs a description of a language as its suggestion. While only being presented powers of three, the learner h might choose to output a description of the set of all powers of three as its suggestion. Once it sees an odd number divisible by three which is no power of three, it may change its mind to the set of all odd numbers divisible by three.

In his pioneer paper, Gold (1967) introduced a first criterion when such learning can be considered successful, called *explanatory* learning. We define when a learner h (a computable function) explanatory learns a target language L (a computably enumerable subset of the natural numbers). The learner is successively presented all and only the elements of L. A list of such elements is called a *text* T of the language L. With every new input, h makes a conjecture (a natural number inter-

preted as code for a computably enumerable set) which language it believes to be presented. Once h sticks to a single, correct description of the target language, we say that h successfully learned the target language L on text T. If h learns L on every text T of L, denoted by $T \in \mathbf{Txt}(L)$, we say that h $\mathbf{TxtGEx}$-*learned* the language L. Here, $\mathbf{Txt}$ indicates that only positive examples of the language are presented, $\mathbf{G}$, for *Gold-style* or *full-information* learning, specifies that the learner has full information on the elements presented so far and $\mathbf{Ex}$ stands for explanatory learning (giving a final, syntactically unchanging hypothesis *explaining* the data). Every *single* language can be learned by a $\mathbf{TxtGEx}$-learner which constantly outputs one and the same correct description of the language. Thus, we are interested in learning *classes* of languages, where a single learner h has to successfully $\mathbf{TxtGEx}$-learn each member of the class.

We strive to investigate the learning power of $\mathbf{TxtGEx}$-learners. To that end, we compare the set of all $\mathbf{TxtGEx}$-learnable classes, denoted as $[\mathbf{TxtGEx}]$, to other learning criteria. Such restrictions may be modelled to reflect an expected behaviour or be inspired by learning observed in nature. To provide an example, it seems natural that the suggestions of a learner always include the information they are based on. This is known as *consistent* learning, denoted as $\mathbf{Cons}$, see Angluin (1980). Although being a seemingly natural learning behaviour, consistent learning is known to severely lessen the learning power of explanatory learners, that is, $[\mathbf{TxtGConsEx}] \subsetneq [\mathbf{TxtGEx}]$. This is known as the *inconsistency phenomenon*, see Bārzdiņš (1977). This is due to $\mathbf{TxtGEx}$-learning requiring syntactic convergence for successful learning. In the semantic counterpart, where the learner needs to converge to a semantically correct, but not syntactically identical description of the target language, known as *behaviourally correct* learning ($\mathbf{Bc}$), see Case and Lynes (1982) and Osherson and Weinstein (1982), this phenomenon does not occur.

In this paper we investigate other seemingly natural learning restrictions, namely those which prohibit overgeneralization. For example, suggesting more than the target language may seem as an unnatural behaviour, especially considering that there is no way to refute this suggestion given only positive information. Learners that are *target cautious* do not show such a behaviour, see Kötzing and Palenta (2016). However, it is well-known that target cautious learners cannot achieve full learning power, see Kötzing and Palenta (2016). We prove that the same is true for the behaviourally correct case. Target cautious learning was proposed as an elegant way to deal with the more restrictive *cautious* learning restriction ($\mathbf{Caut}$), see Osherson et al. (1982), where the learner may never suggest a proper subset of any of its previous conjectures. As both cautious learning restrictions present a proper constraint, Kötzing and Palenta (2016) also investigated slightly different versions, in order to understand where this unexpected loss in learning power comes from. They proposed learning restrictions which are cautious only on finite, respectively infinite, suggestions, called *finitely cautious* learning ($\mathbf{Caut_{Fin}}$), respectively *infinitely cautious* learning ($\mathbf{Caut}_\infty$). While the behaviour of these learners is well-known in the explanatory case, see Kötzing and Palenta (2016), we provide the picture in the behaviourally correct case.

Another widely studied question in literature is whether learners need full information, $\mathbf{G}$, to maintain full learning power. For example, a learner may only be presented the set of elements shown so far, called *set-driven* learning ($\mathbf{Sd}$), see Wexler and Culicover (1980). This is known to severely weaken unrestricted explanatory learning, see Fulk (1990). However, additionally providing the total amount of elements shown so far, called *partially set-driven* or *rearrangement-independent* learning ($\mathbf{Psd}$), see Schäfer-Richter (1984) and Blum and Blum (1975), is enough to retain full learning power in this unrestricted case, see Fulk (1990). The question naturally trans-

lates to restricted explanatory learning. While only partial results to this question are known, we complete the picture for explanatory learning in Section 2, see Figure 1.

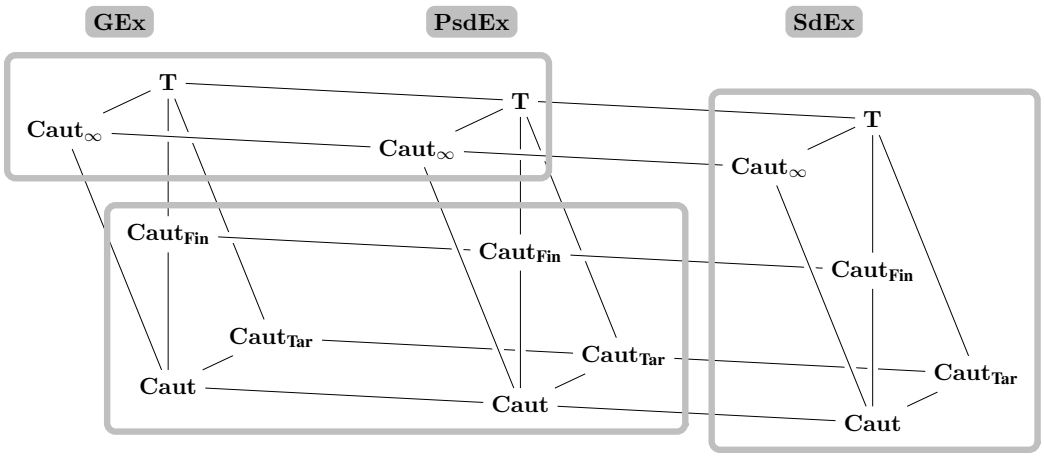

Figure 1: Relation of $[\mathbf{Txt}\beta\delta\mathbf{Ex}]$ for $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$ and various learning restrictions δ. While $\mathbf{T}$ indicates the absence of a restriction, black solid lines imply trivial inclusions (bottom-to-top, right-to-left) and greyly edged areas illustrate a collapse of the enclosed learning criteria. Our contributions are collected in Corollary 3, Lemma 4 and Theorem 5.

As semantic learners show a different behaviour in general, the question how cautious restrictions interfere with behaviourally correct learning is only natural. While obtaining the results for Figure 1, searching for locking sequences turned out to be a fruitful approach. Intuitively, a ($\mathbf{Bc}$) *locking sequence* contains sufficient information for the learner to guess the target language correctly and to prevent the learner from ever (semantically) changing its mind whatever information of the language is yet to come, see Jain et al. (1999). The partially set-driven and set-driven counterparts are called ($\mathbf{Bc}$) *locking information* and ($\mathbf{Bc}$) *locking set*, respectively. We also use the term ($\mathbf{Bc}$) locking information to subsume all three concepts. While it is known that there are learners and texts where no initial sequence is locking, Kötzing and Palenta (2016); Kötzing et al. (2017) showed when this undesired property can be forgone. A learner where every text has an initial sequence that serves as ($\mathbf{Bc}$) locking sequence is called *strongly* ($\mathbf{Bc}$) *locking*, see Kötzing and Palenta (2016). While, for explanatory learning, many ways to search for locking information are known, methods to do so in the behaviourally correct case are still sparse. In Section 3, we propose first approaches on how to search for $\mathbf{Bc}$-locking information, in order to obtain a full map depicting how cautious learning restrictions interfere with $\mathbf{Bc}$-learners, see Figure 2.
Furthermore, we have seen that consistency does not lessen the power of unrestricted $\mathbf{Bc}$-learners given full-information or (partially) set-driven information. In this case, we say that the restriction *allows for consistent* $\mathbf{Bc}$*-learning*. We ask, whether this holds true when adding further restrictions. While it is known that $\mathbf{Caut_{Tar}}$ allows for consistent $\mathbf{Bc}$-learning, see Kötzing et al. (2017), in Section 3, we provide the results for the remaining restrictions of interest. Most notably, we hereby solve an open problem stated by Kötzing et al. (2017).

Although syntactic learners are severely weaker than their semantic counterpart in general, comparing their learning power also gives interesting insights. Given the full pictures for explanatory

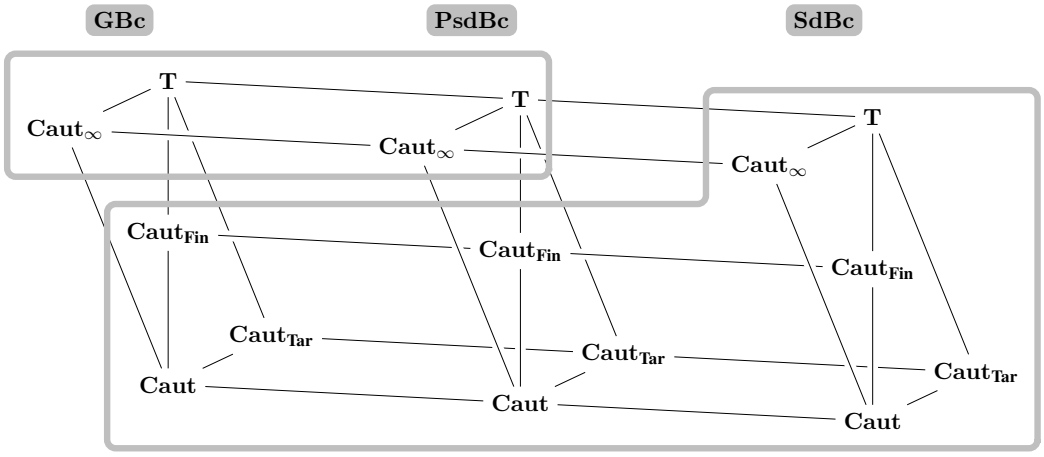

Figure 2: Relation of $[\mathbf{Txt}\beta\delta\mathbf{Bc}]$ for $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$ and various learning restrictions δ. While $\mathbf{T}$ indicates the absence of a restriction, black solid lines imply trivial inclusions (bottom-to-top, right-to-left) and greyly edged areas illustrate a collapse of the enclosed learning criteria. The only previously known results were $[\mathbf{TxtSdCaut_{Tar}Bc}] = [\mathbf{TxtPsdCaut_{Tar}Bc}]$, see Kötzing et al. (2017), and $[\mathbf{TxtSdBc}] \subsetneq [\mathbf{TxtPsdBc}] = [\mathbf{TxtGBc}]$, see Fulk (1990) and Carlucci et al. (2006). Furthermore, all learning can be assumed to be done consistently.

and behaviourally correct learning, see Figures 1 and 2, respectively, we draw the picture showing the full comparison, see Figure 3. The only non-trivial result here is the separation for $\mathbf{TxtGEx}$- and $\mathbf{TxtSdBc}$-learning, which follows from Fulk (1990) and Kötzing and Schirneck (2016).

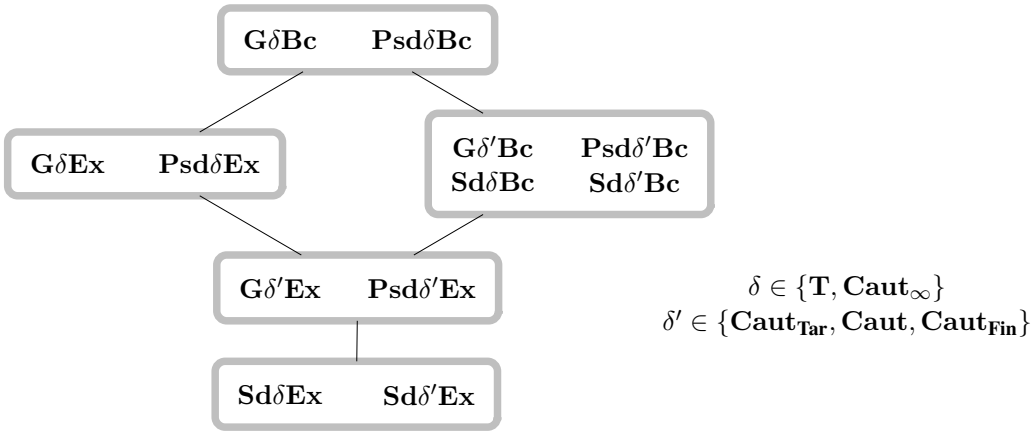

Figure 3: Relation of all considered learning restrictions ($\mathbf{Txt}$ is omitted for convenience). While $\mathbf{T}$ indicates the absence of a restriction, black solid lines imply trivial inclusions (bottom-to-top) and greyly edged areas illustrate a collapse of the enclosed learning criteria.

