Mapping Restrictions in Behaviourally Correct Learning

Vanja Doskoč

Universitätsdissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

in der Wissenschaftsdisziplin
Informatik

eingereicht an der
Digital-Engineering-Fakultät
der Universität Potsdam

Datum der Disputation: 21. April 2023

Betreuer

Prof. Dr. Tobias Friedrich
Hasso Plattner Institute, University of Potsdam, Germany

Gutachter

Prof. Dr. Frank Stephan
National University of Singapore, Singapore

Prof. Dr. Luca San Mauro
Sapienza University of Rome, Italy

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-59311
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-593110

Abstract

In this thesis, we investigate language learning in the formalisation of Gold [Gol67]. Here, a learner, being successively presented all information of a *target language*, conjectures which language it believes to be shown. Once these hypotheses converge *syntactically* to a correct explanation of the target language, the learning is considered successful. Fittingly, this is termed *explanatory* learning. To model learning strategies, we impose *restrictions* on the hypotheses made, for example requiring the conjectures to follow a monotonic behaviour. This way, we can study the impact a certain restriction has on learning.

Recently, the literature shifted towards *map charting*. Here, various seemingly unrelated restrictions are contrasted, unveiling interesting relations between them. The results are then depicted in *maps*. For explanatory learning, the literature already provides maps of common restrictions for various forms of data presentation.

In the case of *behaviourally correct* learning, where the learners are required to converge *semantically* instead of syntactically, the same restrictions as in explanatory learning have been investigated. However, a similarly complete picture regarding their interaction has not been presented yet.

In this thesis, we transfer the map charting approach to behaviourally correct learning. In particular, we complete the partial results from the literature for many well-studied restrictions and provide *full* maps for behaviourally correct learning with different types of data presentation. We also study properties of learners assessed important in the literature. We are interested whether learners are *consistent*, that is, whether their conjectures include the data they are built on. While learners cannot be assumed consistent in explanatory learning, the opposite is the case in behaviourally correct learning. Even further, it is known that learners following different restrictions may be assumed consistent. We contribute to the literature by showing that this is the case for *all* studied restrictions.

We also investigate mathematically interesting properties of learners. In particular, we are interested in whether learning under a given restriction may be done with *strongly* **Bc**-*locking learners*. Such learners are of particular value as they allow to apply *simulation arguments* when, for example, comparing two learning paradigms to each other. The literature gives a rich ground on when learners may be assumed strongly **Bc**-locking, which we *complete* for all studied restrictions.

Zusammenfassung

In dieser Arbeit untersuchen wir das Sprachenlernen in der Formalisierung von Gold [Gol67]. Dabei stellt ein Lerner, dem nacheinander die volle Information einer *Zielsprache* präsentiert wird, Vermutungen darüber auf, welche Sprache er glaubt, präsentiert zu bekommen. Sobald diese Hypothesen *syntaktisch* zu einer korrekten Erklärung der Zielsprache konvergieren, wird das Lernen als erfolgreich angesehen. Dies wird passenderweise als *erklärendes* Lernen bezeichnet. Um Lernstrategien zu modellieren, werden den aufgestellten Hypothesen *Einschränkungen* auferlegt, zum Beispiel, dass die Vermutungen einem monotonen Verhalten folgen müssen. Auf diese Weise können wir untersuchen, welche Auswirkungen eine bestimmte Einschränkung auf das Lernen hat.

In letzter Zeit hat sich die Literatur in Richtung *Kartographie* verlagert. Hier werden verschiedene, scheinbar nicht zusammenhängende Restriktionen einander gegenübergestellt, wodurch interessante Beziehungen zwischen ihnen aufgedeckt werden. Die Ergebnisse werden dann in so genannten *Karten* dargestellt. Für das erklärende Lernen gibt es in der Literatur bereits Karten geläufiger Einschränkungen für verschiedene Formen der Datenpräsentation.

Im Falle des verhaltenskorrekten Lernens, bei dem die Lerner nicht syntaktisch, sondern *semantisch* konvergieren sollen, wurden die gleichen Einschränkungen wie beim erklärenden Lernen untersucht. Ein ähnlich vollständiges Bild hinsichtlich ihrer Interaktion wurde jedoch noch nicht präsentiert.

In dieser Arbeit übertragen wir den Kartographie-Ansatz auf das verhaltenskorrekte Lernen. Insbesondere vervollständigen wir die Teilergebnisse aus der Literatur für viele gut untersuchte Restriktionen und liefern Karten für verhaltenskorrektes Lernen mit verschiedenen Arten der Datenpräsentation. Wir untersuchen auch Eigenschaften von Lernern, die in der Literatur als wichtig eingestuft werden. Uns interessiert, ob die Lerner *konsistent* sind, das heißt ob ihre Vermutungen die Daten einschließen, auf denen sie aufgebaut sind. Während man beim erklärenden Lernen nicht davon ausgehen kann, dass die Lerner konsistent sind, ist beim verhaltenskorrekten Lernen das Gegenteil der Fall. Es ist sogar bekannt, dass Lerner, die verschiedenen Einschränkungen folgen, als konsistent angenommen werden können. Wir tragen zur Literatur bei, indem wir zeigen, dass dies für *alle* untersuchten Restriktionen der Fall ist.

Wir untersuchen auch mathematisch interessante Eigenschaften von Lernern.

Insbesondere interessiert uns, ob das Lernen unter einer gegebenen Restriktion mit *stark* **Bc**-*sperrenden Lernern* durchgeführt werden kann. Solche Lerner sind von besonderem Wert, da sie es erlauben, *Simulationsargumente* anzuwenden, wenn man zum Beispiel zwei Lernparadigmen miteinander vergleicht. Die Literatur bietet eine reichhaltige Grundlage dafür, wann Lerner als stark **Bc**-sperrend angenommen werden können, die wir auf alle untersuchten Einschränkungen *erweitern*.

Acknowledgments

That's the game we play.

A few years ago, I set out to Potsdam in order to obtain my doctors degree. Admittedly, I had the math, fame and travelling in mind, but I obtained something way more valuable, *memories*. I first would like to thank the amazing Algorithm Engineering group (the Æ-team), who were particularly involved in these.

There is Tobias Friedrich, the head of the Æ-team, who proved to be a great leader and entertainer over and over again. He always protected us from the harsh outside world, a favour which we repaid in papers. Then, there is Timo Kötzing, my advisor, who always had an open ear for my questions regarding the researcher's life and an open eye for my texts and proofs. While he contributed a lot to my growth as a researcher, I also enjoyed all the teaching activities we did together. I gained a lot of new experiences, for example from contributing to a "flipped classroom". The teaching, researching and fun conversations I had with Sarel Cohen will also remain special to me. It was quite chaotic most of the time, but, then again, I liked the chaos.

A lot of memories involve the A-1.13 crew, an intensely fun flock consisting of Merlin de la Haye, Maximilian Katzmann, Ralf Rothenberger, Martin Schirneck and Ziena Zeif. I happily remember all the discussions about research, style of writing, proofs, random results and daily life we had! In Summer 2022, the Æ-team moved to a new building and, as such, we were assigned new offices. While I did not spend as much time with my new office mates from K-2.9/10, that is, Michelle Luise Döring, Gregor Lagodzinski, Louise Molitor, Leila Parsaei-Majd and, again, Ziena Zeif, the time we had was certainly filled with a lot of laughter and (vertical and horizontal) movement of our tables. Over the years, I spent particularly much time with the real estate gang, that is, Maximilian Katzmann, Martin Krejca, Louise Molitor and Aishwarya Radhakrishnan. I do believe we all could sell a house now.

While I cannot mention all further memories with the 30+ members of the Æ-team, I can assure you that I value all the research, barbecues, board-games, sports, evening activities and late-night paper-writing sessions we did together. That was a lot of fun. Thank you all!

Collecting all these memories would not be possible without the support of my parents, who I am very grateful to. They supported me way before Potsdam was

even a topic for me. I also love to remember doing various kinds of sports, watching different sport events, travelling, the parties, the discussions about life, the gaming nights (PES and Yahtzee, in particular) with my family and my friends (including the new ones I made all over the globe). That was a hell of an experience, thank you all!

Lastly, I would like to thank *you*, the reader of this thesis, for taking the time to view my work. I hope you enjoy reading it as much as I enjoyed writing it.

Thank you!

Contents

1 Introduction

Motivation

Humans are usually neither presented all rules of a language nor the complete list of its words. Yet, they are able to infer and to use the language. Thus, sufficient information to *learn a language* must come in an implicit way. To investigate this learning process, Gold [Gol67] proposes an initial formalism for language learning. Here, a *learner* (a computable function) successively receives information of a *target language* (a subset of the natural numbers) from a *text* (a list of all and only the positive information of the target language). After receiving a new datum, the learner produces a *conjecture* (a Gödel number) which language it believes to be presented. For example, when learning the set of all prime numbers $\mathbb{P}$, a learner may conjecture a Gödel number of the set of all odd numbers while it is presented odd elements from $\mathbb{P}$. Once it receives a 2 as input, the learner changes its guess to a code of the set $\mathbb{P}$.

Once these guesses converge to a *single* correct explanation of the target language, we say that the learner successfully learned said target language. Note that this is a *syntactic* requirement for convergence as the learner needs to eventually fix a *single* correct conjecture. This learning criterion is known as *explanatory learning* and denoted as **TxtGEx**.[1] Naturally, each single language may be learnt by a learner which always suggests a conjecture for this specific language. Thus, we consider classes of languages which can be learnt by a single learner. We refer to the set of classes of languages **TxtGEx**-learnable by some learner as the *learning power* of **TxtGEx**-learning or **TxtGEx**-learners.

We focus on a generalisation of explanatory learning. As each language has multiple correct explanations, that is, there exist multiple Gödel numbers for the same language, inference may be considered correct if the learner outputs correct but possibly different descriptions of the target language. That is, the learner is required to converge *semantically* instead of syntactically. This criterion is termed *behaviourally correct* learning [CL82; OW82] and, analogously to **TxtGEx**, denoted as **TxtGBc**.

1 In particular, a *text* (**Txt**) provides all and only the positive information of the target language, from which *Gold-style* (**G**) learners, which know the order and frequency of the information presented, then infer their conjectures. Lastly, **Ex** stands for explanatory learning.

These learning criteria can be further generalised or altered in order to study the impact of certain restrictions on the learning power. Affecting the *mode of data presentation*, we limit the amount of data presented to the learner to mimic memory sensitive learners. For example, in *set-driven* learning (abbreviated as **Sd**, Wexler and Culicover [WC80]), the learner is only presented the set of data and has to infer the language knowing neither the frequency nor the order of the given information. Providing the learner with an additional counter results in *partially set-driven* learning (**Psd**, Blum and Blum [BB75] and Schäfer-Richter [Sch84]). These alterations to the mode of data presentation affect Gold-style learning, that is, **G**-learning, and are accordingly denoted as **TxtSdBc** and **TxtPsdBc**, respectively.

Another approach requires the learners to follow a desired behaviour. On one hand, this behaviour can be motivated naturally, such as requiring the conjectures to correctly reflect the information they are built on. In the formalisation of Gold [Gol67], this is referred to as *consistent* learning (**Cons**, Angluin [Ang80]). On the other hand, the desired behaviour may be inspired by other sciences. To give an example, psychologists observe that children, upon learning a language [MM69; RB99], show certain patterns [Mar+92]. When learning to use the third form of irregular English verbs, children first learn the correct use, for example "catch – caught". However, they overgeneralise the use of the regular form, where the verbs end with "-ed", and start using the incorrect form "catched", just to correct their wrong behaviour again. Fittingly, such a learning behaviour is referred to as a *U-shape*. To understand whether U-shapes are needed, one can require the learner in the formalisation of Gold [Gol67] to omit using U-shapes, resulting in *non-U-shaped* learning (**NU**, Baliga et al. [Bal+08]). Both requirements can be attached to the ones above, resulting in **TxtGConsBc**- or **TxtGNUBc**-learning.