Throughout this paper we follow standard notations used in computability theory, for an overview see Rogers Jr. (1987). For the learning restrictions we follow Kötzing (2009). For a full overview on the formal setting and used notation we refer the reader to the appendix. There, the reader can also find the detailed proofs which are only sketched in the paper for reasons of space.

2. Cautious Ex-Learning

In this section, we provide a full map depicting the learning power of explanatory learners under various cautious restrictions. Whatever information is provided, i.e. full-information, set-driven or partially set-driven, cautious learners are known to achieve equal learning power as target cautious learners, see Kötzing and Palenta (2016); Kötzing and Schirneck (2016). Furthermore, in the full-information and set-driven case, these are as powerful as finitely cautious learners, see Kötzing and Palenta (2016). In Corollary 3, we show that the same holds true for the partially set-driven case. In addition, we show that partially set-driven information suffices, see Theorem 2 for the general result and Corollary 3 for the application to the current case, while set-driven information does not, see Lemma 4. Lastly, we show that just as in the full-information case, see Kötzing and Palenta (2016), partially set-driven infinitely cautious learning is as powerful as its unrestricted counterpart, see Theorem 5. As it is known that unrestricted partially set-driven learners are equally powerful as their Gold-style counterpart while set-driven learners are not, see Fulk (1990), this concludes the map. For convenience, we first gather the discussed known results in the next theorem.

Theorem 1 *Let* $\delta \in \{\mathbf{T}, \mathbf{Caut}_\infty, \mathbf{Caut}_{\mathbf{Tar}}, \mathbf{Caut}_{\mathbf{Fin}}, \mathbf{Caut}\}$. *Then, we have*

$$[\mathbf{TxtPsdCautEx}] = [\mathbf{TxtPsdCaut}_{\mathbf{Tar}}\mathbf{Ex}], \tag{1}$$

$$[\mathbf{TxtGCaut}_{\mathbf{Tar}}\mathbf{Ex}] = [\mathbf{TxtGCautEx}] = [\mathbf{TxtGCaut}_{\mathbf{Fin}}\mathbf{Ex}], \tag{2}$$

$$[\mathbf{TxtSd}\delta\mathbf{Ex}] = [\mathbf{TxtSdEx}]. \tag{3}$$

All of those are known to be separated from $[\mathbf{TxtPsdEx}] = [\mathbf{TxtGEx}] = [\mathbf{TxtGCaut}_\infty\mathbf{Ex}]$.

To get the full map for cautions **Ex**-learning as shown in Figure 1, we first show that target cautious learners do not need all information on the data given in order to maintain learning power, i.e. the learning restrictions from Equations (1) and (2) are equal in learning power. To that end, we make use of an idea from Fulk (1990), where an analogous result is shown for unrestricted explanatory learning. There, the partially set-driven learner mimics the Gold-style learner on possible locking sequences. It succeeds in learning once it finds the minimal such sequence.
We strive to generalize this result to a wide range of learning restrictions, namely, to restrictions δ where each hypothesis fulfils a predicate P also depending on languages. Formally, a learner h learns the language L under the restriction δ if and only if for every text T of the target language L the sequence p of hypotheses made by h fulfils predicate P pointwise. Notationally, that is, $\delta(p, T)$ if and only if for all i we have $P(p(i), \mathrm{content}(T))$. As every suggested hypothesis has to fulfil P, mimicking the learner will also maintain this property. We state this insight in the next theorem.

Theorem 2 *Let* P *be a predicate on hypotheses and languages. Let* δ *be a learning restriction such that*

$$\delta(p, T) \Leftrightarrow \forall i\colon P(p(i), \mathrm{content}(T)).$$

Then, $[\mathbf{TxtPsd}\delta\mathbf{Ex}] = [\mathbf{TxtG}\delta\mathbf{Ex}]$.

As target cautiousness depends only on the target language and the current hypothesis at a time, Theorem 2 can be applied to show that the learning restrictions from Equations (1) and (2) coincide in learning power.

Corollary 3 *We have* $[\mathbf{TxtPsdCaut_{Tar}Ex}] = [\mathbf{TxtGCaut_{Tar}Ex}]$ *and, in particular, for* $\beta \in \{\mathbf{G}, \mathbf{Psd}\}$ *and* $\delta \in \{\mathbf{Caut_{Tar}}, \mathbf{Caut_{Fin}}, \mathbf{Caut}\}$*, we have* $[\mathbf{Txt}\beta\delta\mathbf{Ex}] = [\mathbf{TxtGCautEx}]$.

Next, we show that the learning restrictions from Equations (1) and (2) separate from the learning restrictions in Equation (3). We do so by showing the existence of a class of languages which can be learned under the first restriction, but not the latter. This is done using *self-learning classes*, see Case and Kötzing (2016), and the Operator Recursion Theorem, see Case (1974).

Lemma 4 *We have* $[\mathbf{TxtPsdCautEx}] \setminus [\mathbf{TxtSdEx}] \neq \emptyset$.

To obtain the missing piece of Figure 1, we show that explanatory learners, if ever, only need to fall back to finite subsets of previous hypotheses. That is, $\mathbf{Caut}_\infty$ does not form a restriction in the partially set-driven case. We carry over the idea from the full-information proof, see Kötzing and Palenta (2016), where infinite sets were only enumerated if the underlying sequence was a locking sequence. As no information on the elements' presented order is available in partially set-driven learning, we have to put some additional work into choosing the right hypothesis to output.

Theorem 5 *We have* $[\mathbf{TxtPsdCaut}_\infty\mathbf{Ex}] = [\mathbf{TxtPsdEx}]$. *Particularly, for* $\beta \in \{\mathbf{G}, \mathbf{Psd}\}$ *and* $\delta \in \{\mathbf{T}, \mathbf{Caut}_\infty\}$*, we have* $[\mathbf{Txt}\beta\delta\mathbf{Ex}] = [\mathbf{TxtGEx}]$.

3. Cautious Bc-Learning

In the last section we have obtained a full comparison of the cautious learning variants for explanatory learning, see Figure 1. In this section, we do the same for the behaviourally correct case, see Figure 2. While the unrestricted learning behaves just as its syntactic counterpart, see Fulk (1990) and Carlucci et al. (2006), target cautious set-driven learners do accomplish the same learning power as their partially set-driven counterparts, see Kötzing et al. (2017). Again, we gather the discussed results for convenience.

Theorem 6 *We have*

$$\begin{aligned} [\mathbf{TxtSdBc}] &\subsetneq [\mathbf{TxtPsdBc}] = [\mathbf{TxtGBc}], \\ [\mathbf{TxtSdCaut_{Tar}Bc}] &= [\mathbf{TxtPsdCaut_{Tar}Bc}]. \end{aligned} \tag{4}$$

Target cautious learning already shows a different behaviour than the explanatory counterpart, see Equation (4), indicating that this map will turn out differently. For the remaining, we elaborate the set-driven part in Section 3.1, and then continue with incorporating the partially set-driven and full-information results in Section 3.2.

3.1. Forward Verification and Backwards Search

We have seen that set-driven explanatory learners can be assumed to be cautious without losing learning power. We show that the same holds true in the behaviourally correct case, see Theorem 9. We will do so stepwise. For the further discussion, let h be a learner, let L be a target language and let $\sigma, \tau \in L_{\#}^{*}$ be finite sequences of elements from $L_{\#} := L \cup \{\#\}$, where $\#$ is a *pause symbol*.

As we have seen, searching for locking sequences for h is a fruitful attempt in order to attain the learning power of h. While in explanatory learning, where syntactic convergence is required, $h(\sigma) \neq h(\sigma\tau)$ implies that σ cannot be a locking sequence for h on L, this does not hold true for the semantic counterpart. We elaborate a way to search for $\mathbf{Bc}$-locking sequences. By doing so, we show that, amongst other useful properties, $\mathbf{Sd}$-learning can be done target cautiously in general. In the set-driven behaviourally correct case, the *Weak Forward Verification* (**WFV**), see Algorithm 1, serves as a first step to circumvent this problem.

Algorithm 1: Weak Forward Verification (**WFV**), h_w

Parameter: $\mathbf{Sd}$-learner h, function enum such that $\forall e\colon W_e = \mathrm{range}(\mathrm{enum}(e, .))$.
Input: Finite set $D \subseteq \mathbb{N}$.
Semantic Output: $W_{h_w(D)} = \bigcup_{i \in \mathbb{N}} E_i$.
Initialization: $E_0 \leftarrow D$.

```
1 for i = 0 to ∞ do
2     x_i ← enum(h(D), i)
3     if x_i ∉ E_i then
4         for D', D ⊆ D' ⊆ E_i ∪ {x_i} do
5             search for t such that E_i ∪ {x_i} ⊆ W^t_{h(D')}
6         end
7     end
8     E_{i+1} ← E_i ∪ {x_i}
9 end
```

The intuition is the following. Given a finite input $D \subseteq L$, **WFV** will start by enumerating D. Now, at step i, let E_i be what **WFV** has enumerated so far and let x_i be the element newly enumerated by $h(D)$, see line 2. If D were a $\mathbf{Bc}$-locking set for h on L, every possible next hypothesis $h(D')$, with $D \subseteq D' \subseteq E_i \cup \{x_i\}$, would have to witness at least $E_i \cup \{x_i\}$, see lines 4 and 5. If all of this is witnessed, chances that D is a $\mathbf{Bc}$-locking set are still sustained, thus, **WFV** enumerates x_i and continues with step $i + 1$.

As every $\mathbf{Sd}$-learner is strongly $\mathbf{Bc}$-locking, see Kötzing et al. (2017), the **WFV** algorithm, upon enumerating the whole target language L, also has to enumerate $\mathbf{Bc}$-locking sets for h on L. These sets, in the checking phase of the **WFV**, see lines 4 and 5, will prevent the algorithm from enumerating more than the target language, resulting in target cautious learning.

Lemma 7 *We have* $[\mathbf{TxtSdCaut}_{\mathbf{Tar}}\mathbf{Bc}] = [\mathbf{TxtSdBc}]$.

The **WFV** approach is extendable. While we wait for every possible hypothesis $h(D')$ to witness at least $E_i \cup \{x_i\}$, other elements could be witnessed as well, that is, for the minimal t in line 5 of Algorithm 1 we have $E_i \cup \{x_i\} \subsetneq W^t_{h(D')}$. We show how to exploit such elements in the

search for $\mathbf{Bc}$-locking sets. If D, and thus every D' as well, were $\mathbf{Bc}$-locking sets, all elements in $W^t_{h(D')}$ must be elements of the target language, and thus every hypothesis $h(D')$ also would have to witness these elements. We capture the idea of extending the check from Algorithm 1, lines 4 and 5, in the *Strong Forward Verification* (**SFV**), see Algorithm 2, lines 4-8. For later usage, we state the algorithm in a generalized form, accepting any β-learner.

Algorithm 2: Strong Forward Verification (**SFV**), h_s

Parameter: Learner h, function enum such that $\forall e\colon W_e = \mathrm{range}(\mathrm{enum}(e,.))$.
Input: Finite sequence σ.
Semantic Output: $W_{h_s(\sigma)} = \bigcup_{i\in\mathbb{N}} E_i$.
Initialization: $E_0 \leftarrow D$.