Throughout this thesis, we consider further important restrictions. We motivate them shortly by giving an intuitive description.

- *Monotonic variants.* Here, we require the output of the learners to follow a monotonic behaviour [Jan91; LZ93; Wie91]. In the case of *strongly monotone* learning (**SMon**), the guesses have to contain all previously conjectured elements. For *monotone* learning (**Mon**), this requirement only needs to be fulfilled on correctly inferred elements. In *weakly monotone* learning (**WMon**) the conjectures have to be strongly monotone while they are consistent.

- *Cautious variants.* In *cautious* learning (**Caut**, Osherson et al. [OSW82]), the conjectures must not be proper subsets of previous guesses. In the case of *cautiously infinite* ($\mathbf{Caut}_\infty$, Kötzing and Palenta [KP16]) and *cautiously finite* ($\mathbf{Caut}_{\mathbf{Fin}}$, Kötzing and Palenta [KP16]) learning, this holds true only on infinite and finite instances, respectively. In *target-cautious* learning (**Caut**, Kötzing and Palenta [KP16]), the conjectures must not overgeneralise the target language.

- *Conservative variants.* In *semantically conservative* learning (**SemConv**, Kötzing et al. [KSS17]) the learners must not change a hypothesis while it is consistent with the data given. In *semantically witness-based* learning (**SemWb**, Kötzing et al. [KSS17]) the learners must justify each mind change they make.

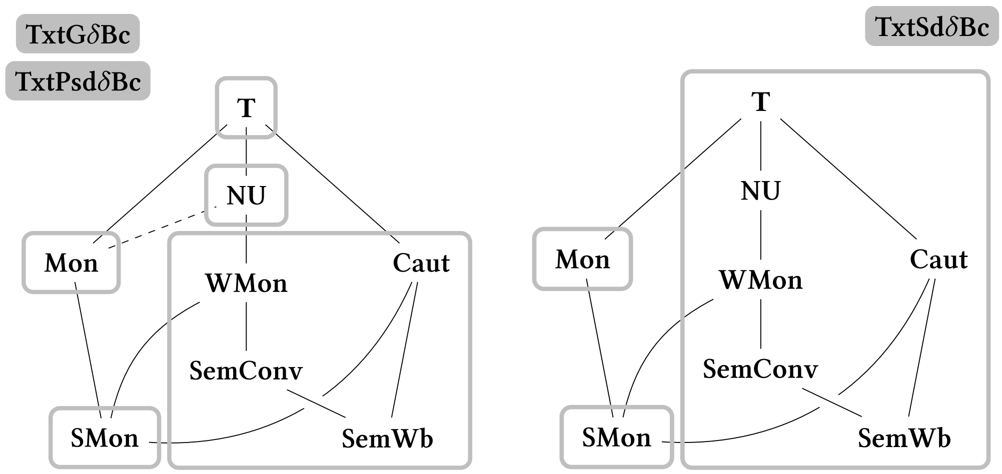
In this work, we focus on behaviourally correct learning under the mentioned restrictions. As such, we do not consider explanatory learning. Furthermore, consistent learning will get special attention. Thus, when referring to *all restrictions considered* or, shortly, *all restrictions* we mean the restrictions above except for **Ex** and **Cons** (if not explicitly stated otherwise).

Often investigated alone or in groups of similar restrictions, the main focus of study is whether the given restriction reduces the learning power of **TxtGEx**- or **TxtGBc**-learners. For example, based on the model of Gold [Gol67], Bārzdiņš [Bār77] unveils that consistency hinders the learning power of **TxtGEx**-learners, restraining them from learning certain languages. This contradicts the natural intuition and is termed the *inconsistency phenomenon*. Similarly, in the model of Gold [Gol67], non-U-shaped learning has been studied [Bal+08; CC13; CK10; CK16a; CM08; FJO94] and found to be restrictive in the case of behaviourally correct learning [Bal+08; FJO94]. This would indicate that the seemingly inefficient U-shapes are actually needed for learning.

While these studies give an insight on particular restrictions, a more global picture remains undrawn. Recently, Jain et al. [Jai+16], Kötzing and Palenta [KP16] and Kötzing and Schirneck [KS16] initiated another branch of research where the focus lies on the thorough comparison of presumably different groups of restrictions, birthing the so-called *map charting*. The term is motivated by the core of the approach. One places the restrictions of interest on a table and arranges them according to trivial implications regarding their learning power. Starting with this *backbone*, one continues to fill the *map* by comparing the different learning powers. The goal is to complete the map and, thus, to understand the relation between different restrictions and foster new insights. For example, in the case of Gold-style explanatory learning, cautious, weakly monotone and conservative [Ang80] learning coincide and monotone learning implies non-U-shaped learning.

Our Contributions

In the case of explanatory learning, different restrictions have been studied under different modes of data presentation. As such, various maps have been presented



(a) The interplay of various restrictions in Gold-style and partially set-driven learning.

(b) The interplay of various restrictions in set-driven learning.

Figure 1.1: A depiction of the relation between the studied restrictions (see Section 2.2 for the detailed definitions) for Gold-style and partially set-driven (see Figure 1.1 (a)) as well as set-driven learning (see Figure 1.1 (b)). Solid and dashed lines imply trivial and non-trivial inclusions (both bottom-to-top), respectively. Areas enclosed in a grey border illustrate a collapse of the enclosed learning criteria. There are no further collapses.

in the literature [Jai+16; KP16; KS16]. However, for behaviourally correct learning only partial results on the pairwise interaction of different restrictions are known [Bal+08; FJO94; Jai+99; KSS17]. We continue this work and *complete* the literature on behaviourally correct learning of formal languages in the following ways. In particular, we are not only interested in the pairwise relations but also present full results regarding interesting properties of the restrictions, fostering the understanding of the similarities and differences between the restrictions even further.

Behaviourally Correct Map. We compare the various important restrictions listed above with each other, providing complete maps. With that, in particular, we observe when certain restrictions can be replaced by others without changing the learning power. This continues the work of Jain et al. [Jai+16], Kötzing and Palenta [KP16] and Kötzing and Schirneck [KS16] in the explanatory case and completes initial results presented in the literature [Bal+08; FJO94; Jai+99; KSS17]. In particular, we show that cautious, semantically conservative and weakly monotone

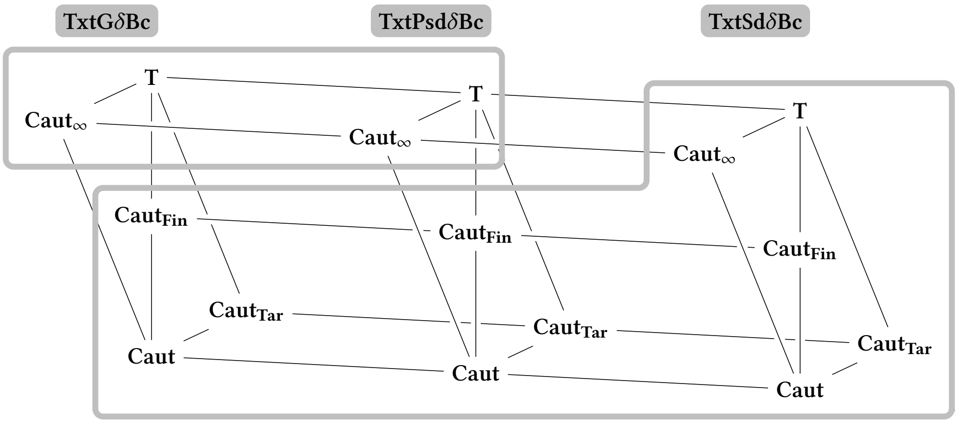


Figure 1.2: A depiction of the relations between the studied variants of cautious learning (see Section 2.2 for the detailed definitions) for Gold-style, partially set-driven and set-driven learning in *one* map. Solid lines imply trivial inclusions (bottom-to-top and right-to-left). Areas enclosed in a grey border illustrate a collapse of the enclosed learning criteria. There are no further collapses.

learning coincides for all studied modes of data representation. Furthermore, we observe that monotone learning implies non-U-shaped learning. The overall picture is presented in Figure 1.1.

Target-cautious learning is a useful variant of cautious learning as it generalises many important restrictions, such as weakly monotone, semantically conservative and cautious learning. Further variants to cautious learning exist [KP16] and we take a particularly close look at them. By studying different variants of this restriction, we aim to understand better which kinds of languages hinder learning. In particular, we find that conjectures for finite languages influence the learners learning abilities. We present the results in Figure 1.2.

Consistent Behaviourally Correct Learning. Consistent learning may be seen as a natural requirement when learning a language. However, explanatory learners cannot achieve full learning power when forced to be consistent [Bār77]. On the other hand, the opposite seems to be the case for behaviourally correct learning. Under plenty of restrictions, behaviourally correct learners are known to be consistent without loss of generality [Car+06; KSS17]. However, for some important restrictions similar results are missing. We provide a complete picture by showing that all restrictions considered in this work allow for consistent learning. In particular, we answer the open question [KSS17] whether cautious learning may be done with consistent learners. The result is collected in Corollary 4.11.

Strongly Bc-Locking Behaviourally Correct Learning. We further focus on normal forms in inductive inference. These make the learners mathematically graspable. **Bc**-locking sequences [BB75; Jai+99] are sequences of elements of the target language containing sufficient information for a learner to correctly infer the said language and never change its mind semantically any more. Such **Bc**-locking sequences are frequently used in various proofs. Unfortunately, in general, there are texts which do not contain **Bc**-locking (sub-)sequences [BB75], making it impossible to use well-known approaches which build on the existence of such sequences. Knowing that **Bc**-locking sequences occur on any text, that is, being *strongly* **Bc**-*locking*, enables us to use such approaches again. Kötzing et al. [KSS17] provide general results when strongly **Bc**-locking learning may be assumed. However, some important restrictions are not covered. In this work, we complete the literature by showing that each considered restriction allows for strongly **Bc**-locking learning. The result is collected in Corollary 5.2.

Methods. To obtain above results, we use methods from computability theory. However, since we are dealing with semantic learning, these differ from the methods used in the explanatory case [Jai+16; KP16; KS16], where the focus lies on syntactic learning. In particular, explanatory learning relies on the syntactic output of the learners, meaning that, if there is an output, one can deduce whether there has been any syntactic change. On the other hand, in behaviourally correct learning, the set enumerated by the conjectures is of more importance. As the sets are enumerated, elements may only occur in the limit. We study ways to approach this issue. First, we focus on different simulation techniques searching for **Bc**-locking sequences and study these for their general properties, see Section 3.1. Throughout the work, we also provide enumeration techniques which are custom made for the given problem. In particular, this parallels the work by Case and Kötzing [CK16a], who provide a general result on separation techniques.

Structure of the Thesis

The remainder of this thesis is structured as follows. In Chapter 2, we discuss necessary notions and preliminary results used in this thesis. Afterwards, we collect all results to complete the picture mentioned above. In Chapter 3, we provide the necessary results for set-driven behaviourally correct learning. We do the same for partially set-driven learning in Chapter 4 and Gold-style learning in Chapter 5. We conclude the thesis in Chapter 6, where we also point to possible future work.

2 Preliminaries

In this chapter we discuss the formal notation and important concepts used in this thesis. In particular, we start with the mathematical notation, see Section 2.1. In Section 2.2, we discuss the framework for language learning in the limit. This is followed by useful normal forms observed in this framework, see Section 2.3.