1 **for** $i = 0$ **to** ∞ **do**
2 $x_i \leftarrow \mathrm{enum}(h(\sigma), i)$
3 **if** $x_i \notin E_i$ **then**
4 **for** $\tau'' \in (E_i \cup \{x_i\})^{\leq i}_{\#}$ **do**
5 $s_{\tau''} \leftarrow \min\{s\colon E_i \cup \{x_i\} \subseteq W^s_{h(\sigma\tau'')}\}$
6 **end**
7 **for** $\tau' \in (E_i \cup \{x_i\})^{\leq i}_{\#}$ **do**
8 search for t such that $\bigcup_{\tau''\in(E_i\cup\{x_i\})^{\leq i}_{\#}} W^{s_{\tau''}}_{h(\sigma\tau'')} \subseteq W^t_{h(\sigma\tau')}$
9 **end**
10 **end**
11 $E_{i+1} \leftarrow E_i \cup \{x_i\}$
12 **end**

The extended forward verification yields useful properties. We gather these in the next proposition, extending some which have been observed already by Carlucci et al. (2006) and providing new ones.

Proposition 8 *Let $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$. Given a learner h and with it the learner h_s as built in Algorithm 2, the following properties hold.*

(i) *If h is a β-learner, then h_s is a β-learner which is consistent on arbitrary input.*

(ii) *If σ_0 is a $\mathbf{Bc}$-locking information for h on some $L \subseteq \mathbb{N}$, then σ_0 is a $\mathbf{Bc}$-locking information for h_s on L.*

(iii) *For[1] $\beta \neq \mathbf{G}$ target cautious learning is preserved by the learner h_s, that is, we have that $\mathbf{Txt}\beta\mathbf{Caut}_{\mathbf{Tar}}\mathbf{Bc}(h) \subseteq \mathbf{Txt}\beta\mathbf{Caut}_{\mathbf{Tar}}\mathbf{Bc}(h_s)$.*

(iv) *If $W_{h_s(\sigma)}$ is infinite, then $W_{h_s(\sigma)} = W_{h(\sigma)} =: L$ and σ is a $\mathbf{Bc}$-locking information for h and h_s on L.*

(v) *If $L \in \mathbf{Txt}\beta\mathbf{Caut}_{\mathbf{Tar}}\mathbf{Bc}(h)$ and σ_0 is a $\mathbf{Bc}$-locking information for h_s on L, then σ_0 is a $\mathbf{Bc}$-locking information for h on L.*

1. As it will turn out, the same holds true for $\beta = \mathbf{G}$, see Corollary 11.

(vi) *Let h, and thus h_s, be* $\mathbf{Sd}$*-learners. Let D_0 be a* $\mathbf{Bc}$*-locking set for h on some L. Then, for D with either (a) $D \subseteq D_0$ or (b) $D_0 \subseteq D \subseteq L$, we have*

$$D_0 \subseteq W_{h_s(D)} \Rightarrow W_{h_s(D)} \subseteq L.$$

So far, we have seen various ways to search for $\mathbf{Bc}$-locking sequences. While this search maintains $\mathbf{Bc}$-learning and provides interesting properties, cautious learning seems to be unattainable this way. We establish a way to solve this problem. As in cautious learning preceding hypotheses remain important, we include these into the enumeration. Let h be a learner and D some finite input. We start by enumerating $E_0 = D$. At step i, let E_i be the elements enumerated so far. It seems like a promising idea to check whether for some $D' \subseteq D$ a previous hypothesis $h(D')$ exceeds what is enumerated so far, i.e. whether we have $E_i \subseteq W^i_{h(D')}$. If so, for the first such occurring hypothesis $h(D')$, enumerate $W^i_{h(D')}$ and proceed with the next step. This idea is captured in the *Backwards Search* (**BS**), see Algorithm 3.

Algorithm 3: Backwards Search (**BS**), h_b

Parameter: $\mathbf{Sd}$-learner h.
Input: Finite set $D \subseteq \mathbb{N}$.
Semantic Output: $W_{h_b(D)} = \bigcup_{i \in \mathbb{N}} E_i$.
Initialization: $E_0 \leftarrow D$.

```
for i = 0 to ∞ do
    if ∃D' ⊆ D: W^i_{h(D')} ⊋ E_i then
        for the first such D': E_{i+1} ← W^i_{h(D')}
    else
        E_{i+1} ← E_i
    end
end
```

Unfortunately, in general, this approach does not provide cautious learning. This is due to more information D yielding more possible previous hypotheses $h(D')$ which can lead the strategy from Algorithm 3 to wrong hypotheses. However, by combining the **SFV** and the **BS** and by exploiting Proposition 8 (iv) and (vi), we can circumvent this problem. In the next theorem, we use $\tau(\delta)$ to indicate that the restriction δ is also satisfied on arbitrary input.

Theorem 9 *We have* $[\tau(\mathbf{Cons})\mathbf{TxtSdCautBc}] = [\mathbf{TxtSdBc}]$.

Proof The inclusion $[\tau(\mathbf{Cons})\mathbf{TxtSdCautBc}] \subseteq [\mathbf{TxtSdBc}]$ follows immediately. For the other direction, let h be a total learner and let $L \in \mathbf{TxtSdBc}(h)$, that is, the language L can be $\mathbf{TxtSdBc}$-learned by h. By Lemma 7, we may assume $L \in \mathbf{TxtSdCaut_{Tar}Bc}(h)$. By Proposition 8 (iii), we may even assume the learning to be done by h_s from Algorithm 2, i.e. $L \in \mathbf{TxtSdCaut_{Tar}Bc}(h_s)$. This way, we are allowed to exploit Proposition 8. Now, let h_b be as in Algorithm 3 with h_s as parameter. We will show $L \in \mathbf{TxtSdConsCautBc}(h_b)$ step by step.

First, we show that $L \in \mathbf{TxtSdBc}(h_b)$. Let $T \in \mathbf{Txt}(L)$. For finite L, let n_0 be such that $\mathrm{content}(T[n_0]) = L$. Then, for all $n \geq n_0$, we get $W_{h_b(\mathrm{content}(T[n]))} = L$ as $W_{h_b(\mathrm{content}(T[n]))}$ starts by enumerating L and never enumerates any more elements as $\neg\left(\exists D' \subseteq L\colon W_{h_s(D')} \supsetneq L\right)$

due to h_s being $\mathbf{Caut}_{\mathbf{Tar}}$.
For infinite L, let n_0 be such that $D_0 \coloneqq \mathrm{content}(T[n_0])$ is a $\mathbf{Bc}$-locking set for h_s on L, see Kötzing et al. (2017). Let $n \geq n_0$ and $D \coloneqq \mathrm{content}(T[n])$. We study the candidates for a possible enumeration, i.e. $D' \subseteq D$, with $W_{h_s(D')} \supseteq D$. We may have the following two situations.

(I) Either $W_{h_s(D')}$ is infinite and, due to[2] Proposition 8 (iv), equal to L,

(II) or $W_{h_s(D')}$ is finite and, due to Proposition 8 (vi), a subset of L.

As $D \supseteq D_0$, there exists a D' fulfilling (I). As these are the only candidates to be enumerated into $W_{h_b(D)}$, we observe $W_{h_b(D)} \subseteq L$.
To prove $L \subseteq W_{h_b(D)}$, assume the opposite, that is, there exists some $x \in L \setminus W_{h_b(D)}$. For each $D' \subseteq D$ with $D \subseteq W_{h_s(D')}$ define $s_{D'}$ in the following way. Either, if x is enumerated into $W_{h_s(D')}$, then $s_{D'}$ is the last step before that very enumeration. Or, if x is never to be enumerated into $W_{h_s(D')}$, then $W_{h_s(D')}$ must be finite as it cannot be equal to L, see (I). In this case, $s_{D'}$ will be the first step where the enumeration of $W_{h_s(D')}$ is finished. Formally, we define

$$s_{D'} \coloneqq \begin{cases} \max\left\{s : x \notin W^s_{h_s(D')}\right\}, & x \in W_{h_s(D')}, \\ \min\left\{s : W^s_{h_s(D')} = W_{h_s(D')}\right\}, & \text{else.} \end{cases}$$

So, no later than at step $\max\{s_{D'} : D' \subseteq D \wedge D \subseteq W_{h_s(D')}\}$ the enumeration of $W_{h_b(D)}$ has to be finished, as any further enumeration would result in x being an element of $W_{h_b(D)}$. However, then $W_{h_b(D)}$ is a finite subset of L. Since there exists at least one D' fulfilling (I), the enumeration would have to continue, and thus enumerate x into $W_{h_b(D)}$, a contradiction to the assumption. Altogether, we have $W_{h_b(D)} = L$ and thus $\mathbf{TxtSdCaut}_{\mathbf{Tar}}\mathbf{Bc}(h_s) \subseteq \mathbf{TxtSdBc}(h_b)$.

Next, we want to show that h_b is $\mathbf{Caut}$. In order to do so, assume the opposite, i.e. there exist $D_1 \subseteq D_2$, with $D_2 \subseteq L$, such that $W_{h_b(D_1)} \supsetneq W_{h_b(D_2)}$. For finite $W_{h_b(D_2)}$, let i_0 be the step where $W_{h_b(D_2)}$ is completely enumerated, that is, $W^{i_0}_{h_b(D_2)} = W_{h_b(D_2)}$. As $W_{h_b(D_1)} \supsetneq W_{h_b(D_2)}$, there also must exist some $i_1 \geq i_0$ such that $W^{i_1}_{h_b(D_1)} \supsetneq W^{i_0}_{h_b(D_2)}$. Without loss of generality, we may assume that i_1 is also the point where $W^{i_1}_{h_b(D_1)}$ got enumerated, i.e. $W^{i_1}_{h_s(D')} = W^{i_1}_{h_b(D_1)}$ for some $D' \subseteq D_1$. But now, since $D' \subseteq D_2$ and $W^{i_1}_{h_s(D')} = W^{i_1}_{h_b(D_1)} \supsetneq W^{i_1}_{h_b(D_2)}$, the enumeration of $W_{h_b(D_2)}$ would have to continue, a contradiction.
If $W_{h_b(D_2)}$ is infinite, then there exists $D'' \subseteq D_2$ such that $W_{h_s(D'')} = W_{h_b(D_2)}$ is infinite and thus, by Proposition 8 (iv), D'' is a $\mathbf{Bc}$-locking set for h_s on $W_{h_s(D'')}$. Analogously, since $W_{h_b(D_1)} \supsetneq W_{h_b(D_2)}$, $W_{h_b(D_1)}$ is infinite too, and there also exists some $D' \subseteq D_1$ such that $W_{h_s(D')} = W_{h_b(D_1)}$ and thus D' is a $\mathbf{Bc}$-locking set for h_s on $W_{h_s(D')}$. However, $D_2 \subseteq W_{h_s(D'')} \subsetneq W_{h_s(D')}$ and D_2 is a superset of both D' and D''. Hence, D_2 is a $\mathbf{Bc}$-locking set for h_s on both $W_{h_s(D')}$ and $W_{h_s(D'')}$, which are different, a contradiction.

Observing that h_b is $\tau(\mathbf{Cons})$ by definition, we get $L \in \tau(\mathbf{Cons})\mathbf{TxtSdCautBc}(h_b)$, which finishes the proof. ■

2. By Proposition 8 (iv), D' must be a $\mathbf{Bc}$-locking set for h_s on $W_{h_s(D')}$. Now, as $D \supseteq D_0$ and $D \supseteq D'$, D must be both a $\mathbf{Bc}$-locking set for h_s on L and $W_{h_s(D')}$, respectively. Thus, $L = W_{h_s(D')}$.

3.2. When full-information learning is necessary

In the previous section we completed the behaviourally correct set-driven map. It remains to study the full-information and partially set-driven case. Firstly, we show that for $\mathbf{Caut}_{\mathbf{Tar}}$ and $\mathbf{Caut}_{\mathbf{Fin}}$ these are equal in learning power, see Corollary 11 and Theorem 12, respectively. Afterwards, we show how $\mathbf{Caut}_{\infty}$ fits into the picture, see Theorem 13. In the end, we get Corollary 14 which provides the whole picture.

We start by showing when full-information and partially set-driven learners may be assumed equally powerful, just as in Theorem 2 in the explanatory case. Unfortunately, the same approach does not bear fruits, as, although performing a search for $\mathbf{Bc}$-locking sequences, we do not mimic the learner. Rather, we enumerate the learner's output on possible $\mathbf{Bc}$-locking sequences, as discussed in private communication with Jain (2017). If, for certain languages, the Gold-style learner refrains from suggesting more than the target language, our enumeration can maintain this behaviour. To formally cover this in the next theorem, recall the setup in Section 2, just before Theorem 2, regarding the notation used. Again, we use $\tau(\delta)$ to indicate that the restriction δ is also satisfied on arbitrary input.