2.1 Mathematical Notation

Regarding mathematical notation, we mainly follow the book by Rogers Jr. [Rog87]. In particular, we denote with $\mathbb{N} = \{0, 1, 2, \dots\}$ the set of all natural numbers and with $\emptyset$ the empty set. Given $x \in \mathbb{N}$, we write $\mathbb{N}_{>x} = \{x+1, x+2, \dots\}$, that is, the set of all natural numbers greater than x. The set $\mathbb{N}_{\geq x}$ is defined analogously. We use $\subseteq$ ($\subsetneq$) to denote the (proper) subset relation for two sets. Furthermore, we use $\setminus$ to denote the difference between two sets. For a set $A \subseteq \mathbb{N}$ and $n \in \mathbb{N}$, we denote with $A^{\leq n}$ the set of all sequences of elements in A of length at most n. The sets $A^{<n}$, $A^{\geq n}$ and $A^{>n}$ are defined analogously. We write A^* for the set of all finite sequences of elements in A. Given two finite sequences $\sigma, \tau \in \mathbb{N}^*$, we write $\sigma \subseteq \tau$ ($\sigma \subsetneq \tau$) if σ is a (proper) subsequence of τ and $\sigma \leq \tau$ ($\sigma < \tau$) if σ, interpreted as a natural number, is less than or equal to (less than but not equal to) τ interpreted as a natural number. We choose the interpretation as natural numbers such that $\sigma \subseteq \tau$ implies $\sigma \leq \tau$. For a finite set $D \subseteq \mathbb{N}$ or sequence $\sigma \in \mathbb{N}^*$, we write $|D|$ or $|\sigma|$ to denote the cardinality of D or length of σ, respectively. We denote the concatenation of two finite sequences $\sigma, \tau \in \mathbb{N}^*$ as $\sigma^\frown\tau$ or (if it is clear from the context) $\sigma\tau$. Additionally, for a non-empty, finite sequence $\sigma \in \mathbb{N}^*$ we write σ^- for the sequence σ without its last element. Furthermore, we let $\mathcal{P}$ ($\mathcal{R}$) be the set of all (total) computable functions $p\colon \mathbb{N} \to \mathbb{N}$. For any function $f\colon \mathbb{N} \to \mathbb{N}$, we let $\mathrm{dom}(f)$ and $\mathrm{range}(f)$ be the domain and range of f, respectively. We fix an effective numbering $\{\varphi_e\}_{e\in\mathbb{N}}$ of all partial computable functions and denote the e-th computable set as $W_e = \mathrm{dom}(\varphi_e)$. We refer to e as the *program* or *index* of W_e. Additionally, for any step $t \in \mathbb{N}$, we write W_e^t for the set of all elements which the program e enumerates in at most t steps. The total computable function $\mathrm{enum}(.,.)$ enumerates all elements of a given program, that is, for all $e \in \mathbb{N}$ we have $W_e = \mathrm{range}(\mathrm{enum}(e, .))$.

We learn (formal) languages $L \subseteq \mathbb{N}$, that is, recursively enumerable sets, using learners, that is, (partial) computable functions. We present data to the learners and write # for the *pause symbol*, that is, no information. Consequently, for any set $S \subseteq \mathbb{N}$, we write $S_\# := S \cup \{\#\}$. A *text* is then a total function $T: \mathbb{N} \to \mathbb{N}_\#$. For any text (or sequence) T, we define the *content* of T as $\mathrm{content}(T) := \mathrm{range}(T) \setminus \{\#\}$ and say that T is a text of the language L if $\mathrm{content}(T) = L$. We write **Txt** for the set of all texts and we write $\mathbf{Txt}(L)$ for the set of all texts of L. Furthermore, for any $n \in \mathbb{N}$ and any text $T \in \mathbf{Txt}$, we let $T[n]$ be the initial sequence of T of length n, that is, $T[0] := \epsilon$ (the empty string) and, if $n > 0$, $T[n] := (T(0), T(1), \dots, T(n-1))$. Additionally, we call the text (or sequence) listing all elements of a set $A \subseteq \mathbb{N}$ in ascending order and exactly once the *canonical text* (or *canonical sequence*) *of* A. In case A is finite, we list the pause symbol after the last element of A occurs in the text. Lastly, for $t, t' \in \mathbb{N}$ and finite sets $D, D' \subseteq \mathbb{N}$, we write $(D, t) \preceq (D', t')$ if and only if $t \leq t'$ and there exists a text $T \in \mathbf{Txt}$ such that $D = \mathrm{content}(T[t])$ and $D' = \mathrm{content}(T[t'])$.

2.2 Language Learning in the Limit

In this section, we discuss the language learning in the limit framework and with it the formalisation of learning criteria we use. For the latter, we follow Kötzing [Köt09]. An *interaction operator* β provides the learner with the information to infer its hypotheses from. Formally, the interaction operator β takes a learner $h \in \mathcal{P}$ and a text $T \in \mathbf{Txt}$ as input and outputs a (possibly partial) function p. We call the function p the *sequence of hypotheses*. We consider the interaction operator **Sd** for *set-driven* learning [WC80], providing the learners only with the set of elements to base their conjecture on, **Psd** for *partially set-driven* learning [BB75; Sch84], where the learners additionally get an iteration-counter, and **G** for *Gold-style* learning [Gol67], where the learners receive full information on the order and amount of elements presented. Formally, we have, for every $i \in \mathbb{N}$,

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

In the case of Gold-style learning, we call $h(T[i])$ the *conjecture, hypothesis* or *guess of* h *on* $T[i]$ and interpret it as the $h(T[i])$-th computably enumerable set $W_{h(T[i])}$. As we are mainly interested in the semantic output, we interchangeably call $W_{h(T[i])}$ the conjecture, hypothesis or guess of h on $T[i]$. Analogous definitions hold for $h(\mathrm{content}(T[i]), i)$ and $h(\mathrm{content}(T[i]))$. We may refer to a β-learner h in its *starred form* h^*, that is, the **G**-learner simulating h. For example, the starred form of an **Sd**-learner h is defined as, for every $i \in \mathbb{N}$, $h^*(T[i]) = h(\text{content}(T[i]))$.

We formalise learning criteria as follows. To meet the criterion of *explanatory* learning (**Ex**, Gold [Gol67]) the learner is expected to converge to a *single*, correct hypothesis. We focus on a relaxation thereof: We expect the learner to be correct from some point onwards. This is referred to as *behaviourally correct* learning (**Bc**, Case and Lynes [CL82] and Osherson and Weinstein [OW82]) and allows the learner to *syntactically* change its mind while outputting a *semantically* correct guess. Formally, a restriction δ is a predicate on a sequence of hypotheses p and a text $T \in \textbf{Txt}$. For the mentioned criteria, we have

$$\textbf{Ex}(p, T) \Leftrightarrow \exists n_0 \in \mathbb{N}\, \forall n \in \mathbb{N}, n \geq n_0 \colon W_{p(n)} = \text{content}(T) \wedge p(n) = p(n_0),$$
$$\textbf{Bc}(p, T) \Leftrightarrow \exists n_0 \in \mathbb{N}\, \forall n \in \mathbb{N}, n \geq n_0 \colon W_{p(n)} = \text{content}(T).$$

In this thesis, we consider further restrictions. In *non-U-shaped* learning (**NU**, Baliga et al. [Bal+08]), learners may never discard a correct guess. Furthermore, for *consistent* learning (**Cons**, Angluin [Ang80]), each hypothesis of the learners must contain the information it is based on. We consider various monotonic restrictions [Jan91; LZ93; Wie91]. For *strongly monotone* learning (**SMon**), the learner may not discard elements present in previous guesses. In *monotone* learning (**Mon**), this applies only to elements correctly inferred, that is, elements belonging to the target language. For *weakly monotone* learning (**WMon**), the learners may not discard any elements while their guess is consistent with the information provided. In *cautious* learning (**Caut**, Osherson et al. [OSW82]), the learners may never conjecture guesses which are proper subsets of previous guesses. Several relaxations thereof have been introduced [KP16]. In *infinitely cautious* ($\textbf{Caut}_\infty$) and *finitely cautious* learning ($\textbf{Caut}_\textbf{Fin}$), cautiousness is demanded only on infinite and finite instances, respectively. Another relaxation requires the learners to never conjecture proper supersets of the target language and is referred to as *target-cautious* learning ($\textbf{Caut}_\textbf{Tar}$). For *semantically conservative* learning (**SemConv**, Kötzing et al. [KSS17]), the learners may not change a hypothesis while it is consistent with the data given and for *semantically witness-based* learning (**SemWb**, Kötzing et al. [KSS17]), the learners must justify each mind change they make. Formally, we have, for any sequence of hypotheses p and any text $T \in \textbf{Txt}$,

$$\textbf{NU}(p, T) \Leftrightarrow \forall i, j, k \in \mathbb{N} \colon \big(i \leq j \leq k \wedge W_{p(i)} = W_{p(k)} = \text{content}(T)\big) \Rightarrow$$
$$\Rightarrow W_{p(i)} = W_{p(j)},$$
$$\textbf{Cons}(p, T) \Leftrightarrow \forall i \in \mathbb{N} \colon \text{content}(T[i]) \subseteq W_{p(i)},$$

$$\begin{aligned}
\mathbf{SMon}(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon i \le j \Rightarrow W_{p(i)} \subseteq W_{p(j)},\\
\mathbf{Mon}(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon i \le j \Rightarrow \mathrm{content}(T) \cap W_{p(i)} \subseteq \mathrm{content}(T) \cap W_{p(j)},\\
\mathbf{WMon}(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon \big(i \le j \wedge \mathrm{content}(T[j]) \subseteq W_{p(i)}\big) \Rightarrow W_{p(i)} \subseteq W_{p(j)},\\
\mathbf{Caut}(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon W_{p(i)} \subsetneq W_{p(j)} \Rightarrow i \le j,\\
\mathbf{Caut}_\infty(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon \big(i < j \wedge W_{p(j)} \subsetneq W_{p(i)}\big) \Rightarrow W_{p(j)} \text{ is finite},\\
\mathbf{Caut}_{\mathbf{Fin}}(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon \big(i < j \wedge W_{p(j)} \subsetneq W_{p(i)}\big) \Rightarrow W_{p(j)} \text{ is infinite},\\
\mathbf{Caut}_{\mathbf{Tar}}(p,T) &\Leftrightarrow \forall i \in \mathbb{N}\colon \neg(\mathrm{content}(T) \subsetneq W_{p(i)}),\\
\mathbf{SemConv}(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon \big(i \le j \wedge \mathrm{content}(T[j]) \subseteq W_{p(i)}\big) \Rightarrow W_{p(i)} = W_{p(j)},\\
\mathbf{SemWb}(p,T) &\Leftrightarrow \forall i,j \in \mathbb{N}\colon \big(\exists k \in \mathbb{N}\colon i < k \le j \wedge W_{p(i)} \neq W_{p(k)}\big) \Rightarrow\\
&\qquad\qquad \Rightarrow \big(\mathrm{content}(T[j]) \cap W_{p(j)}\big) \setminus W_{p(i)} \neq \emptyset.
\end{aligned}$$

We combine two restrictions δ and δ' by intersecting them and denote this using the juxtaposition $\delta\delta'$. With $\mathbf{T}$ we denote the predicate which is always true and interpret it as absence of a learning restriction. If not stated otherwise, if we refer to *all restrictions*, we mean *all but* **Ex** *and* **Cons**. Furthermore, with *cautious variants* we mean **Caut**, $\mathbf{Caut}_{\mathbf{Tar}}$, $\mathbf{Caut}_{\mathbf{Fin}}$ and $\mathbf{Caut}_\infty$.