Theorem 10 *Let P be a predicate on languages. Let δ be a learning restriction such that*

$$\delta(p,T) \Leftrightarrow \big(P(\mathrm{content}(T)) \Rightarrow \mathbf{Caut}_{\mathbf{Tar}}(p,T)\big).$$

Then,

1. *δ allows for consistent $\mathbf{Bc}$-learning, that is, for any interaction operator $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$ we have $[\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}]$, and*
2. *$[\mathbf{TxtPsd}\delta\mathbf{Bc}] = [\mathbf{TxtG}\delta\mathbf{Bc}]$.*

In Theorem 10, choosing $\top$ as predicate P results in target cautious learning, immediately providing the following corollary.

Corollary 11 *We have*

$$[\mathbf{TxtPsdCaut}_{\mathbf{Tar}}\mathbf{Bc}] = [\mathbf{TxtGCaut}_{\mathbf{Tar}}\mathbf{Bc}].$$

To deal with $\mathbf{Caut}_{\mathbf{Fin}}$, we introduce a slightly less restrictive version on which we can apply results established throughout this paper. In its core, this is a similar approach to Kötzing and Palenta (2016) introducing $\mathbf{Caut}_{\mathbf{Tar}}$ in order to deal with $\mathbf{Caut}$.

Theorem 12 *We have*

$$[\mathbf{TxtSdCaut}_{\mathbf{Fin}}\mathbf{Bc}] = [\mathbf{TxtPsdCaut}_{\mathbf{Fin}}\mathbf{Bc}] = [\mathbf{TxtGCaut}_{\mathbf{Fin}}\mathbf{Bc}].$$

To conclude the behaviourally correct cautious map, it remains to show that infinitely cautious learning, i.e. $\mathbf{Caut}_{\infty}$, does not restrict the learning power of full-information and partially set-driven learners. We use the same idea as in the explanatory case, namely by ensuring that infinite suggestions only occur when the underlying information is a $\mathbf{Bc}$-locking information. We use the **SFV**, see Algorithm 2, to do so.

Lemma 13 *We have* $[\tau(\mathbf{Cons})\mathbf{TxtPsdCaut}_\infty\mathbf{Bc}] = [\mathbf{TxtPsdBc}]$.

Using the results obtained throughout Sections 3.1 and 3.2, we can sum up the map depicted in Figure 2 in the following corollary.

Corollary 14 *For* $\delta \in \{\mathbf{Caut}, \mathbf{Caut}_{\mathbf{Tar}}, \mathbf{Caut}_{\mathbf{Fin}}\}$ *and* $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$ *as well as* $\delta' \in \{\mathbf{T}, \mathbf{Caut}_\infty\}$ *and* $\beta' \in \{\mathbf{G}, \mathbf{Psd}\}$*, we have*

$$[\tau(\mathbf{Cons})\mathbf{TxtSdCautBc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{TxtSdBc}],$$
$$[\tau(\mathbf{Cons})\mathbf{TxtPsdCaut}_\infty\mathbf{Bc}] = [\mathbf{Txt}\beta'\delta'\mathbf{Bc}] = [\mathbf{TxtGBc}].$$

In particular, the previous result shows that any cautious learning restriction can be assumed consistent in general. This answers an open problem stated by Kötzing et al. (2017), namely whether $\mathbf{Caut}$ learning can be done consistently. Furthermore, it answers the same question for all considered cautious restrictions.

Corollary 15 *Let* $\delta \in \{\mathbf{T}, \mathbf{Caut}_\infty, \mathbf{Caut}_{\mathbf{Tar}}, \mathbf{Caut}_{\mathbf{Fin}}, \mathbf{Caut}\}$*. Then,* δ *allows for consistent* $\mathbf{Bc}$*-learning, that is, for* $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$ *we have* $[\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}]$,

4. Conclusion and Future Work

We have shown how cautious learning restrictions interfere with learning power in several learning settings. In particular, we give a full overview of all pairwise relations of the learning restrictions considered, as depicted in Figures 1, 2 and 3. To obtain the syntactic cautious map, namely Figure 1, we conducted searches for locking sequences as done in previous literature, see Blum and Blum (1975); Fulk (1990); Kötzing and Palenta (2016) for example. However, ways to exploit $\mathbf{Bc}$-locking sequences in the semantic counterpart are not prevalent. We propose first approaches to do so, namely with the Weak and Strong Forward Verification, see Algorithms 1 and 2, respectively. While these only serve as a first step to search for $\mathbf{Bc}$-locking sequences, it remains open how these searches can be beneficial in other settings, that is, when investigating other learning criteria, see Jain et al. (1999) for an overview.

Furthermore, we also focused on consistency. While syntactic learning is known to be restricted by consistency, many semantic learning restrictions are not. Unrestricted behaviourally correct learning, in addition to target cautious learning are only a few examples, see Kötzing et al. (2017). We extend this list, adding all restrictions considered in this paper. In particular, this solves an open problem stated by Kötzing et al. (2017). Extending this list further is left for future research.

Acknowledgments

We would like to thank Sanjay Jain for providing valuable input and insights through private communication. Furthermore, we are thankful for the helpful and constructive comments from the anonymous reviewers. This work was supported by the German Research Foundation (DFG) under Grant KO 4635/1-1 (SCL).

References

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

Jānis M. Bārzdiņš. Inductive inference of automata, functions and programs. In *American Mathematical Society Translations*, pages 107–122, 1977.

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

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

Lorenzo Carlucci, Sanjay Jain, Efim B. Kinber, and Frank Stephan. Variations on U-shaped learning. *Information and Computation*, 204:1264–1294, 2006.

John Case. Periodicity in generations of automata. *Mathematical Systems Theory*, 8:15–32, 1974.

John Case and Timo Kötzing. Strongly non-U-shaped language learning results by general techniques. *Information and Computation*, 251:1–15, 2016.

John Case and Christopher Lynes. Machine inductive inference and language identification. In *Proceedings of the 9th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 107–115, 1982.

John Case and Samuel E. Moelius. Optimal language learning from positive data. *Information and Computation*, 209:1293–1311, 2011.

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

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

Sanjay Jain. Personal communication, 2017.

Sanjay Jain, Daniel Osherson, James S. Royer, and Arun Sharma. *Systems that Learn: An Introduction to Learning Theory*. MIT Press, Cambridge (MA), Second Edition, 1999.

Timo Kötzing and Raphaela Palenta. A map of update constraints in inductive inference. *Theoretical Computer Science*, 650:4–24, 2016.

Timo Kötzing and Martin Schirneck. Towards an atlas of computational learning theory. In *Proceedings of the 33rd Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 47:1–47:13, 2016.

Timo Kötzing, Martin Schirneck, and Karen Seidel. Normal forms in semantic language identification. In *Proceedings of the 28th International Conference on Algorithmic Learning Theory (ALT)*, pages 76:493–76:516, 2017.

Timo Kötzing. *Abstraction and Complexity in Computational Learning in the Limit*. PhD thesis, University of Delaware, 2009.

Daniel Osherson, Michael Stob, and Scott Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, Cambridge (MA), 1986.

Daniel N. Osherson and Scott Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.

Daniel N. Osherson, Michael Stob, and Scott Weinstein. Learning strategies. *Information and Control*, 53:32–51, 1982.

Hartley Rogers Jr. *Theory of recursive functions and effective computability*. Reprinted by MIT Press, Cambridge (MA), 1987.

Gisela Schäfer-Richter. *Über Eingabeabhängigkeit und Komplexität von Inferenzstrategien*. PhD thesis, RWTH Aachen University, Germany, 1984.

Kenneth Wexler and Peter W. Culicover. Formal principles of language acquisition. *MIT Press, Cambridge (MA)*, 1980.

Appendix A. Language Learning in the Limit

In this section we collect the notations and preliminary results used throughout this paper. For a background on computability theory we refer to Rogers Jr. (1987). For the learning restrictions, we follow the system given by Kötzing (2009).

A.1. Preliminaries

Starting with the mathematical notation, we use $\subsetneq$ and $\subseteq$ to denote the proper subset and subset relation between sets, respectively. With $\subseteq_{\text{Fin}}$ we denote finite subsets. With $\mathbb{N} = \{0, 1, 2, \dots\}$ we denote the set of all natural numbers and, if not stated otherwise, e, i, j, k, n, s, t are elements thereof. We let $\mathfrak{P}$ and $\mathfrak{R}$ be the set of all partial and total functions $p\colon \mathbb{N} \to \mathbb{N}$, respectively. The subset of all computable (partial) functions is $(\mathcal{P})\ \mathcal{R}$. We fix an effective numbering $\{\varphi_e\}_{e\in\mathbb{N}}$ of $\mathcal{P}$ and let $W_e = \operatorname{dom}(\varphi_e)$ denote the e-th computably enumerable set. This way, we interpret the natural number e as a hypothesis for the set W_e.

We aim to learn recursively enumerable sets $L \subseteq \mathbb{N}$, also called *languages*. A *learner* is a partial computable function $h \in \mathcal{P}$. By $\#$ we denote the *pause* symbol and for any set S we denote $S_{\#} \coloneqq S \cup \{\#\}$. We use an *enumeration function* $\operatorname{enum}(.,.)$, where for every $e \in \mathbb{N}$ we have $W_e = \operatorname{range}(\operatorname{enum}(e,.))$. A *text* is a total function $T\colon \mathbb{N} \to \mathbb{N} \cup \{\#\}$, the collection of all texts is $\mathbf{Txt}$. For any text or sequence T, we let $\operatorname{content}(T) \coloneqq \operatorname{range}(T) \setminus \{\#\}$ be the *content* of T. A text T of a language L is such that $\operatorname{content}(T) = L$, the collection of all texts of L is $\mathbf{Txt}(L)$. By $T[n]$ we denote the initial sequence of T of length n, i.e. $T[n] \coloneqq (T(0), \dots, T(n-1))$ and $T[0] \coloneqq \varepsilon$. On finite sequences, we use $\subseteq$ to denote the *extension relation* and $\leq$ to denote the order on sequences interpreted as natural numbers. Also, we define an order $\preceq$ on tuples of the form (D, t), where $D \subseteq \mathbb{N}$ and $t \in \mathbb{N}$, as $(D, t) \preceq (D', t')$ iff $t \leq t'$ and there is a text $T \in \mathbf{Txt}$ such that $\operatorname{content}(T[t]) = D$ and $\operatorname{content}(T[t']) = D'$.

For learning, an *interaction operator* is an operator β which takes a learner $h \in \mathcal{P}$ and a text $T \in \mathbf{Txt}$ as arguments and outputs a possibly partial function p. Intuitively, β defines what kind of information the learner will have available to produce its guesses. For example, the interaction operators $\mathbf{G}$ for *full-information* or *Gold-style* learning, see Gold (1967), $\mathbf{Psd}$ for *partially set-driven* or *rearrangement independent* learning, see Schäfer-Richter (1984) and Blum and Blum (1975), and $\mathbf{Sd}$ for *set-driven* learning, see Wexler and Culicover (1980), are defined as

$$\begin{aligned}
\mathbf{G}(h, T)(i) &\coloneqq h(T[i]),\\
\mathbf{Psd}(h, T)(i) &\coloneqq h(\operatorname{content}(T[i]), i),\\
\mathbf{Sd}(h, T)(i) &\coloneqq h(\operatorname{content}(T[i])).
\end{aligned}$$

While the Gold-style learner has full information on the input, the partially set-driven learner has no information on the order of the input, and the set-driven learner only has access to the elements presented.

We can distinguish between different criteria for successful learning. E.g., one could require the learner to syntactically converge to the correct hypothesis, known as *explanatory* learning $\mathbf{Ex}$, see Gold (1967), or to semantically converge to the correct hypothesis, which then is called *behaviourally correct* learning $\mathbf{Bc}$, see Case and Lynes (1982) and Osherson and Weinstein (1982). Formally, a *learning restriction* is a predicate δ defined on a total learning sequence, i.e. total func-

tion, p and a text $T \in \mathbf{Txt}$. So, we have

$$\mathbf{Ex}(p,T) :\Leftrightarrow \exists n_0\, \forall n \geq n_0 \colon p(n) = p(n_0) \wedge W_{p(n)} = \mathrm{content}(T),$$
$$\mathbf{Bc}(p,T) :\Leftrightarrow \exists n_0\, \forall n \geq n_0 \colon W_{p(n)} = \mathrm{content}(T).$$

To model certain learning behaviours found in nature, one can add further restrictions to the hypotheses on the way. For example, it may seem reasonable not to suggest strictly less than any previous hypothesis. We call such behaviour *cautious* learning ($\mathbf{Caut}$), see Osherson et al. (1986). Formally,

$$\mathbf{Caut}(p,T) :\Leftrightarrow \forall i < j \colon \neg(W_{p(j)} \subsetneq W_{p(i)}).$$

In order to deal with a restriction that affects more than one hypothesis at a time, it proved useful to add stepwise more lenient restrictions. In the case of cautious learning, this has been done in Kötzing and Palenta (2016). There, three different types of cautious learning were introduced, namely *target cautious* ($\mathbf{Caut}_{\mathbf{Tar}}$), *infinitely cautious* ($\mathbf{Caut}_{\infty}$) and *finitely cautious* ($\mathbf{Caut}_{\mathbf{Fin}}$) learning. Intuitively, target cautious learning prevents the learner from outputting proper supersets of the target language, while infinite and finite cautiousness demand cautiousness only on infinite or finite instances, respectively. Formally, we define

$$\mathbf{Caut}_{\mathbf{Tar}}(p,T) :\Leftrightarrow \forall i \colon \neg(\mathrm{content}(T) \subsetneq W_{p(i)}),$$
$$\mathbf{Caut}_{\infty}(p,T) :\Leftrightarrow \left(\forall i < j \colon W_{p(j)} \subsetneq W_{p(i)} \Rightarrow W_{p(j)} \text{ is finite}\right),$$
$$\mathbf{Caut}_{\mathbf{Fin}}(p,T) :\Leftrightarrow \left(\forall i < j \colon W_{p(j)} \subsetneq W_{p(i)} \Rightarrow W_{p(j)} \text{ is infinite}\right).$$

Finally, the constantly true predicate $\mathbf{T}$ denotes the absence of a learning restriction.

Now, a *learning criterion* is a tuple $(\alpha, \mathcal{C}, \beta, \delta)$, where $\mathcal{C}$ is a set of admissible learners, typically $\mathcal{P}$ or $\mathcal{R}$, β is an interaction operator and α and δ are learning restrictions. We write $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$ to denote this learning criterion, omitting $\mathcal{C}$ in case of $\mathcal{C} = \mathcal{P}$ and the learning restriction if it equals $\mathbf{T}$. For an admissible learner $h \in \mathcal{C}$, we say that h *$\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learns* a language L iff on arbitrary texts $T \in \mathbf{Txt}$ we have $\alpha(\beta(h,T),T)$, and on texts for the target language $T \in \mathbf{Txt}(L)$ we have $\delta(\beta(h,T),T)$. With $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta(h)$ we denote the class of languages $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learned by h, and with $[\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta]$ the set of all $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learnable classes of languages.

A.2. Normal Forms

When mathematically dealing with learners, certain properties come in handy. For example, it is more convenient if the learner may be assumed to be total. Kötzing and Palenta (2016); Kötzing et al. (2017) state when this is the case. For example, Kötzing and Palenta (2016) show that this is the case for full-information delayable restrictions. Informally, a restriction is delayable if hypotheses can be postponed arbitrarily, but not indefinitely. Formally, a learning restriction δ is *delayable* iff for all texts T and T' with $\mathrm{content}(T) = \mathrm{content}(T')$, all learning sequences p and all unbounded non-decreasing functions r, if $\delta(p,T)$ and, for all n, $\mathrm{content}(T[r(n)]) \subseteq \mathrm{content}(T'[n])$, then $\delta(p \circ r, T')$. A learning restriction δ is called *semantic* if for any learning sequences p, p' and text T, $\delta(p,T)$ and, for all n, $W_{p(n)} = W_{p'(n)}$ implies $\delta(p',T)$. Intuitively, any hypothesis could be replaced by any semantically equivalent hypothesis. By Kötzing et al. (2017), the learners for any semantic learning restrictions can be assumed to be total. Thus, we may assume all semantic learners considered to be total.

Another very useful object of desire are so-called locking sequences. They encapsulate sufficient information for the learner to identify the target language and prevent it from changing its mind anymore. Formally, let $h \in \mathcal{P}$ be a $\mathbf{G}$-learner. A sequence $\sigma \in L^*_\#$ is called *locking sequence* for h on L iff for every sequence $\tau \in L^*_\#$ we have $h(\sigma) = h(\sigma\tau)$ and $W_{h(\sigma\tau)} = L$, see Blum and Blum (1975). The transfer to the $\mathbf{Psd}$ and $\mathbf{Sd}$ case is immediate. An information (D, t) is called *locking information* for h on L iff for every $(D', t') \succeq (D, t)$, with $D' \subseteq L$, we have $h(D, t) = h(D', t')$ and $W_{h(D',t')} = L$. A set D is called *locking set* for h on L iff for every D', with $D \subseteq D' \subseteq L$, we have $h(D) = h(D')$ and $W_{h(D')} = L$. We use the term locking information to subsume all three cases. The three $\mathbf{Bc}$-equivalents are defined analogously, for completeness, we state the $\mathbf{G}$-case. A sequence $\sigma \in L^*_\#$ is called $\mathbf{Bc}$*-locking sequence* for h on L iff for every sequence $\tau \in L^*_\#$ we have $W_{h(\sigma\tau)} = L$, see Jain et al. (1999). By an important observation by Blum and Blum (1975), every learner has a ($\mathbf{Bc}$-) locking sequence. However, it is well-known that there are learners where no initial sequence of a given text serves as locking sequence. In certain cases, such undesired behaviour can be bypassed, as shown in Kötzing and Palenta (2016); Kötzing and Schirneck (2016); Kötzing et al. (2017). Following them, we call a learner h *strongly (*$\mathbf{Bc}$*-) locking* on some language L iff for every text $T \in \mathbf{Txt}(L)$ there exists a position n_0 such that $T[n_0]$ is a ($\mathbf{Bc}$-) locking sequence. If h is strongly ($\mathbf{Bc}$-) locking on every language it learns, then we call h strongly ($\mathbf{Bc}$-) locking. The transfer to the $\mathbf{Psd}$- and $\mathbf{Sd}$-case is omitted because it is immediate.

Lastly, consistency will play a key role. We say that learning is *consistent* if the hypotheses always include the current information. Formally,

$$\mathbf{Cons}(p, T) :\Leftrightarrow \forall i\colon \operatorname{content}(T[i]) \subseteq W_{p(i)}.$$

Although being a natural requirement, consistency can form a severe restriction at times, see Fulk (1990). The picture changes when considering the $\mathbf{Bc}$-case. We say a restriction δ allows for *consistent* $\mathbf{Bc}$*-learning* iff, for every $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$, every language learned $\mathbf{Txt}\beta\delta\mathbf{Bc}$ can be learned $\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}$. By Kötzing et al. (2017), we already know that $\mathbf{T}$ and $\mathbf{Caut}_{\mathbf{Tar}}$ allow for consistent $\mathbf{Bc}$-learning. We will later extend this to all considered restrictions.

Appendix B. Omitted Proofs of Section 2

Theorem 2 *Let P be a predicate on hypotheses and languages. Let δ be a learning restriction such that*

$$\delta(p, T) \Leftrightarrow \forall i\colon P(p(i), \operatorname{content}(T)).$$

Then, $[\mathbf{TxtPsd}\delta\mathbf{Ex}] = [\mathbf{TxtG}\delta\mathbf{Ex}]$.

Proof The inclusion $[\mathbf{TxtPsd}\delta\mathbf{Ex}] \subseteq [\mathbf{TxtG}\delta\mathbf{Ex}]$ is trivial. For the other, let h be a learner and $L \in \mathbf{TxtG}\delta\mathbf{Ex}(h)$. As δ is delayable, we may assume h to be total without losing generality, see Kötzing and Palenta (2016). Now, we define a learner h' to search for the minimal, possible locking sequence given a finite set D and $t \geq 0$ as information. Formally, with $D^{\leq t}_\#$ being the set of all sequences of elements in $D_\# := D \cup \{\#\}$ of at most length t, we define h' as

$$M_{D,t} := \left\{\sigma \in D^{\leq t}_\# \mid \forall \tau \in D^{\leq t}_\#\colon h(\sigma) = h(\sigma\tau)\right\},$$

$$h'(D, t) := \begin{cases} h\left(\min(M_{D,t})\right), & M_{D,t} \neq \emptyset, \\ h(\varepsilon), & \text{else.} \end{cases}$$

To show that $L \in \mathbf{TxtPsd}\delta\mathbf{Ex}(h')$, we first show that $L \in \mathbf{TxtPsdEx}(h')$. To that end, let $T \in \mathbf{Txt}(L)$. By Blum and Blum (1975) there exists a locking sequence σ for h on L. Let σ_0 be a minimal such locking sequence. Now, let n_0 be large enough such that, using $D_0 \coloneqq \mathrm{content}(T[n_0])$ for notational convenience,

- $\mathrm{content}(\sigma_0) \subseteq D_0$,
- $|\sigma_0| \leq n_0$ and
- for all $\sigma' < \sigma_0$ there exists $\tau' \in (D_0)^{\leq n_0}_{\#}$ witnessing $\sigma' \notin M_{D_0,n_0}$.

Then, $\min(M_{D_0,n_0}) = \sigma_0$. Thus, for $n \geq n_0$ we have $h'(\mathrm{content}(T[n]), n) = h(\sigma_0)$, and $W_{h'(\mathrm{content}(T[n]),n)} = W_{h(\sigma_0)} = L$. Thus, $L \in \mathbf{TxtPsdEx}(h')$.

It remains to show that h' retains the restriction δ. As $\mathrm{content}(T) = L$, it suffices to show that for every $\sigma' \in L^*_{\#}$ we have $P(h'(\sigma'), L)$. By definition, there exists $\sigma \in L^*_{\#}$ such that $h'(\sigma') = h(\sigma)$. As $P(h(\sigma), L)$, we also have $P(h'(\sigma'), L)$, concluding the proof. ■

Lemma 4 *We have* $[\mathbf{TxtPsdCautEx}] \setminus [\mathbf{TxtSdEx}] \neq \emptyset$.

Proof The same proof as for a more restrictive case works for this setting as well, see Kötzing and Schirneck (2016). We include it here for completeness. We interpret natural numbers as coded triples of natural numbers. Let π_i denote the projection of such triples onto their i-th coordinate. Furthermore, let Φ denote a fixed Blum complexity measure, see Blum (1967). In particular, there is an algorithm which, given a program p, an input value x and a time t, decides whether $\Phi_p(x) > t$ holds. Let p_0 be an index of the empty set and p_1 be an index of the set $\mathbb{N}$ of all natural numbers. By the S-m-n Theorem, there is a total computable function $\mathrm{join} \in \mathcal{R}$ such that, for all numbers e and all finite sets D, we have $W_{\mathrm{join}(e,D)} = W_e \cup D$. For any number t and any finite set D, we consider the following total learner

$$h(D,t) \coloneqq \begin{cases} p_0, & D = \emptyset, \\ p_1, & \text{else, if } \exists x, y \in D\ \exists i \in \{1,2\} \colon \pi_i(x) \neq \pi_i(y), \\ e, & \text{else, if } \exists p \colon ((\forall x \in D\ \exists i \colon x = \langle e,p,i\rangle) \wedge \Phi_p(0) > t), \\ \mathrm{join}(e,D), & \text{else.} \end{cases}$$

First, we argue why the learner h is cautious on arbitrary texts. As long as all presented data is of the form $\langle e,p,i\rangle$, for some fixed e and p and various i, and the length of the initial sequence shown so far does not extend $\Phi_p(0)$, the set W_e is proposed. Once value $\Phi_p(0)$ is reached by parameter t, if ever, h switches to the superset $W_e \cup D$. If multiple first or second coordinates occur, h conjectures $\mathbb{N}$ as its final guess. As the conjectured sets only become potentially larger, i.e. supersets, h is cautious. Let $\mathcal{L} = \mathbf{TxtPsdCautEx}(h)$ be the class of languages h infers.