Finally, a learning criterion is a tuple $(\alpha, \mathcal{C}, \beta, \delta)$ consisting of learning restrictions α and δ, a set of admissible learners $\mathcal{C}$, usually $\mathcal{P}$ or $\mathcal{R}$, and an interaction operator β. We denote this learning criterion as $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$ and omit $\mathcal{C}$ if it equals $\mathcal{P}$ and the learning restrictions α or δ in case they equal $\mathbf{T}$. We say an admissible learner $h \in \mathcal{C}$ $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learns a language L if, for any text $T \in \mathbf{Txt}$, we have $\alpha(\beta(h,T),T)$ and, for all $T \in \mathbf{Txt}(L)$, we have $\delta(\beta(h,T),T)$. Intuitively, we require α to hold on any arbitrary text, while δ needs to be true on texts of the target language. We write $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta(h)$ for the class of all languages the learner h $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learns and $[\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta]$ for the set containing, for all $h' \in \mathcal{C}$, all classes $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta(h')$. We refer to $[\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta]$ as the *learning power* of $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learning or $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learners.

We collect the existing results for behaviourally correct learning needed in this thesis. It is known that partially set-driven learning is no restriction [Car+06; KR88]. On the other hand, set-driven learning is a restriction [Ful90].

▶ **Theorem 2.1 ([Car+06; Ful90; KR88]).** We have

$$[\mathbf{TxtSdBc}] \subsetneq [\mathbf{TxtPsdBc}] = [\mathbf{TxtGBc}].$$

◀

Even stronger separations from set-driven learning hold [KSS17].

▶ **Theorem 2.2 ([KSS17]).** We have

$$[\mathbf{TxtPsdNUMonBc}] \setminus [\mathbf{TxtSdBc}] \neq \emptyset.$$

◀

In Gold-style behaviourally correct learning, non-U-shaped learning is restrictive [Bal+08]. Due to partially set-driven and Gold-style learning coinciding [Car+06; KR88], the analogous result holds for partially set-driven learning.

▶ **Theorem 2.3 ([Bal+08; Car+06; KR88]).** We have, for all $\beta \in \{\mathbf{G}, \mathbf{Psd}\}$,

$$[\mathbf{Txt}\beta\mathbf{NUBc}] \subsetneq [\mathbf{Txt}\beta\mathbf{Bc}].$$

◀

Furthermore, we know that monotone learning is a proper restriction but strictly more powerful than strongly monotone learning [Jai+99]. Note that these separations are *topological*, that is, they do not rely on arguments from computability theory [CK16b]. Furthermore, we note that the class of languages not learnable by any monotone learner [Jai+99] is learnable by a weakly monotone (and non-U-shaped) one.

▶ **Theorem 2.4 ([Jai+99]).** We have, for all $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$,

$$[\mathbf{Txt}\beta\mathbf{SMonBc}] \subsetneq [\mathbf{Txt}\beta\mathbf{MonBc}] \subsetneq [\mathbf{Txt}\beta\mathbf{Bc}],$$
$$[\mathbf{Txt}\beta\mathbf{WMonBc}] \setminus [\mathbf{Txt}\beta\mathbf{MonBc}] \neq \emptyset.$$

◀

Lastly, for any interaction operator, semantically conservative and semantically witness-based learning coincide [KSS17]. Furthermore, weakly monotone learning implies cautious learning [KSS17].

▶ **Theorem 2.5 ([KSS17]).** We have, for all $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$,

$$[\mathbf{Txt}\beta\mathbf{SemWbBc}] = [\mathbf{Txt}\beta\mathbf{SemConvBc}],$$
$$[\mathbf{Txt}\beta\mathbf{WMonBc}] \subseteq [\mathbf{Txt}\beta\mathbf{CautBc}].$$

◀

2.3 Normal Forms in Language Learning

Learners may have particular *normal forms* which are useful when studying them mathematically. For example, proofs are simplified if the considered learners are total. However, this property cannot be expected in general. We discuss when it can be assumed. We call a learning restriction *semantic* if conjectures may be replaced with semantically equivalent ones without violating the restriction itself [Köt17; KSS17]. Formally, a learning restriction δ is semantic if, for any text $T \in \mathbf{Txt}$ and any sequences of hypotheses p, p' which satisfy, for any $n \in \mathbb{N}$, that $p(n)\downarrow$ if and only if $p'(n)\downarrow$ and if $p(n)\downarrow$ then $W_{p(n)} = W_{p'(n)}$, we have that $(p, T) \in \delta$ implies $(p', T) \in \delta$. Note that all restrictions considered in this work except **Ex** are semantic. It is known that learning under semantic restrictions may be done with *total* learners.

▶ **Theorem 2.6 ([KSS17]).** Let δ be a semantic learning restriction and β an interaction operator. Then, every $\mathbf{Txt}\beta\delta\mathbf{Bc}$-learner may be assumed total. ◀

We also consider **Bc**-*locking sequences* which contain enough information for a learner to infer the target language correctly and never change its mind semantically regardless what information from the target language it is presented [Jai+99]. This is a generalisation of *locking sequences* in the syntactic counterpart [BB75]. Formally, for a language $L \subseteq \mathbb{N}$ and a Gold-style learner $h \in \mathcal{P}$, a sequence $\sigma \in L_\#^*$ is a **Bc**-*locking sequence for h on L* if, for any $\tau \in L_\#^*$, we have that $W_{h(\sigma\tau)} = L$. For partially set-driven learners h, we call a finite set $D \subseteq \mathbb{N}$ and $t \in \mathbb{N}_{\geq |D|}$ a **Bc**-*locking information for h on L* if for any $(D', t') \geq (D, t)$ with $D' \subseteq L$, we have that $W_{h(D',t')} = L$. Analogously, for set-driven learners h, we call a finite set $D \subseteq \mathbb{N}$ a **Bc**-*locking set for h on L* if, for any finite $D' \subseteq \mathbb{N}$, with $D \subseteq D' \subseteq L$, we have that $W_{h(D')} = L$. We use the term **Bc**-*locking data* to subsume all three concepts.

While it is known that every **Bc**-learner has a **Bc**-locking sequence on every language it learns [BB75], there are learners and texts where no initial sequence of the text is a **Bc**-locking sequence [BB75]. We call a learner $h \in \mathcal{P}$ *strongly* **Bc**-*locking on a language L* if for each text $T \in \mathbf{Txt}(L)$ there exists an $n \in \mathbb{N}$ such that $T[n]$ is a **Bc**-locking sequence for h on L. We call h *strongly* **Bc**-*locking* [KP16; KS16; KSS17] if it is strongly **Bc**-locking on every language it learns. The transition to partially set-driven and set-driven learning is immediate and, thus, omitted. Synonymously, we say that a restriction $\tau(\alpha)C\mathbf{Txt}\beta\delta$ *allows for strongly* **Bc**-*locking learning* if every $\tau(\alpha)C\mathbf{Txt}\beta\delta$-learner may be assumed strongly **Bc**-locking. To collect which learners are known to be strongly **Bc**-locking, we need the following concept.

All considered restrictions have in common that they *allow for simulation on equivalent text* [KSS17]. Intuitively, a learner $h' \in \mathcal{P}$ seeing a text $T \in \mathbf{Txt}$ may simulate a learner $h \in \mathcal{P}$ on a different text $T' \in \mathbf{Txt}$ for the same language in the following way. The learner h' may use later and later hypotheses of h on T', but always only hypotheses of h which are based on the same data that is already presented to h'. Formally, a learning restriction δ allows for simulation on equivalent text if, for all texts $T, T' \in \mathbf{Txt}$ with $\mathrm{content}(T) = \mathrm{content}(T')$, all sequences of hypotheses p and all unbounded non-decreasing functions $r\colon \mathbb{N} \to \mathbb{N}$, we have that $\delta(p, T')$ and, for all $n \in \mathbb{N}$, $\mathrm{content}(T[n]) = \mathrm{content}(T'[r(n)])$ implies $\delta(p \circ r, T)$. We now state which learners are known to be strongly **Bc**-locking.

▶ **Theorem 2.7 ([KSS17]).** The following learners are strongly **Bc**-locking.

1. Every **Sd**-learner is strongly **Bc**-locking.
2. Let δ be a restriction which allows for simulation on equivalent text. Then, every class of languages that can be $\mathbf{TxtPsd}\delta\mathbf{Bc}$-learned can be learned so by a strongly **Bc**-locking learner.
3. Let β be an interaction operator and let $\delta \subseteq \mathbf{NU}$ be a restriction. Then, every class of languages that can be $\mathbf{Txt}\beta\delta\mathbf{Bc}$-learned can be learned so by a strongly **Bc**-locking learner. ◀

The theorem, in particular, states that, except for **TxtGMonBc**, **TxtGCautBc**, $\mathbf{TxtGCaut}_{\infty}\mathbf{Bc}$, $\mathbf{TxtGCaut}_{\mathbf{Fin}}\mathbf{Bc}$ and $\mathbf{TxtGCaut}_{\mathbf{Tar}}\mathbf{Bc}$, we may assume the learners for all learning criteria considered in this work to be strongly **Bc**-locking. Throughout this work, we will complete this list.

Consistency is also an important property which many semantic learners seem to exhibit. We say that a learning restriction δ *allows for consistent* **Bc**-*learning* [KSS17] if, for any interaction operator $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$, we have

$$[\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}].$$

Intuitively, the learning may be assumed to be consistent on arbitrary text. Note that some of the discussed restrictions allow for consistent **Bc**-learning. The remaining ones will be dealt with throughout this work.

▶ **Theorem 2.8 ([KSS17]).** The following restrictions allow for consistent **Bc**-learning: **T**, $\mathbf{Caut}_{\mathbf{Tar}}$, **Mon**, **SMon**, **WMon**, **SemConv**, **SemWb**. ◀

3 Set-Driven Learning

In this chapter, we study set-driven behaviourally correct learning. It is based on three different papers, all of which are joint work with Timo Kötzing. In particular, in Section 3.1 we focus on variations of cautious learning. This section includes results from Doskoč and Kötzing [DK20]. Solely Lemma 3.2 stems from Doskoč and Kötzing [DK22]. Furthermore, in Section 3.2 we study semantically witness-based behaviourally correct learning. Here, we include all results from Doskoč and Kötzing [DK21b]. Lastly, in Section 3.3 we connect the results of the previous two sections, that is, Sections 3.1 and 3.2. This section is based on results from Doskoč and Kötzing [DK22].

In set-driven behaviourally correct learning, monotone learning is a restriction, which, in turn, is strictly more powerful than strongly monotone learning [Jai+99]. Furthermore, it is known that semantically conservative and semantically witness-based learning coincide [KSS17]. To obtain results separating two learning paradigms, classes of languages have been provided which can be learnt respecting one learning paradigm but not the other. To show equalities between two learning paradigms, one usually takes a learner following one paradigm and then mimics what it learns while maintaining the other, desired restriction.

Throughout this chapter, we study various ways to mimic learning paradigms and, in particular, learners. Popular approaches to do so in the setting of explanatory learning involve searching for locking sequences. These sequences contain enough information about the target language so that the learner can infer it correctly, making them a valuable object of interest. In explanatory learning, one can deduce that a sequence is *not* locking from the output of the learner. However, in behaviourally correct learning, one cannot rule out a **Bc**-locking sequence (the semantic pendant to locking sequences) from the learners (syntactic) output. Rather, one has to consider the set enumerated by the program output by the learner. This raises the need for different searching strategies.