Assume that $\mathcal{L} \subseteq \mathbf{TxtSdEx}(h')$ for some learner h'. As $\mathbb{N} \in \mathcal{L}$, the learner h' is total. Using the Operator Recursion Theorem, see Case (1974), there are indices e and p such that, with $\langle\!\langle e,p,j\rangle\!\rangle \coloneqq \{\langle e,p,i\rangle : i < j\}$ for any natural number j,

$$W_e = \left\{\langle e,p,i\rangle \mid \forall j \leq i \colon h'(\langle\!\langle e,p,j\rangle\!\rangle) \neq h'(\langle\!\langle e,p,j+1\rangle\!\rangle)\right\},$$

$$\varphi_p(0) = \begin{cases} 1, & \exists j \colon h'(\langle\!\langle e,p,j\rangle\!\rangle) = h'(\langle\!\langle e,p,j+1\rangle\!\rangle), \\ \uparrow, & \text{else.} \end{cases}$$

In order to get to a contradiction, we distinguish between the following two cases. First, assume the set W_e is infinite. In this case, $W_e = \{\langle e, p, i\rangle \mid i \in \mathbb{N}\}$ and $\Phi_p(0) > t$ even for arbitrarily large t. Thus, $W_e \in \mathcal{L}$. However, h' cannot learn W_e from the text $(\langle e, p, i\rangle)_{i\in\mathbb{N}}$ as it makes infinitely many mind changes. For the second case, assume the set W_e is finite. Then, there exists k such that $W_e = \langle\!\langle e, p, k\rangle\!\rangle$. As $\langle e, p, k\rangle$ is not in W_e, we have $h'(W_e) = h'(W_e \cup \{\langle e, p, k\rangle\})$. So, there are t large enough such that $\Phi_p(0) \leq t$. Let $L := W_e$ and $L' := W_e \cup \{\langle e, p, k\rangle\}$. Since the learner h converges correctly to the hypotheses $\mathrm{join}(e, L)$ and $\mathrm{join}(e, L')$ on arbitrary texts of L and L', respectively, we have $L, L' \in \mathcal{L}$. On the other hand, h' cannot distinguish between L and L', as $h'(L) = h'(L')$, a contradiction. ■

Theorem 5 *We have* $[\mathbf{TxtPsdCaut}_\infty\mathbf{Ex}] = [\mathbf{TxtPsdEx}]$. *Particularly, for* $\beta \in \{\mathbf{G}, \mathbf{Psd}\}$ *and* $\delta \in \{\mathbf{T}, \mathbf{Caut}_\infty\}$, *we have* $[\mathbf{Txt}\beta\delta\mathbf{Ex}] = [\mathbf{TxtGEx}]$.

Proof The inclusion $[\mathbf{TxtPsdCaut}_\infty\mathbf{Ex}] \subseteq [\mathbf{TxtPsdEx}]$ follows immediately. For the other inclusion, let h be a learner and $L \in \mathbf{TxtPsdEx}(h)$. By Case and Kötzing (2016), we may assume h to be total and strongly non-U-shaped[3]. We will define a learner h' to learn L in a $\mathbf{Caut}_\infty$ manner, i.e. $L \in \mathbf{TxtPsdCaut}_\infty\mathbf{Ex}(h')$. First, we need an auxiliary function. Using the S-m-n Theorem we obtain a total, computable function $p \in \mathcal{R}$ such that, for all finite $D \subseteq \mathbb{N}$ and $s, t \geq 0$,

$$H^s_{D,t} := \left\{(D', t') : (D, t) \preceq (D', t') \preceq (W^s_{h(D,t)}, t + s)\right\},$$
$$W_{p(D,t)} = D \cup \bigcup_{s\in\mathbb{N}} \begin{cases} W^s_{h(D,t)}, & D \subseteq W^s_{h(D,t)} \wedge \forall (D', t') \in H^s_{D,t} \colon h(D, t) = h(D', t'), \\ \emptyset, & \text{else.} \end{cases} \tag{5}$$

Informally, $p(D, t)$ enumerates $W_{h(D,t)}$ as long as (D, t) acts like a locking information. Formally, let (D_0, t_0) be a locking information for h on L and let $(D, t) \succeq (D_0, t_0)$, with $D \subseteq L$. We want to show that $W_{p(D,t)} = L$. By definition, $W_{p(D,t)} \subseteq D \cup W_{h(D,t)} = L$. For the other inclusion, let s be such that $D \subseteq W^s_{h(D,t)}$. Such s must exist as $D \subseteq_{\text{Fin}} L = W_{h(D,t)}$. As $D \cup W^s_{h(D,t)} \subseteq_{\text{Fin}} L$, for every $(D', t') \in H^s_{D,t}$ we have $h(D, t) = h(D', t')$. Thus, $W^s_{h(D,t)}$ gets enumerated into $W_{p(D,t)}$ and, in the end, we have $L = \bigcup_{s\in\mathbb{N}} W^s_{h(D,t)} \subseteq W_{p(D,t)}$. Altogether, we have $W_{p(D,t)} = L$.
We show another property of $p(D, t)$ which will be needed later. Namely,

$$\text{if } W_{p(D,t)} \text{ is infinite, then } W_{p(D,t)} = W_{h(D,t)}. \tag{6}$$

As $W_{p(D,t)}$ is infinite, and D is finite, additional elements must have been enumerated through the case distinction in the union in the Term (5). Thus, $D \subseteq W_{h(D,t)}$ must have been witnessed and then, by definition, $W_{p(D,t)} \subseteq W_{h(D,t)}$. For the other direction, assume there exists $x \in W_{h(D,t)} \setminus W_{p(D,t)}$, and let s_0 be minimal such that $x \in W^{s_0}_{h(D,t)}$. As $x \notin W_{p(D,t)}$, we have $W_{p(D,t)} \subseteq D \cup \bigcup_{s<s_0} W^s_{h(D,t)}$, which is finite, a contradiction.

Before we define h', we fix some notations to ease readability. For any function g, let $g^*(\sigma) := g(\mathrm{content}(\sigma), |\sigma|)$. Also, let $\sigma_{D,t}$ be the canonical sequence of the set D of length t, that is, the sequence of elements of D in ascending order, possibly continued by pause symbols to fit the length.

3. Formally, $\mathbf{SNU}(p, T) :\Leftrightarrow \forall i, j, k \colon (i \leq j \leq k \wedge W_{p(i)} = W_{p(k)} = \mathrm{content}(T)) \Rightarrow p(i) = p(j)$, see Case and Moelius (2011). Informally, in strongly non-U-shaped learning, once the target language is suggested correctly, no more syntactic mind changes are allowed.

Now, we can define h'. Intuitively, given the information (D,t), h' will search for the shortest initial part of the canonical sequence $\sigma_{D,t}[n_{D,t}]$ that looks like a locking sequence for h. Formally, for any finite $D \subseteq \mathbb{N}$, $t \geq 0$ and $0 \leq n \leq t$, we define h' as

$$\begin{aligned} I^n_{D,t} &\coloneqq \{(D',t') : (\text{content}(\sigma_{D,t}[n]), n) \preceq (D',t') \preceq (D,t)\},\\ n_{D,t} &\coloneqq \min\{n \mid 0 \leq n \leq t \wedge \forall (D',t') \in I^n_{D,t} \colon h(D',t') = h(D,t)\},\\ h'(D,t) &\coloneqq p^*(\sigma_{D,t}[n_{D,t}]). \end{aligned}$$

We start by showing that $L \in \mathbf{TxtPsdEx}(h')$. To that end, let $T \in \mathbf{Txt}(L)$ and T_c be the canonical text for L, that is, the text containing all elements of L in ascending order, possibly continued by infinitely many pause symbols if L is finite. Then, as h is strongly non-U-shaped, every initial sequence $T_c[n]$ with $W_{h^*(T_c[n])} = L$ is a locking information for h on L. Let n_0 be minimal such that $\sigma_0 \coloneqq T_c[n_0]$ is a locking information for h on L. Now, let $n_1 \geq n_0$ such that $\text{content}(T[n_1]) \supseteq \text{content}(\sigma_0)$. Then, for $n \geq n_1$, $h'(T[n], n) = p^*(\sigma_0)$ with $W_{p^*(\sigma_0)} = L$, showing that $L \in \mathbf{TxtPsdEx}(h')$.

It remains to show that h' is $\mathbf{Caut}_\infty$ on L. To that end, assume the opposite, namely that there exist $(D_1,t_1) \preceq (D_2,t_2)$, with $D_2 \subseteq L$, such that $W_{h'(D_1,t_1)} \supsetneq W_{h'(D_2,t_2)}$ and $W_{h'(D_2,t_2)}$ is infinite. For $i \in \{1,2\}$ let

$$\begin{aligned} \sigma_i &\coloneqq \sigma_{D_i,t_i}[n_{D_i,t_i}],\\ (D'_i, t'_i) &\coloneqq (\text{content}(\sigma_i), |\sigma_i|). \end{aligned}$$

Basically, σ_i is the sequence h' searches back to, i.e. $h'(D_i,t_i) = p^*(\sigma_i)$. This changes the assumption to $W_{p^*(\sigma_1)} \supsetneq W_{p^*(\sigma_2)}$ and $W_{p^*(\sigma_2)}$ is infinite. Additionally, we have that

- $W_{p^*(\sigma_1)}$ is infinite, as $W_{p^*(\sigma_2)}$ is, and
- $W_{p^*(\sigma_1)} \supseteq D'_1$ and $W_{p^*(\sigma_2)} \supseteq D'_2$, due to the definition of p. In particular, we have that $W_{p^*(\sigma_1)} \supseteq D'_1 \cup D'_2$.

For $t^* \coloneqq \max\{t'_1, t'_2, |D'_1 \cup D'_2|\}$ the following hold.

$(*)$: By the definition of $p^*(\sigma_1)$, we have $h^*(\sigma_1) = h(D,t)$ for every (D,t) such that $D'_1 \subseteq D \subseteq W_{p^*(\sigma_1)}$ and $t'_1 \leq t$. In particular, this holds true for $(D,t) = (D'_1 \cup D'_2, t^*)$.

$(**)$: As, by definition of σ_2, $h'(D_2,t_2) = p^*(\sigma_2)$, we have for each $(D'',t'') \in I^{n_{D_2,t_2}}_{D_2,t_2}$ that $h^*(\sigma_2) = h(D'',t'')$. In particular, this holds true for $(D'',t'') = (D'_1 \cup D'_2, t^*)$.

Thus, we have

$$h^*(\sigma_1) \overset{(*)}{=} h(D'_1 \cup D'_2, t^*) \overset{(**)}{=} h^*(\sigma_2). \tag{7}$$

Now, we have the contradiction

$$W_{p^*(\sigma_2)} \subsetneq W_{p^*(\sigma_1)} \overset{(6)}{=} W_{h^*(\sigma_1)} \overset{(7)}{=} W_{h^*(\sigma_2)} \overset{(6)}{=} W_{p^*(\sigma_2)}.$$

Altogether, we get $L \in \mathbf{TxtPsdCaut}_\infty\mathbf{Ex}(h')$ and thus the desired. ■

Appendix C. Omitted Proofs of Section 3

C.1. Omitted Proofs of Section 3.1

Lemma 7 *We have* $[\mathbf{TxtSdCaut_{Tar}Bc}] = [\mathbf{TxtSdBc}]$.

Proof The inclusion $[\mathbf{TxtSdCaut_{Tar}Bc}] \subseteq [\mathbf{TxtSdBc}]$ follows immediately. For the other inclusion, let h be a $\mathbf{Sd}$-learner. We will show that $\mathbf{TxtSdBc}(h) \subseteq \mathbf{TxtSdCaut_{Tar}Bc}(h_w)$ for h_w from Algorithm 1. To that end, let $L \in \mathbf{TxtSdBc}(h)$ and $T \in \mathbf{Txt}(L)$.

First, we show that $L \in \mathbf{TxtSdBc}(h_w)$. As h is strongly $\mathbf{Bc}$-locking, see Kötzing et al. (2017), there exists n_0 such that $D_0 := \mathrm{content}(T[n_0])$ is a $\mathbf{Bc}$-locking set for h on L. We show that for every $n \geq n_0$ and $D := \mathrm{content}(T[n])$ we have $W_{h_w(D)} = L$. Since only $x \in W_{h(D)} = L$ are considered for the enumeration, see line 2, we get $W_{h_w(D)} \subseteq L$. For the other direction, we show that the algorithm runs through every step i successfully. Let $E_0 = D$, and let i be the next step in Algorithm 1. If $x_i \in E_i$, then step i is completed and x_i is enumerated into $E_{i+1} \subseteq W_{h_w(D)}$. In the other case, we have $x_i \notin E_i$. Since $E_i \cup \{x_i\}$ is a finite subset of $W_{h(D)} = L$, for every D', with $(D_0 \subseteq)D \subseteq D' \subseteq E_i \cup \{x_i\}(\subseteq L)$, $W_{h(D')} = L$ will witness $E_i \cup \{x_i\}$, i.e. there exists some t such that $E_i \cup \{x_i\} \subseteq W^t_{h(D')}$. Thus, x_i will be enumerated into $E_{i+1} \subseteq W_{h_w(D)}$, and step i is completed in this case as well. So, every $x \in W_{h(D)} = L$ will also be enumerated into $W_{h_w(D)}$, and we get $W_{h_w(D)} \supseteq L$. Altogether, we have $W_{h_w(D)} = L$, concluding the first part of the proof.

To prove that h_w learns L respecting $\mathbf{Caut_{Tar}}$, assume the opposite, namely the existence of $D' \subseteq L$ such that $L \subsetneq W_{h_w(D')}$. Let $x \in W_{h_w(D')} \setminus L$ be a witness and let D_0 be a $\mathbf{Bc}$-locking set for h on L such that $D' \subseteq D_0 \subseteq L$. Let i be the step[4] where $D_0 \cup \{x\}$ is enumerated into $W_{h_w(D')}$, i.e. $D_0 \cup \{x\} \not\subseteq E_i$ and $D_0 \cup \{x\} \subseteq E_{i+1}$. Then, by lines 4 and 5, for $D'' = D_0$, we have $x \in E_{i+1} \subseteq W_{h(D'')}$, a contradiction. ∎

Proposition 8 *Let* $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$. *Given a learner* h *and with it the learner* h_s *as built in Algorithm 2, the following properties hold.*

(i) *If* h *is a* β*-learner, then* h_s *is a* β*-learner which is consistent on arbitrary input.*

(ii) *If* σ_0 *is a* $\mathbf{Bc}$*-locking information for* h *on some* $L \subseteq \mathbb{N}$*, then* σ_0 *is a* $\mathbf{Bc}$*-locking information for* h_s *on* L*.*

(iii) *For*[5] $\beta \neq \mathbf{G}$ *target cautious learning is preserved by the learner* h_s*, that is, we have that* $\mathbf{Txt}\beta\mathbf{Caut_{Tar}Bc}(h) \subseteq \mathbf{Txt}\beta\mathbf{Caut_{Tar}Bc}(h_s)$.

(iv) *If* $W_{h_s(\sigma)}$ *is infinite, then* $W_{h_s(\sigma)} = W_{h(\sigma)} =: L$ *and* σ *is a* $\mathbf{Bc}$*-locking information for* h *and* h_s *on* L*.*

(v) *If* $L \in \mathbf{Txt}\beta\mathbf{Caut_{Tar}Bc}(h)$ *and* σ_0 *is a* $\mathbf{Bc}$*-locking information for* h_s *on* L*, then* σ_0 *is a* $\mathbf{Bc}$*-locking information for* h *on* L*.*

(vi) *Let* h*, and thus* h_s*, be* $\mathbf{Sd}$*-learners. Let* D_0 *be a* $\mathbf{Bc}$*-locking set for* h *on some* L*. Then, for* D *with either (a)* $D \subseteq D_0$ *or (b)* $D_0 \subseteq D \subseteq L$*, we have*

$$D_0 \subseteq W_{h_s(D)} \Rightarrow W_{h_s(D)} \subseteq L.$$

4. Note that x and x_i may differ.
5. As it will turn out, the same holds true for $\beta = \mathbf{G}$, see Corollary 11.

Proof

(i) Let h be a β-learner. By definition, h_s is consistent on arbitrary input. As all inquiries to sequences occur within h, namely $h(\sigma)$ in line 2, $h(\sigma\tau'')$ in line 5 and $h(\sigma\tau'')$ and $h(\sigma\tau')$ in line 8, h_s requires the same form of information. Thus, h_s is a β-learner which is consistent on arbitrary input.

(ii) Let σ_0 be a $\mathbf{Bc}$-locking information for h on some $L \subseteq \mathbb{N}$ and let $\sigma \in L^*_\#$ such that $\sigma_0 \subseteq \sigma$. We want to show that $W_{h_s(\sigma)} = L$. By definition, $W_{h_s(\sigma)} \subseteq W_{h(\sigma)} = L$. Now, let i be the current step in the algorithm and let $x_i = \mathrm{enum}(h(\sigma), i)$. Either $x_i \in E_i$, then this step is completed and x_i will be enumerated into E_{i+1}. Otherwise, for every $\tau'' \in (E_i \cup \{x_i\})^{\leq i}_\#$ we can find $s_{\tau''}$ such that $E_i \cup \{x_i\} \subseteq W^{s_{\tau''}}_{h(\sigma\tau'')}$, as $E_i \cup \{x_i\} \subseteq_{\mathrm{Fin}} L = W_{h(\sigma\tau'')}$. Then, again, for every $\tau' \in (E_i \cup \{x_i\})^{\leq i}_\#$ we can find t such that

$$\bigcup_{\tau'' \in D^{\leq i}_\#} W^{s_{\tau''}}_{h(\sigma\tau'')} \subseteq W^t_{h(\sigma\tau')},$$

as the big union is a finite subset of $L = W_{h(\sigma\tau')}$. Thus, x_i will be enumerated into E_{i+1}. As every $x \in W_{h(\sigma)} = L$ will be enumerated into $W_{h_s(\sigma)}$, we also get $L = W_{h(\sigma)} \subseteq W_{h_s(\sigma)}$, concluding the proof.

(iii) For $\beta \neq \mathbf{G}$, let $L \in \mathbf{Txt}\beta\mathbf{Caut_{Tar}Bc}(h)$. First, we show that h_s from Algorithm 2 preserves $\mathbf{Txt}\beta\mathbf{Bc}$-learning, i.e. $L \in \mathbf{Txt}\beta\mathbf{Bc}(h_s)$. To do so, let $T \in \mathbf{Txt}(L)$. As h is strongly $\mathbf{Bc}$-locking, see Kötzing et al. (2017), there exists n_0 such that $T[n_0]$ is a $\mathbf{Bc}$-locking information for h on L. Then, by Proposition 8 (ii), $T[n_0]$ is also a $\mathbf{Bc}$-locking information for h_s. Thus, $\mathbf{Txt}\beta\mathbf{Bc}(h) \subseteq \mathbf{Txt}\beta\mathbf{Bc}(h_s)$.

To show that h_s also preserves $\mathbf{Caut_{Tar}}$ while learning L, assume the opposite, i.e. there exists $\sigma \in L^*_\#$ such that $L \subsetneq W_{h_s(\sigma)}$. Then, by definition, $L \subsetneq W_{h_s(\sigma)} \subseteq W_{h(\sigma)}$, contradicting the target cautiousness of h.

(iv) Let $W_{h_s(\sigma)}$ be infinite. First, we show that $W_{h_s(\sigma)} = W_{h(\sigma)}$. By definition, $W_{h_s(\sigma)} \subseteq W_{h(\sigma)}$. Now, assume there exists $x \in W_{h(\sigma)} \setminus W_{h_s(\sigma)}$, and also assume that x is the first such with respect to $\mathrm{enum}(h(\sigma), .)$. As $x \notin W_{h_s(\sigma)}$, the enumeration must be stuck either at finding a minimal s in the lines 4 and 5 or in the check in the lines 7 and 8, and thus $W_{h_s(\sigma)}$ must be finite, a contradiction.

For the second property, we first show that σ is a $\mathbf{Bc}$-locking information for h on $L := W_{h_s(\sigma)}$. Assume the existence of some $\tilde{\tau} \in L^*_\#$ such that $W_{h(\sigma\tilde{\tau})} \neq L$. We distinguish between the following two cases.

1.Case: $\exists x \in W_{h(\sigma\tilde{\tau})} \setminus L$: Let t_0 be such that $x \in W^{t_0}_{h(\sigma\tilde{\tau})}$. Let i_0 be the step such that $|E_{i_0}| > |W^{t_0}_{h(\sigma\tilde{\tau})}|$, $E_{i_0} \supseteq \mathrm{content}(\sigma\tilde{\tau})$ and $\tilde{\tau} \in (E_{i_0+1})^{\leq i_0}_\#$. Such i_0 exists as $|E_i| \overset{i\to\infty}{\to} \infty$ and $L = W_{h_s(\sigma)} \supseteq \mathrm{content}(\sigma\tilde{\tau})$. As the check in the lines 7 and 8 must be successful, we have for $\tau' = \varepsilon \in (E_{i_0+1})^{\leq i_0}_\#$ that

$$(x \in) \bigcup_{\tau'' \in (E_{i_0+1})^{\leq i_0}_\#} W^{s_{\tau''}}_{h(\sigma\tau'')} \subseteq W_{h(\sigma\tau')}.$$

The element x is in the union, as $|E_{i_0}| > |W^{t_0}_{h(\sigma\tilde{\tau})}|$ implies $s_{\tilde{\tau}} > t_0$, and, thus, we have $x \in W^{s_{\tilde{\tau}}}_{h(\sigma\tilde{\tau})}$. Altogether, we get $x \in W_{h(\sigma)} = L$, a contradiction.

2.Case: $\exists x \in L \setminus W_{h(\sigma\tilde{\tau})}$: Let i_0 be the step[6] such that $\tilde{\tau} \in (E_{i_0+1})^{\leq i_0}_{\#}$ and $\mathrm{content}(\sigma\tilde{\tau}) \cup \{x\} \subseteq E_{i_0+1}$. Then, by the lines 7 and 8 in the **SFV**, for $\tau' = \tilde{\tau} \in (E_{i_0+1})^{\leq i_0}_{\#}$ we have

$$(x \in E_{i_0+1} \subseteq) \bigcup_{\tau'' \in (E_{i_0+1})^{\leq i_0}_{\#}} W^{s_{\tau''}}_{h(\sigma\tau'')} \subseteq W_{h(\sigma\tau')}.$$

This yields $x \in W_{h(\sigma\tilde{\tau})}$, a contradiction.

Altogether, we get that σ is a $\mathbf{Bc}$-locking information for h on $W_{h_s(\sigma)} = W_{h(\sigma)}$. By Proposition 8 (ii), it also is for h_s.

(v) Let $L \in \mathbf{Txt}\beta\mathbf{Caut_{Tar}Bc}(h)$ and let σ_0 be a $\mathbf{Bc}$-locking information for h_s on L. Assume that σ_0 is no $\mathbf{Bc}$-locking information for h on L, i.e. there exists some $\tau' \in L^*_{\#}$ such that $W_{h(\sigma\tau')} \neq L$. As $L = W_{h_s(\sigma\tau')} \subseteq W_{h(\sigma\tau')}$, we get $L \subsetneq W_{h(\sigma\tau')}$, a contradiction to h being $\mathbf{Caut_{Tar}}$.