In Section 3.1, we study general approaches to do so, starting with the *weak forward verification* which we then extend to the *strong forward verification*. Doing so, we obtain interesting insights into the learners and learning paradigms. In particular, we show that set-driven learners may be assumed target-cautious and non-U-shaped at the same time. Searching for **Bc**-locking sequences includes looking for future hypotheses. However, plenty of restrictions rather affect previous hypotheses. For example, in cautious learning, the current hypothesis must not be a proper subset of any previously made hypothesis. In this case, we conduct a *backwards search* and follow previous hypotheses rather than future ones. Combined with properties obtained from the forward verification, we make set-driven behaviourally correct learners cautious.

In Section 3.2, we apply further variations of forward searches to study set-driven behaviourally correct learning from the "other side". We start with the presumably weak semantically witness-based learning and show that it coincides with (Gold-style) semantically conservative learning. We do so by providing step-wise generalisations. In each of those, we apply custom search paradigms, combining forward and backward searches. In contrast to Section 3.1, this is a more direct approach as the methods chosen fit the problem at hand.

This chapter is concluded by Section 3.3, where we combine the used approaches and obtained results. In particular, we bridge the gap between (target cautious and non-U-shaped) set-driven behaviourally correct learning and semantically witness-based learning. This shows that set-driven behaviourally correct learners may be assumed semantically witness-based without loss of generality.

3.1 Forward Verification and Backwards Search

In this section, we study various techniques to mimic learning paradigms in behaviourally correct learning. In particular, we show that set-driven behaviourally correct learning may be assumed to be

1. target-cautious and non-U-shaped or
2. cautious

without losing learning power,[2] see Lemma 3.2 and Theorem 3.4, respectively. We do so stepwise. For the further discussion, let $h \in \mathcal{R}$ be a total learner, let $L \subseteq \mathbb{N}$ be a target language and let $\sigma, \tau \in L_{\#}^{*}$ be finite sequences thereof.

In general, 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 study ways to search for **Bc**-locking sequences. By doing so, we show that, amongst other useful properties, **Sd**-learning is target-cautious and non-U-shaped in general. The *Weak*

2 Later, we show that set-driven behaviourally correct learners may be assumed semantically witness-based, see Theorem 3.5. Note that this result implies all three mentioned restrictions.

Algorithm 1: Weak Forward Verification (**WFV**), $h_w \in \mathcal{R}$

Parameter: An **Sd**-learner $h \in \mathcal{R}$.
Input: A finite set $D \subseteq \mathbb{N}$.
Semantic Output: $W_{h_w(D)} = \bigcup_{i \in \mathbb{N}} E_i$.
Initialisation: $E_0 \leftarrow D$.

1 **for** $i = 0$ **to** ∞ **do**
2 $\quad x_i \leftarrow \mathrm{enum}(h(D), i)$
3 $\quad$ **if** $x_i \notin E_i$ **then**
4 $\quad\quad$ **for** $D', D \subseteq D' \subseteq E_i \cup \{x_i\}$ **do**
5 $\quad\quad\quad$ search for $t \in \mathbb{N}$ such that $E_i \cup \{x_i\} \subseteq W^t_{h(D')}$
6 $\quad E_{i+1} \leftarrow E_i \cup \{x_i\}$

Forward Verification (**WFV**), see Algorithm 1, serves as a first step to search for **Bc**-locking sets.

The intuition is the following. Given a finite input $D \subseteq L$, **WFV** starts by enumerating D. Now, at step $i \in \mathbb{N}$, let E_i be what **WFV** has enumerated so far and let x_i be the element newly enumerated by (the program) $h(D)$, see line 2. If D were a **Bc**-locking set for h on L, then, for D' with $D \subseteq D' \subseteq E_i \cup \{x_i\}$, every possible next hypothesis $h(D')$ 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 **Bc**-locking set are still sustained, thus, **WFV** enumerates x_i and continues with step $i + 1$.

As every **Sd**-learner is strongly **Bc**-locking, see Kötzing et al. [KSS17], the **WFV** algorithm, upon enumerating the whole target language L, also has to enumerate **Bc**-locking sets for h on L. These sets, in the checking phase of the **WFV**, see lines 4 and 5, prevent the algorithm from enumerating more than the target language, resulting in target-cautious learning.

▶ **Lemma 3.1.** We have

$$[\tau(\mathbf{Cons})\mathbf{TxtSdCaut_{Tar}Bc}] = [\mathbf{TxtSdBc}].$$

◀

Proof. The inclusion $[\tau(\mathbf{Cons})\mathbf{TxtSdCaut_{Tar}Bc}] \subseteq [\mathbf{TxtSdBc}]$ follows immediately. For the other inclusion, let $h \in \mathcal{R}$ be a **Sd**-learner. We show that $\mathbf{TxtSdBc}(h) \subseteq \tau(\mathbf{Cons})\mathbf{TxtSdCaut_{Tar}Bc}(h_w)$ for $h_w \in \mathcal{R}$ from Algorithm 1. Note that, due to the initialisation, h_w is consistent on any input by definition. Now, 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. [KSS17], there exists $n_0 \in \mathbb{N}$ 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 $i \in \mathbb{N}$ 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 $E_i \cup \{x_i\}$ eventually, that is, there exists some $t \in \mathbb{N}$ such that $E_i \cup \{x_i\} \subseteq W^t_{h(D')}$. Thus, the element 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 this part of the proof.

To prove that h_w learns L respecting $\mathbf{Caut_{Tar}}$, assume the opposite, namely the existence of a finite set $D'' \subseteq L$ 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 $i \in \mathbb{N}$ be the step[3] where $D_0 \cup \{x\}$ is enumerated into $W_{h_w(D'')}$, that is, $D_0 \cup \{x\} \not\subseteq E_i$ and $D_0 \cup \{x\} \subseteq E_{i+1}$. Then, by lines 4 and 5, for $D' = D_0$, we have $x \in E_{i+1} \subseteq W_{h(D')} = L$, a contradiction. ■

In fact, we may apply the weak forward verification (Algorithm 1) onto the resulting learner *another* time and obtain an equally powerful learner which is *both* target-cautious and non-U-shaped.

▶ **Lemma 3.2.** We have

$$[\tau(\mathbf{Cons})\mathbf{TxtSdCaut_{Tar}NUBc}] = [\mathbf{TxtSdBc}].$$

◀

Proof. The inclusion $[\tau(\mathbf{Cons})\mathbf{TxtSdCaut_{Tar}NUBc}] \subseteq [\mathbf{TxtSdBc}]$ is immediate. For the other, let $h \in \mathcal{R}$ be a learner and let $\mathcal{L} = \mathbf{TxtSdBc}(h)$. Applying a *weak forward search* algorithm (Algorithm 1) we may assume h to be target-cautious and everywhere consistent, see Lemma 3.1. We show that, by applying the same algorithm again, we get a learner which also $\mathbf{NU}$-learns $\mathcal{L}$.

Let $h_w \in \mathcal{R}$ be given as in Algorithm 1 with parameter h. By Lemma 3.1, the learner h_w $\mathbf{TxtSdBc}$-learns $\mathcal{L}$. We further remark that h_w remains target-cautious and consistent as, in particular, h is a $\mathbf{Sd}$-learner. It remains to be shown that h_w is also non-U-shaped. To that end, let $L \in \mathcal{L}$ and assume there exists a finite $D \subseteq L$

3 Note that x and x_i may differ.

with $W_{h_w(D)} = L$. We show that for all finite D'', with $D \subseteq D'' \subseteq L$, we have $W_{h_w(D'')} = L$.

We first show that D is a **Bc**-locking set for h on L. For finite D' with $D \subseteq D' \subseteq L$ and $x \in W_{h_w(D)} = L$, let $i \in \mathbb{N}$ be the step[4] in Algorithm 1 such that $D' \cup \{x\}$ is enumerated into $W_{h_w(D)}$, that is, $D' \cup \{x\} \nsubseteq E_i$ and $D' \cup \{x\} \subseteq E_{i+1}$. Then, as $D \subseteq D' \subseteq E_{i+1}$ we have, by lines 4 and 5,

$$x \in E_{i+1} \subseteq W_{h(D')}.$$

Thus, for each finite D' with $D \subseteq D' \subseteq L$ we get for all $x \in L$ that $x \in W_{h(D')}$. So we have $L \subseteq W_{h(D')}$ and, since h is target-cautious, even $L = W_{h(D')}$. Altogether, the set D is a **Bc**-locking set for h on L.

Now we show that for finite D'', with $D \subseteq D'' \subseteq L$, the algorithm runs through every step $i \in \mathbb{N}$ successfully. This way, we obtain $W_{h_w(D'')} = W_{h(D'')} = L$. Let $E_0 = D''$ and let i be the next step in Algorithm 1. If $x_i \in E_i$, step i is completed successfully. Otherwise, the algorithm checks whether for each finite D', with $D'' \subseteq D' \subseteq L$, we have some $t \in \mathbb{N}$ such that $E_i \cup \{x_i\} \subseteq W^t_{h(D')}$. As $W_{h(D')} = L$ and as $E_i \cup \{x_i\}$ is a finite subset of L, such t will eventually be found. Thus, x_i will be enumerated into E_{i+1} and, hence, into $W_{h_w(D'')}$. This concludes the proof. ■

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 $t \in \mathbb{N}$ in line 5 of Algorithm 1 we have $E_i \cup \{x_i\} \subsetneq W^t_{h(D')}$. We discuss how to exploit such elements in the search for **Bc**-locking sets. If D, and thus every $D' \subseteq E_i \cup \{x_i\}$ as well, were **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 to 7. For later usage, we state the algorithm in a generalised form, accepting any β-learner. Note that we omit using the starred notation of the considered learners to ease readability.

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. [Car+06] and providing new ones.

▶ **Proposition 3.3.** Let $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$. Given a learner $h \in \mathcal{R}$ and with it the learner $h_s \in \mathcal{R}$ 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.

4 Note that x and x_i may differ.

Algorithm 2: Strong Forward Verification (**SFV**), $h_s \in \mathcal{R}$

Parameter: A β-learner $h \in \mathcal{R}$.
Input: A finite sequence $\sigma \subseteq \mathbb{N}^*$.
Semantic Output: $W_{h_s(\sigma)} = \bigcup_{i \in \mathbb{N}} E_i$.
Initialisation: $E_0 \leftarrow D$.

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

(ii) Let σ_0 be **Bc**-locking data for h on some $L \subseteq \mathbb{N}$. Then σ_0 is **Bc**-locking data 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) Let $\sigma \in \mathbb{N}^*$ be a finite sequence. If $W_{h_s(\sigma)}$ is infinite, then $W_{h_s(\sigma)} = W_{h(\sigma)}$ and σ is **Bc**-locking data for h and h_s on $W_{h_s(\sigma)} = W_{h(\sigma)}$.

(v) Let $L \in \mathbf{Txt}\beta\mathbf{Caut_{Tar}Bc}(h)$ and let σ_0 be **Bc**-locking data for h_s on L. Then σ_0 is **Bc**-locking data for h on L.

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

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

◀

Proof. We proceed to prove each of the statements.

(i) Let h be a β-learner. As all inquiries to sequences occur within h in its starred form, namely $h(\sigma)$ in line 2, $h(\sigma\tau'')$ in line 5 and $h(\sigma\tau'')$ and $h(\sigma\tau')$

5 In Corollary 4.2 we see that the same holds true for $\beta = \mathbf{G}$.

in line 7, the learner h_s requires the same form of information. Furthermore, by definition, h_s is consistent on arbitrary input. Altogether, h_s is a β-learner which is consistent on arbitrary input.