(vi) Let D_0 be a $\mathbf{Bc}$-locking set for h on L. For D, with (b) $D_0 \subseteq D \subseteq L$, we have $W_{h_s(D)} \subseteq W_{h(D)} = L$ by definition. For D, with (a) $D \subseteq D_0$, assume the existence of some $x \in W_{h_s(D)} \setminus L$. Let i_0 be the step[7] of Algorithm 2 such that $D_0 \cup \{x\} \not\subseteq E_{i_0}$ and $D_0 \cup \{x\} \subseteq E_{i_0+1}$. Then, by the lines 7 and 8, for $D' = D_0$, we have $x \in \bigcup_{D \subseteq D'' \subseteq E_{i_0+1}} W^{s_{D''}}_{h(D'')} \subseteq W_{h(D')} = L$, a contradiction. ■

C.2. Omitted Proofs in Section 3.2

Theorem 10 *Let P be a predicate on languages. Let δ be a learning restriction such that*

$$\delta(p, T) \Leftrightarrow \big(P(\mathrm{content}(T)) \Rightarrow \mathbf{Caut_{Tar}}(p, T)\big).$$

Then,

1. *δ allows for consistent $\mathbf{Bc}$-learning, that is, for any interaction operator $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$ we have $[\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}]$, and*
2. *$[\mathbf{TxtPsd}\delta\mathbf{Bc}] = [\mathbf{TxtG}\delta\mathbf{Bc}]$.*

Proof

1. We show that δ allows for consistent $\mathbf{Bc}$-learning. We follow the proof of Kötzing et al. (2017). For a total learner h let $L \in \mathbf{Txt}\beta\delta\mathbf{Bc}(h)$. Omitting the interaction operators for clarity, we define h' on finite sequences σ as

$$W_{h'(\sigma)} = \mathrm{content}(\sigma) \cup \bigcup_{s \in \mathbb{N}} \begin{cases} W^s_{h(\sigma)}, & \mathrm{content}(\sigma) \subseteq W^s_{h(\sigma)}, \\ \emptyset, & \text{else.} \end{cases}$$

6. Note that x and x_{i_0} may differ.
7. Note that x and x_{i_0} may differ.

Obviously, learner h' is consistent on arbitrary input, and if $W_{h(\sigma)} = L$, then $W_{h'(\sigma)} = W_{h(\sigma)}$, preserving $\mathbf{Bc}$-learning. To show that h' obeys the restriction δ, assume the opposite, i.e. there exists $\sigma \in L^*_\#$ such that $P(L)$ and $L \subsetneq W_{h'(\sigma)}$. Since this cannot be the case if $W_{h'(\sigma)} = \mathrm{content}(\sigma)$, there must have been some additional enumerations, i.e. $\mathrm{content}(\sigma) \subseteq W_{h(\sigma)}$ must have been witnessed at some point. Thus, $W_{h'(\sigma)} = W_{h(\sigma)}$, and now $L \subsetneq W_{h'(\sigma)} = W_{h(\sigma)}$, a contradiction.

2. To show that $[\mathbf{TxtPsd}\delta\mathbf{Bc}] = [\mathbf{TxtG}\delta\mathbf{Bc}]$, observe that the inclusion $[\mathbf{TxtPsd}\delta\mathbf{Bc}] \subseteq [\mathbf{TxtG}\delta\mathbf{Bc}]$ follows immediately. For the other, we follow the idea how $\mathbf{TxtGBc}$-learning can be done partially set-driven, discussed in private communication with Jain (2017). We expand that idea so that the restriction δ is also preserved. To that end, let $L \in \mathbf{TxtG}\delta\mathbf{Bc}(h)$ for a learner h. Now, define the $\mathbf{Psd}$-learner h' as follows. With the S-m-n Theorem, we get a total computable function p such that, for finite $D \subseteq \mathbb{N}$ and $t \geq 0$,

$$A_{D,t} \coloneqq W_{p(D,t)} = \bigcup_{\sigma \in D^*_\#} \left(\bigcap_{\tau \in D^{\leq t}_\#} W_{h(\sigma\tau)} \cap \bigcap_{\sigma' < \sigma, \sigma' \in D^*_\#} \bigcup_{\tau' \in D^*_\#} W_{h(\sigma'\tau')} \right), \tag{8}$$

$$W_{h'(D,t)} = \bigcup_{s \in \mathbb{N}} \begin{cases} A^s_{D,t}, & \exists \rho \in D^{\leq t}_\# \colon A^s_{D,t} \subseteq W_{h(\rho)}, \\ \emptyset, & \text{else.} \end{cases}$$

Intuitively, $A_{D,t}$ checks whether the information given is enough to witness a (minimal) $\mathbf{Bc}$-locking sequence. Then, at every step of the enumeration of $W_{h'(D,t)}$, there is a check whether there is a possible hypothesis of h which would enumerate the same. This will ensure to maintain the restriction δ.

We start by proving $L \in \mathbf{TxtPsdBc}(h')$. For that, let $T \in \mathbf{Txt}(L)$. By Blum and Blum (1975), there exists a $\mathbf{Bc}$-locking sequence for h on L. Let α be the least such $\mathbf{Bc}$-locking sequence with respect to $<$. By Osherson et al. (1986), for each $\alpha' < \alpha$ such that $\mathrm{content}(\alpha') \subseteq L$, there exists $\tau_{\alpha'}$ such that $\alpha'\tau_{\alpha'}$ is a $\mathbf{Bc}$-locking sequence for h on L. Now, let n_0 be large enough such that

- $n_0 \geq |\alpha|$,
- $\mathrm{content}(\alpha) \subseteq \mathrm{content}(T[n_0])$ and
- for all $\alpha' < \alpha$ such that $\mathrm{content}(\alpha') \subseteq L$, we have $\mathrm{content}(\alpha'\tau_{\alpha'}) \subseteq \mathrm{content}(T[n_0])$ and $|\tau_{\alpha'}| \leq n_0$.

We claim that for $t \geq n_0$ and $D = \mathrm{content}(T[t])$, we have $W_{h'(D,t)} = L$. In order to do so, we first have to show $A_{D,t} = L$.

$\subseteq$: To show $A_{D,t} \subseteq L$, let $x \in A_{D,t}$ and let σ be the witness of enumerating x into $A_{D,t}$. We will distinguish between the following two cases.

$\sigma \leq \alpha$: As x must be an element of the left hand intersection of (8), and as $\tau_\sigma \in D^{\leq t}_\#$ for $\sigma \leq \alpha$, we get $x \in W_{h(\sigma\tau_\sigma)} = L$.

$\sigma > \alpha$: Here, we exploit that x must be an element of the right hand intersection of (8). As $\alpha < \sigma$ and $\alpha \in D^{\leq t}_\#$, we have $x \in W_{h(\alpha\tau)} = L$ for any τ.

In both cases we have $x \in L$, thus $A_{D,t} \subseteq L$.

$\supseteq$: Next, we show $L \subseteq A_{D,t}$. Let $x \in L$. As D and t are chosen sufficiently large, α is a candidate for the enumeration of $A_{D,t}$. Since α is a $\mathbf{Bc}$-locking sequence, we will witness $x \in W_{h(\alpha\tau)} = L$ for every $\tau \in D_{\#}^{\leq t}$. Thus, the left hand intersection of (8) will contain x.
For the right hand intersection of (8), observe that for every $\sigma' < \alpha$, with $\mathrm{content}(\sigma') \subseteq D$, we have $\tau_{\sigma'} \in D_{\#}^{*}$. So, the intersection will contain at least $W_{\sigma'\tau_{\sigma'}} = L$, of which x is an element. Thus, we have $L \subseteq A_{D,t}$.

Now that we have shown $A_{D,t} = L$, it remains to show that $W_{h'(D,t)} = A_{D,t} = L$. By definition, we have $W_{h'(D,t)} \subseteq A_{D,t}$. For the other direction, let s be the next step in the enumeration of $W_{h'(D,t)}$. We want to check whether we can enumerate $A_{D,t}^{s}$. As $A_{D,t}^{s} \subseteq L = W_{h(\alpha)}$ with $\alpha \in D_{\#}^{\leq t}$, we have a witness that we can enumerate $A_{D,t}^{s}$. Thus, for all s we have $A_{D,t}^{s} \subseteq W_{h'(D,t)}$ and so we get $W_{h'(D,t)} = A_{D,t}$. In the end, $L \in \mathbf{TxtPsdBc}(h')$.

Finally, to see $L \in \mathbf{TxtPsd}\delta\mathbf{Bc}(h')$, assume there exists some (D,t) such that $P(L)$ and $L \subsetneq W_{h'(D,t)}$. By definition of $W_{h'(D,t)}$, there exists some $\rho \in D_{\#}^{\leq t}$ such that $W_{h'(D,t)} \subseteq W_{h(\rho)}$. Thus, we have $P(L)$ and $L \subsetneq W_{h'(D,t)} \subseteq W_{h(\rho)}$, a contradiction to h learning L according to δ. ■

Theorem 12 *We have*

$$[\mathbf{TxtSdCaut}_{\mathbf{Fin}}\mathbf{Bc}] = [\mathbf{TxtPsdCaut}_{\mathbf{Fin}}\mathbf{Bc}] = [\mathbf{TxtGCaut}_{\mathbf{Fin}}\mathbf{Bc}].$$

Proof To prove the theorem, we apply the same idea as Kötzing and Palenta (2016) when dealing with $\mathbf{Caut}$, that is, we introduce a weaker version of $\mathbf{Caut}_{\mathbf{Fin}}$, namely

$$(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}(p,T) :\Leftrightarrow \left(\mathrm{content}(T) < \infty \Rightarrow \forall i\colon \neg(\mathrm{content}(T) \subsetneq W_{p(i)})\right).$$

Intuitively, $(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}$ has to be $\mathbf{Caut}_{\mathbf{Tar}}$ only on finite target languages. It follows immediately that $\mathbf{Caut}_{\mathbf{Tar}}$ as well as $\mathbf{Caut}_{\mathbf{Fin}} \cap \mathbf{Bc}$ imply $(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}$.

By Theorem 10, we already have $[\mathbf{TxtPsd}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}] = [\mathbf{TxtG}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}]$. To show $[\mathbf{TxtSd}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}] = [\mathbf{TxtPsd}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}]$, let h be a learner and let $L \in \mathbf{TxtPsd}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}(h)$. We first observe that, by Theorem 10, we may assume h to be consistent. Now, we follow the idea from Kötzing et al. (2017) and introduce $h'(D) \coloneqq h(D, |D|)$. First, we show that h' learns L. If L is infinite, then we get $L \in \mathbf{TxtSdBc}(h)$ by Kötzing et al. (2017). For finite L, let $T \in \mathbf{Txt}(L)$ and n_0 be such that $\mathrm{content}(T[n_0]) = L$. Now, for $n \geq n_0$ and $D \coloneqq \mathrm{content}(T[n]) = L$, we will show $L = W_{h'(D)}$. Firstly, we have $L \subseteq W_{h(D,|D|)} = W_{h'(D)}$ by consistency of h. By $(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}$, we also have $\neg(L \subsetneq W_{h(D,|D|)} = W_{h'(D)})$, and thus $L = W_{h'(D)}$.
To show that h' follows the restriction $(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}$, assume the opposite, i.e. there exist a finite target language L and $D \subseteq L$ such that $L \subsetneq W_{h'(D)}$. As $h'(D) = h(D, |D|)$, we get $L \subsetneq W_{h'(D)} = W_{h(D,|D|)}$, a contradiction.

Now, the following inclusion chain closes the proof.

$$[\mathbf{TxtSdCautBc}] \subseteq [\mathbf{TxtSdCaut_{Fin}Bc}] \subseteq [\mathbf{TxtPsdCaut_{Fin}Bc}] \subseteq [\mathbf{TxtGCaut_{Fin}Bc}] \subseteq$$
$$\subseteq [\mathbf{TxtG(Caut_{Tar})_{Fin}Bc}] = [\mathbf{TxtSd(Caut_{Tar})_{Fin}Bc}] =$$
$$= [\mathbf{TxtSdCautBc}].$$ ■

Lemma 13 *We have* $[\tau(\mathbf{Cons})\mathbf{TxtPsdCaut}_\infty\mathbf{Bc}] = [\mathbf{TxtPsdBc}]$.

Proof By definition, we get $[\tau(\mathbf{Cons})\mathbf{TxtPsdCaut}_\infty\mathbf{Bc}] \subseteq [\mathbf{TxtPsdBc}]$. For the other direction, let $L \in \mathbf{TxtPsdBc}(h)$ for some learner h. For the $\mathbf{Psd}$-learner h_s from Algorithm 2, we will show that $L \in \tau(\mathbf{Cons})\mathbf{TxtPsdCaut}_\infty\mathbf{Bc}(h_s)$. By Proposition 8 (i), h_s is consistent on arbitrary input. As in the proof of Proposition 8 (iii), we get $L \in \mathbf{TxtPsdBc}(h_s)$. To show that h_s is $\mathbf{Caut}_\infty$, assume the opposite, i.e. there exists $(D,t) \preceq (D',t')$ with $D' \subseteq L$ such that $W_{h_s(D,t)} \supsetneq W_{h_s(D',t')}$ and $W_{h_s(D',t')}$ is infinite. Then, $W_{h_s(D,t)}$ is infinite, too. By Proposition 8 (iv), (D',t') must be a $\mathbf{Bc}$-locking information both for $W_{h_s(D',t')}$ and, as $(D,t) \preceq (D',t')$ and (D,t) is a $\mathbf{Bc}$-locking information for $W_{h_s(D,t)}$, for $W_{h_s(D,t)}$ as well, which are not equal, yielding a contradiction. ■