(ii) Let σ_0 be **Bc**-locking data for h on some $L \subseteq \mathbb{N}$ and let $\sigma \in L_\#^*$ be such that $\sigma_0 \subseteq \sigma$. We show that $W_{h_s(\sigma)} = L$. By definition, $W_{h_s(\sigma)} \subseteq W_{h(\sigma)} = L$. Now, let $i \in \mathbb{N}$ 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 $\tau'' \in (E_i \cup \{x_i\})_\#^{\leq i}$, as $E_i \cup \{x_i\}$ is a finite subset of $L = W_{h(\sigma\tau'')}$, we find $s_{\tau''} \in \mathbb{N}$ such that $E_i \cup \{x_i\} \subseteq W_{h(\sigma\tau'')}^{s_{\tau''}}$. Then, again, for every $\tau' \in (E_i \cup \{x_i\})_\#^{\leq i}$ we find $t \in \mathbb{N}$ such that

$$\bigcup_{\tau'' \in D_\#^{\leq i}} W_{h(\sigma\tau'')}^{s_{\tau''}} \subseteq W_{h(\sigma\tau')}^{t},$$

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 $\beta \neq$ **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, that is, $L \in \mathbf{Txt}\beta\mathbf{Bc}(h_s)$. To do so, let $T \in \mathbf{Txt}(L)$. As h is strongly **Bc**-locking, see Kötzing et al. [KSS17], there exists $n_0 \in \mathbb{N}$ such that $T[n_0]$ is **Bc**-locking data for h on L. By Proposition 3.3 (ii), $T[n_0]$ is also **Bc**-locking data 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, that is, 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 $s \in \mathbb{N}$ in lines 4 and 5 or in the check in lines 6 and 7, and thus $W_{h_s(\sigma)}$ must be finite, a contradiction.

For the second property, we first show that σ is **Bc**-locking data 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.C.: There exists $x \in W_{h(\sigma\tilde{\tau})} \setminus L$. Let $t_0 \in \mathbb{N}$ be such that $x \in W_{h(\sigma\tilde{\tau})}^{t_0}$. Let $i_0 \in \mathbb{N}$ be the step such that $|E_{i_0}| > |W_{h(\sigma\tilde{\tau})}^{t_0}|$, $E_{i_0} \supseteq \mathrm{content}(\sigma\tilde{\tau})$ as well as $\tilde{\tau} \in (E_{i_0+1})_\#^{\leq i_0}$. Such i_0 exists as $|E_i|$ increases to infinity (with increasing i) and $L = W_{h_s(\sigma)} \supseteq \mathrm{content}(\sigma\tilde{\tau})$. As the check in lines 6 and 7 must be successful, we have for $\tau' = \varepsilon \in (E_{i_0+1})_\#^{\leq i_0}$ that

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

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

2.C.: There exists $x \in L \backslash W_{h(\sigma\tilde{\tau})}$. Let $i_0 \in \mathbb{N}$ 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 lines 6 and 7 in the **SFV**, for $\tau' = \tilde{\tau} \in (E_{i_0+1})_\#^{\leq i_0}$ we have

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

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

Altogether, we get that σ is **Bc**-locking data for h on $W_{h_s(\sigma)} = W_{h(\sigma)}$. By Proposition 3.3 (ii), it also is for h_s.

(v) Let $L \in \mathbf{Txt}\beta\mathbf{Caut}_{\mathbf{Tar}}\mathbf{Bc}(h)$ and let σ_0 be **Bc**-locking data for h_s on L. Assume that σ_0 is no **Bc**-locking data for h on L, that is, 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 $\mathbf{Caut}_{\mathbf{Tar}}$.

(vi) Let D_0 be a **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) $D \subsetneq D_0$, assume the existence of some $x \in W_{h_s(D)} \setminus L$. Let $i_0 \in \mathbb{N}$ be the step[6] 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 lines 6 and 7, for $D' = D_0$, we have $x \in \bigcup_{D \subseteq D'' \subseteq E_{i_0+1}} W_{h(D'')}^{s_{D''}} \subseteq W_{h(D')} = L$, a contradiction. ■

So far, we have seen various ways to search for **Bc**-locking sequences. While this search maintains **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. Recall that h is a learner and D is some finite input. We

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

Algorithm 3: Backwards Search (**BS**), $h_b \in \mathcal{R}$

Parameter: An **Sd**-learner $h \in \mathcal{R}$.
Input: A finite set $D \subseteq \mathbb{N}$.
Semantic Output: $W_{h_b(D)} = \bigcup_{i \in \mathbb{N}} E_i$.
Initialisation: $E_0 \leftarrow D$.

```
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
```

start by enumerating $E_0 = D$. At step $i \in \mathbb{N}$, let E_i be the elements enumerated so far. It seems like a promising idea to check whether for some $D' \subseteq D$ the output of a previous hypothesis $h(D')$ exceeds what is enumerated so far, that is, whether we have $E_i \subsetneq W^i_{h(D')}$. 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.

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 3.3 (iv) and (vi), we can circumvent this problem.

▶ **Theorem 3.4.** 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 $h \in \mathcal{R}$ be a total learner and let $L \in \mathbf{TxtSdBc}(h)$, that is, the language L can be **TxtSdBc**-learned by h. By Lemma 3.1, we may assume $L \in \mathbf{TxtSdCaut_{Tar}Bc}(h)$. By Proposition 3.3 (iii), we may even assume the learning to be done by $h_s \in \mathcal{R}$ from Algorithm 2, that is, $L \in \mathbf{TxtSdCaut_{Tar}Bc}(h_s)$. This way, we are allowed to exploit Proposition 3.3 further. Now, let $h_b \in \mathcal{R}$ be as in Algorithm 3 with h_s as parameter. We proceed to 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 $n_0 \in \mathbb{N}$ 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 $\neg(\exists D' \subseteq L\colon W_{h_s(D')} \supsetneq L)$ due to h_s being $\mathbf{Caut}_{\mathbf{Tar}}$. For infinite L, let $n_0 \in \mathbb{N}$ 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. [KSS17]. Let $n \geq n_0$ and $D \coloneqq \mathrm{content}(T[n])$. We study the candidates for a possible enumeration, that is, $D' \subseteq D$, 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[7] Proposition 3.3 (iv), equal to L, or

(II) $W_{h_s(D')}$ is finite and, due to Proposition 3.3 (vi), a subset of L.

Note that, as $D \supseteq D_0$, there exists such D' fulfilling Condition (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 $s_{D'} \in \mathbb{N}$ 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 Condition (I). In this case, $s_{D'}$ will be the first step where the enumeration of $W_{h_s(D')}$ is finished. Formally, we define

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

So, no later than at step $\max\{s_{D'} \mid D' \subseteq D \wedge D \subseteq W_{h_s(D')}\}$ 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 Condition (I), the enumeration would have to continue, and thus enumerate x into $W_{h_b(D)}$, a contradiction. 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 show that h_b is $\mathbf{Caut}$ when learning L. In order to do so, assume the opposite, that is, there exist D_1, D_2 with $D_1 \subseteq D_2 \subseteq L$ such that $W_{h_b(D_1)} \supsetneq W_{h_b(D_2)}$. For finite $W_{h_b(D_2)}$, let $i_0 \in \mathbb{N}$ 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

7 By Proposition 3.3 (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, $L = W_{h_s(D')}$.

also the point where $W^{i_1}_{h_b(D_1)}$ got enumerated, that is, $W^{i_1}_{h_s(D')} = W^{i_1}_{h_b(D_1)}$ 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 3.3 (iv), D'' is a **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 **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 **Bc**-locking set for h_s on two different languages $W_{h_s(D')}$ and $W_{h_s(D'')}$, a contradiction.

Since h_b is τ(**Cons**) by definition, we get $L \in \tau(\mathbf{Cons})\mathbf{TxtSdCautBc}(h_b)$. Thus, the proof is concluded. ■

3.2 Semantically Witness-Based Learning

In this section, we consider the problem of generalising set-driven behaviourally correct learners from *below*. In particular, we study semantically witness-based learners and show that τ(**ConsSemWb**)**TxtSdBc**-learners are as powerful as **TxtGSemConvBc** ones (Theorem 3.5). We prove this result stepwise. We start by showing that each **TxtGSemConvBc**-learner may be assumed semantically conservative on arbitrary text (Theorem 3.6). Afterwards, we prove that such learners base their guesses solely on the content given (Theorem 3.7). Lastly, we observe that they remain equally powerful when being globally semantically witness-based and consistent (Theorem 3.8).

▶ **Theorem 3.5.** We have

$$[\tau(\mathbf{ConsSemWb})\mathbf{TxtSdBc}] = [\mathbf{TxtGSemConvBc}].$$

◀

We make a **TxtGSemConvBc**-learner $h \in \mathcal{R}$ *globally* semantically conservative first.

▶ **Theorem 3.6.** We have

$$[\tau(\mathbf{SemConv})\mathbf{TxtGBc}] = [\mathbf{TxtGSemConvBc}].$$

◀

Proof. The inclusion $[\tau(\mathbf{SemConv})\mathbf{TxtGBc}] \subseteq [\mathbf{TxtGSemConvBc}]$ is immediate. For the other direction, let $h \in \mathcal{R}$ be a consistent learner [KSS17] and let $\mathcal{L} = \mathbf{TxtGSemConvBc}(h)$. We provide a learner $h' \in \mathcal{R}$ which $\tau(\mathbf{SemConv})\mathbf{TxtGBc}$-learns $\mathcal{L}$.

We do so with the help of an auxiliary $\tau(\mathbf{SemConv})\mathbf{TxtGBc}$-learner $\hat{h} \in \mathcal{R}$, which only operates on sequences without repetitions or pause symbols. For convenience, we subsume these using the term *duplicates*. When h' is given a sequence with duplicates, say $(7, 1, 5, 1, 4, \#, 3, 1)$, it mimics $\hat{h}$ given the same sequence without duplicates, that is, $h'(7, 1, 5, 1, 4, \#, 3, 4) = \hat{h}(7, 1, 5, 4, 3)$. First, note that this mapping of sequences preserves the $\subseteq$-relation on sequences, thus making h' also a $\tau(\mathbf{SemConv})$-learner. Furthermore, it suffices to focus on sequences without duplicates since consistent, semantically conservative learners cannot change their mind when presented a datum they have already witnessed (or a pause symbol). Thus, $\hat{h}$ will be presented sufficient information for the learning task, which then again is transferred to h'. With this in mind, we only consider **sequences without duplicates**, that is, without repetitions or pause symbols, for the entirety of this proof. Sequences where duplicates may potentially still occur (for example when looking at the initial sequence of a text) are also replaced as described above. To ease notation, given a set A, we write $\mathbb{S}(A)$ for the subset of $A^*_\#$ where the sequences do not contain duplicates. Now, we define the auxiliary learner $\hat{h}$.

Consider the learner $\hat{h}$ as in Algorithm 4 with parameter h. Given some input sequence $\sigma \in \mathbb{N}^*$, the intuition is the following. Once $\hat{h}$, on any previous sequence $\sigma' \subseteq \sigma$, is consistent with the currently given information $\mathrm{content}(\sigma)$, the learner only enumerates the same as such hypotheses (lines 2 to 4). While no such hypothesis is found, $\hat{h}$ does a forward search (lines 5 to 9) and only enumerates elements if all visible future hypotheses also witness these elements. As already discussed, $\hat{h}$ operates only on sequences without repetitions or pause symbols, thus making it possible to check *all* necessary future hypotheses.

First we show that for any $L \in \mathcal{L}$ and any $T \in \mathbf{Txt}(L)$ we have, for $n \in \mathbb{N}$,

$$W_{\hat{h}(T[n])} \subseteq W_{h(T[n])}. \tag{3.1}$$

Note that by our assumption, while the (infinite) text T may contain duplicates, the (finite) sequence $T[n]$ does not. Now, we show Equation (3.1) by induction on $n \in \mathbb{N}$. The case $n = 0$ follows immediately. Assume Equation (3.1) holds up to n. As $\mathrm{content}(T[n+1]) \subseteq W_{h(T[n+1])}$ by consistency of h and as, for $n' \leq n$,

Algorithm 4: The auxiliary $\tau(\mathbf{SemConv})$-learner $\hat{h} \in \mathcal{R}$.

Parameter: A **TxtGSemConv**-learner $h \in \mathcal{R}$.
Input: A finite sequence $\sigma \in \mathbb{S}(\mathbb{N})$.
Initialisation: $t' \leftarrow 0$, $E_0 \leftarrow \mathrm{content}(\sigma)$ and, for all $t > 0$, $E_t \leftarrow \emptyset$.

1 **for** $t = 0$ **to** ∞ **do**
2   **if** $\exists \sigma' \subsetneq \sigma \colon \mathrm{content}(\sigma) \subseteq W^t_{\hat{h}(\sigma')}$ **then**
3     $\Sigma'_t \leftarrow \left\{\sigma' \subsetneq \sigma : \mathrm{content}(\sigma) \subseteq W^t_{\hat{h}(\sigma')}\right\}$
4     $E_{t+1} \leftarrow E_t \cup \bigcup_{\sigma' \in \Sigma'_t} W^t_{\hat{h}(\sigma')}$
5   **else if** $\forall \sigma' \subsetneq \sigma \colon \mathrm{content}(\sigma) \not\subseteq W^t_{h(\sigma')}$ **then**
6     $S(\sigma, t') \leftarrow \mathbb{S}\left(W^{t'}_{h(\sigma)} \setminus \mathrm{content}(\sigma)\right)$
7   **if** $\forall \tau \in S(\sigma, t') \colon \bigcup_{\tau' \in S(\sigma,t')} W^{t'}_{h(\sigma\tau')} \subseteq W^t_{h(\sigma\tau)}$ **then**
8       $E_{t+1} \leftarrow E_t \cup W^{t'}_{h(\sigma)}$
9       $t' \leftarrow t' + 1$
10   **else**
11     $E_{t+1} \leftarrow E_t$

$W_{h(T[n'])} = W_{h(T[n+1])}$ whenever $\mathrm{content}(T[n+1]) \subseteq W_{h(T[n'])}$, we get

$$W_{\hat{h}(T[n+1])} \subseteq \bigcup_{\substack{n' \leq n, \\ \mathrm{content}(T[n+1]) \subseteq W_{\hat{h}(T[n'])}}} W_{\hat{h}(T[n'])} \cup W_{h(T[n+1])} \subseteq W_{h(T[n+1])}.$$

The first inclusion follows as the big union contains all previous hypotheses found in the first if-clause (lines 2 to 4) and as $W_{h(T[n+1])}$ contains all elements possibly enumerated by the second if-clause (lines 5 to 9). Note that the latter also contains $\mathrm{content}(T[n+1])$, thus covering the initialisation. The second inclusion follows by the induction hypothesis and semantic conservativeness of h.

We continue by showing that $\hat{h}$ **TxtGBc**-learns $\mathcal{L}$. To that end, let $L \in \mathcal{L}$ and $T \in \mathbf{Txt}(L)$. We distinguish the following two cases.

1.C.: L is finite. Then there exists $n_0 \in \mathbb{N}$ with $\mathrm{content}(T[n_0]) = L$. Let $n \geq n_0$. As h is **SemConv** and consistent, we have $L = W_{h(T[n])}$. By Equation (3.1), we have $L = W_{h(T[n])} \supseteq W_{\hat{h}(T[n])}$ and, as $\hat{h}$ is consistent, $W_{\hat{h}(T[n])} \supseteq \mathrm{content}(T[n]) = L$. Altogether we have $W_{\hat{h}(T[n])} = L$ as required.

2.C.: L is infinite. Let $n_0 \in \mathbb{N}$ be minimal such that $W_{h(T[n_0])} = L$. Then, as h is semantically conservative, $T[n_0]$ is a **Bc**-locking sequence for h on L and we have

$$\forall i < n_0 \colon \mathrm{content}(T[n_0]) \nsubseteq W_{h(T[i])}.$$

Thus, elements enumerated by $W_{\hat{h}(T[n_0])}$ cannot be enumerated by the first if-clause (lines 2 to 4) but only by the second one (lines 5 to 9). We show $W_{\hat{h}(T[n_0])} = L$. The $\subseteq$-direction follows immediately from Equation (3.1). For the other direction, let $t' \in \mathbb{N}$ be the current step of enumeration and let

$$S(T[n_0], t') = \mathbb{S}\Big(W^{t'}_{h(T[n_0])} \setminus \mathrm{content}(T[n_0])\Big).$$

As $T[n_0]$ is a **Bc**-locking sequence, we have, for all $\tau \in S(T[n_0], t')$,

$$\bigcup_{\tau' \in S(T[n_0], t')} W^{t'}_{h(T[n_0])^\frown \tau'} \subseteq W_{h(T[n_0]^\frown \tau)} = L.$$

Thus, at some step $t \in \mathbb{N}$, E_{t+1} is set to $W^{t'}_{h(T[n_0])}$ and then the enumeration continues with $t' + 1$. In the end we have $L \subseteq W_{\hat{h}(T[n_0])}$ and, altogether, $L = W_{\hat{h}(T[n_0])}$.

We now show that, for any $n > n_0$, $L = W_{\hat{h}(T[n])}$ holds. Note that at some point $\mathrm{content}(T[n]) \subseteq W_{\hat{h}(T[n_0])}$ will be witnessed. Thus, $W_{\hat{h}(T[n])}$ will enumerate the same as $W_{\hat{h}(T[n_0])} = L$, and it follows that $L \subseteq W_{\hat{h}(T[n])}$. By Equation (3.1), $W_{\hat{h}(T[n])}$ will not enumerate more than $W_{h(T[n])} = L$, that is, $W_{\hat{h}(T[n])} \subseteq W_{h(T[n])} = L$, concluding this part of the proof.

It remains to be shown that $\hat{h}$ is **SemConv** on arbitrary text $T \in$ **Txt**. The problem is that, when a previous hypothesis becomes consistent with information currently given, the learner may have already enumerated incomparable data in its current hypothesis. This is prevented by closely monitoring the time of enumeration, namely by waiting until the enumerated data will certainly not cause such problems. We prove that $\hat{h}$ is τ(**SemConv**) formally. Let $n, n' \in \mathbb{N}$ be such that $n < n'$ and $\mathrm{content}(T[n']) \subseteq W_{\hat{h}(T[n])}$. We show that $W_{\hat{h}(T[n])} = W_{\hat{h}(T[n'])}$ by separately looking at each inclusion.

$\subseteq$: The inclusion $W_{\hat{h}(T[n])} \subseteq W_{\hat{h}(T[n'])}$ follows immediately since by assumption $\mathrm{content}(T[n']) \subseteq W_{\hat{h}(T[n])}$, meaning that at some point the first if-clause (lines 2 to 4) will find $T[n]$ as a candidate and then $W_{\hat{h}(T[n'])}$ will enumerate $W_{\hat{h}(T[n])}$.

$\supseteq$: Assume there exists $x \in W_{\hat{h}(T[n'])} \setminus W_{\hat{h}(T[n])}$. Let x be the first such enumerated and let $t_x \in \mathbb{N}$ be the step of enumeration with respect to $h(T[n'])$, that is, $x \in W^{t_x}_{h(T[n'])}$ but $x \notin W^{t_x-1}_{h(T[n'])}$. Furthermore, let $t_{\text{content}} \in \mathbb{N}$ be the step where $\text{content}(T[n']) \subseteq W_{\hat{h}(T[n])}$ is witnessed for the first time. Now, by the definition of $\hat{h}$, we have

$$W_{\hat{h}(T[n'])} \subseteq W^{t_{\text{content}}-1}_{h(T[n'])} \cup W_{\hat{h}(T[n])},$$

as $W_{\hat{h}(T[n'])}$ enumerates at most $W^{t_{\text{content}}-1}_{h(T[n'])}$ until it sees the consistent prior hypothesis, namely $\hat{h}(T[n])$. This happens exactly at step $t_{\text{content}}-1$, at which the set $W_{\hat{h}(T[n'])}$ stops enumerating elements from $W^{t_{\text{content}}-1}_{h(T[n'])}$ and continues to follow $W_{\hat{h}(T[n])}$. Now, observe that $t_x < t_{\text{content}}$ since $x \in W_{\hat{h}(T[n'])}$ but $x \notin W_{\hat{h}(T[n])}$. Let

$$S(T[n], t_{\text{content}}) = \mathbb{S}\left(W^{t_{\text{content}}}_{h(T[n])} \setminus \text{content}(T[n])\right).$$

But then, in order for $W_{\hat{h}(T[n])}$ to enumerate $\text{content}(T[n'])$ via the second if-clause (lines 5 to 9), that is, to get $\text{content}(T[n']) \subseteq W_{\hat{h}(T[n])}$, we witness

$$x \in \bigcup_{\tau' \in S(T[n], t_{\text{content}})} W^{t_{\text{content}}}_{h(T[n]^\frown \tau')} \subseteq W_{\hat{h}(T[n])}.$$

This contradicts $x \notin W_{\hat{h}(T[n])}$, concluding the proof. ∎

This result proves that h may be assumed semantically conservative on arbitrary text. Next, we show that h does not rely on the order or amount of information given.

▶ **Theorem 3.7.** We have

$$[\tau(\mathbf{SemConv})\mathbf{TxtSdBc}] = [\tau(\mathbf{SemConv})\mathbf{TxtGBc}].$$

◀

Proof. Let $h \in \mathcal{R}$ be a learner and $\mathcal{L} = \tau(\mathbf{SemConv})\mathbf{TxtGBc}(h)$. We may assume h to be globally consistent, see Kötzing et al. [KSS17]. We provide a learner $h' \in \mathcal{R}$ which $\tau(\mathbf{SemConv})\mathbf{TxtSdBc}$-learns $\mathcal{L}$. To that end, we introduce the following auxiliary notation used throughout this proof. For each finite set $D \subseteq \mathbb{N}$ and each

$x \in \mathbb{N}$, let σ_D be the canonical sequence of D and

$$d := \max(D),$$
$$D_{<x} := \{y \in D \mid y < x\}.$$

Note that the definition of $D_{<x}$ can be extended to $\leq, >$ and $\geq$ as well as infinite sets in a natural way. Now, let $h' \in \mathcal{R}$ be such that, for each finite set $D \subseteq \mathbb{N}$,

$$W_{h'(D)} = D \cup \left(W_{h(\sigma_D)}\right)_{>d} \cup \left\{x \in \left(W_{h(\sigma_D)}\right)_{<d} : D \cup \{x\} \subseteq W_{h(\sigma_{(D_{<x})})}\right\}.$$

Intuitively, $h'(D)$ simulates h assuming it got the information in the canonical order, that is, $h'(D)$ simulates $h(\sigma_D)$. All elements $x \in W_{h(\sigma_D)}$ such that $x > d$ can be enumerated, as any later, consistent hypothesis will do so as well. If $x < d$, then we check whether the learner h given the canonical sequence up to x is consistent with $D \cup \{x\}$, that is, whether $D \cup \{x\} \subseteq W_{h(\sigma_{(D_{<x})})}$. If so, we enumerate x as it will be done by the previous hypotheses as well. Note that, for each finite $D \subseteq \mathbb{N}$, we have

$$W_{h'(D)} \subseteq W_{h(\sigma_D)}. \tag{3.2}$$

We proceed by proving that h' τ(**SemConv**)**TxtSdBc**-learns $\mathcal{L}$. First, we show the **TxtSdBc**-convergence. The idea here is to find a **Bc**-locking sequence of the canonical text. Doing so ensures that even if elements are shown out of order they will be enumerated as h will not make a mind change and thus the consistency condition will be observed. To that end, let $L \in \mathcal{L}$. We distinguish whether L is finite or not.

1.C.: L is finite. We show that $W_{h'(L)} = L$. By definition of h', we have $L \subseteq W_{h'(L)}$. For the other inclusion, note that as h is consistent and semantically conservative (which, in particular, implies it being target-cautious), we have that $W_{h(\sigma_L)} = L$. Then, by Equation (3.2), we have $W_{h'(L)} \subseteq W_{h(\sigma_L)} = L$, concluding this case.

2.C.: L is infinite. Let T_c be the canonical text of L and let σ_0 be a **Bc**-locking sequence for h on T_c. Such a **Bc**-locking sequence exists as h is strongly **Bc**-locking, see Kötzing et al. [KSS17, Thm. 7]. Let $D_0 := \mathrm{content}(\sigma_0)$. For any finite input $D \subseteq L$ such that $D \supseteq D_0$, we show that $W_{h'(D)} = L$. By Equation (3.2), we get $W_{h'(D)} \subseteq W_{h(\sigma_D)} = L$. To show $L \subseteq W_{h'(D)}$, let $x \in L$. We distinguish the relative position of x and d.

$x > d$: In this case we have $x \in W_{h'(D)}$ by definition of h'.

$x \leq d$: In this case either $x \in D$ and we immediately get $x \in W_{h'(D)}$, or we have to check whether $D \cup \{x\} \subseteq W_{h(\sigma_{(D_{<x})})}$. Since σ_0 is an initial segment of the canonical text of L, it holds that $x > \max(\text{content}(\sigma_0))$ and, thus, we get $\sigma_0 \subseteq \sigma_{(D_{<x})}$. Now $W_{h(\sigma_{(D_{<x})})} = L$, meaning that $D \cup \{x\} \subseteq W_{h(\sigma_{(D_{<x})})}$ will be observed at some point in the computation. Thus, $x \in W_{h'(D)}$.

Altogether, we get $W_{h'(D)} = L$ and thus **TxtSdBc**-convergence. It remains to be shown that h' is $\tau(\mathbf{SemConv})$. Let two finite sets D', D'' with $D' \subseteq D''$ and $D'' \subseteq W_{h'(D')}$ be given. The trick here is that, upon checking for consistency with elements shown out of order, the learner has to check the same, minimal sequence regardless whether the input is D' or D''. We proceed with the formal proof. Therefore, we expand the initially introduced notation of this proof. For any $x \in \mathbb{N}$ define

$$\begin{aligned} \sigma' &:= \sigma_{D'}, \\ d' &:= \max(D'), \\ \sigma'_{<x} &:= \sigma_{(D'_{<x})}. \end{aligned}$$

Analogously, we use σ'', d'' and $\sigma''_{<x}$ when D'' is the underlying set. First, we show that $W_{h(\sigma')} = W_{h(\sigma'')}$. Since $W_{h'(D')}$ enumerates D'', that is, $D'' \subseteq W_{h'(D')}$, we have for all $y \in (D'' \setminus D')_{<d'}$ that $D' \cup \{y\} \subseteq W_{h(\sigma'_{<y})}$ by definition of h'. Thus, we have

$$W_{h(\sigma'_{<y})} = W_{h(\sigma')}. \tag{3.3}$$

Note that, if $(D'' \setminus D')_{<d'} = \emptyset$, then $\sigma'_{<d'+1} = \sigma'$. Thus, Equation (3.3) also holds for

$$m := \begin{cases} \min(D''_{<d'} \setminus D'), & \text{if } D''_{<d'} \setminus D' \neq \emptyset, \\ d' + 1, & \text{otherwise.} \end{cases}$$

Furthermore, it holds true that for any $x \leq m$ we have

$$\sigma'_{<x} = \sigma''_{<x}. \tag{3.4}$$

By Equations (3.2) and (3.3), we have $D'' \subseteq W_{h'(D')} \subseteq W_{h(\sigma')} = W_{h(\sigma'_{<m})}$. As, by Equation (3.4), $\sigma'_{<m} = \sigma''_{<m} \subseteq \sigma''$ and h is $\tau(\mathbf{SemConv})$, we get

$$W_{h(\sigma')} = W_{h(\sigma'')}. \tag{3.5}$$

We conclude the proof by showing that $W_{h'(D')} = W_{h'(D'')}$. We check each direction separately by checking every possible position of an element, which is a candidate for enumeration, relative to the given information D' and D''.

$\supseteq$: Let $x \in W_{h'(D'')}$. For $x \in D''$ we have $x \in W_{h'(D')}$ by assumption. Otherwise, by Equations (3.2) and (3.5), we get $x \in W_{h(\sigma')}$. Thus, x will be considered in the enumeration of $W_{h'(D')}$. We distinguish the relation between x and d'.

$x > d'$: In this case $x \in (W_{h(\sigma')})_{>d'} \subseteq W_{h'(D')}$.

$x < d'$: As $d' \leq d''$ and since x is enumerated into $W_{h'(D'')}$, we have $D'' \cup \{x\} \subseteq W_{h(\sigma''_{<x})}$. We, again, distinguish the relative position of x and m.

$x < m$: We have

$$D' \cup \{x\} \subseteq D'' \cup \{x\} \subseteq W_{h(\sigma''_{<x})} \overset{\text{Eq. (3.4)}}{=} W_{h(\sigma'_{<x})}.$$

$m < x < d'$: We use h being $\tau(\mathbf{SemConv})$ (marked by $(*)$) and obtain

$$\begin{aligned} D' \cup \{x\} \subseteq D'' \cup \{x\} \subseteq W_{h(\sigma''_{<x})} &\overset{(*)}{=} W_{h(\sigma'')} \overset{\text{Eq. (3.5)}}{=} W_{h(\sigma')} \overset{\text{Eq. (3.3)}}{=} \\ &\overset{\text{Eq. (3.3)}}{=} W_{h(\sigma'_{<m})} \overset{(*)}{=} W_{h(\sigma'_{<x})}. \end{aligned}$$

In both cases the checks pass and, thus, $x \in W_{h'(D')}$.

Altogether, we have $W_{h'(D'')} \subseteq W_{h'(D')}$.

$\subseteq$: Let $x \in W_{h'(D')}$. For $x \in D''$ we have $x \in W_{h'(D'')}$ by definition of h'. Otherwise,

$$x \in D'' \cup \{x\} \subseteq W_{h'(D')} \subseteq W_{h(\sigma')} \overset{\text{Eq. (3.5)}}{=} W_{h(\sigma'')}.$$

Thus, x will be considered in the enumeration of $W_{h'(D'')}$. We now distinguish between the possible relations of x and d''.

$x > d''$: In this case $x \in W_{h'(D'')}$ by definition of h'.

$x < d''$: We show that $D'' \cup \{x\} \subseteq W_{h(\sigma''_{<x})}$ and, thus, x is enumerated by $W_{h'(D'')}$.

$x < m$: We have

$$D'' \cup \{x\} \subseteq W_{h(\sigma'_{<x})} \overset{\text{Eq. (3.4)}}{=} W_{h(\sigma''_{<x})}.$$

$m < x < d'$: We use h being $\tau(\mathbf{SemConv})$ in the steps marked by $(*)$ and obtain

$$D'' \cup \{x\} \subseteq W_{h(\sigma'_{<x})} \overset{(*)}{=} W_{h(\sigma'_{<m})} \overset{\text{Eq. (3.4)}}{=} W_{h(\sigma''_{<m})} \overset{(*)}{=} W_{h(\sigma''_{<x})}.$$

$d' < x < d''$: We have

$$D'' \cup \{x\} \subseteq W_{h(\sigma')} = W_{h(\sigma'_{<m})} = W_{h(\sigma''_{<m})} \overset{\text{Eq. (3.4)}}{=} W_{h(\sigma''_{<x})}.$$

As the check passes in each case, we get, in the end, $x \in W_{h'(D'')}$. ■

Hence, we may assume h to be a $\tau(\mathbf{SemConv})\mathbf{TxtSdBc}$-learner. Note that the learner h may be assumed globally consistent, see Kötzing et al. [KSS17, Thm. 8]. Lastly, we observe that h may even be assumed globally semantically witness-based. This concludes the proof of Theorem 3.5.

▶ **Theorem 3.8.** We have

$$[\tau(\mathbf{ConsSemWb})\mathbf{TxtSdBc}] = [\tau(\mathbf{SemConv})\mathbf{TxtSdBc}].$$

◀

Proof. Let $\delta \in \{\mathbf{SemWb}, \mathbf{SemConv}\}$. Since δ-learners may be assumed to be consistent, see Kötzing et al. [KSS17, Thm. 8], which also holds true when the restrictions are required globally, we have

$$[\tau(\mathbf{Cons}\delta)\mathbf{TxtSdBc}] = [\tau(\delta)\mathbf{TxtSdBc}].$$

Then, the theorem holds since $\mathbf{Cons} \cap \mathbf{SemWb} = \mathbf{Cons} \cap \mathbf{SemConv}$, see Kötzing et al. [KSS17, Lem. 11]. ■

3.3 Completing the Set-Driven Map

In Section 3.1, we step-wisely obtain specialisations for set-driven behaviour correct learners. In Section 3.2 on the other hand, we study behaviourally correct learners from the "other" side, starting with the presumably weak semantically witness-based learners. In this section, we show that all of these learners are equal regarding their learning power.

In particular, we obtain semantically conservative learners (see Theorem 3.9) in a similar fashion as when making semantically conservative learners so *everywhere*, see Theorem 3.6. Using the results from Section 3.1, we may start with target-cautious and non-U-shaped learners. Then, in particular, we can override wrong hypotheses of the learners using elements as witnesses (as, by target-cautious learning, incorrect guesses cannot overgeneralise the target language) and right hypotheses are never discarded (by non-U-shaped learning).

▶ **Theorem 3.9.** We have

$$[\tau(\mathbf{ConsSemWb})\mathbf{TxtSdBc}] = [\mathbf{TxtSdBc}].$$

◀

Proof. We immediately get $[\tau(\mathbf{ConsSemWb})\mathbf{TxtSdBc}] \subseteq [\mathbf{TxtSdBc}]$. For the other, observe that, by Theorem 3.5, we have

$$[\tau(\mathbf{ConsSemWb})\mathbf{TxtSdBc}] = [\mathbf{TxtGSemWbBc}] = [\mathbf{TxtGSemConvBc}].$$

Furthermore, by Lemma 3.2, we have

$$[\tau(\mathbf{Cons})\mathbf{TxtSdCaut}_{\mathbf{Tar}}\mathbf{NUBc}] = [\mathbf{TxtSdBc}].$$

Thus, to conclude the proof, it suffices to be shown that

$$[\mathbf{TxtGSemConvBc}] \supseteq [\tau(\mathbf{Cons})\mathbf{TxtSdCaut}_{\mathbf{Tar}}\mathbf{NUBc}].$$

Let $h \in \mathcal{R}$ $\tau(\mathbf{Cons})\mathbf{TxtSdCaut}_{\mathbf{Tar}}\mathbf{NUBc}$-learn $\mathcal{L}$. We now provide a learner $h' \in \mathcal{R}$ which **TxtGSemConvBc**-learns $\mathcal{L}$, that is, $\mathcal{L} \subseteq \mathbf{TxtGSemConvBc}(h')$. To that end, we use the learner h' as described in Algorithm 5.

We discuss the learner h' obtained from Algorithm 5 with parameter h and a finite sequence $\sigma \in \mathbb{N}^*$ as input. First note that the outer if-clause (starting at line 1) checks whether the current information σ contains a new datum or is empty. If not, then the learner outputs just the same as when given σ^-. This way, the learner only may change its mind when a new datum occurs. Otherwise, h' checks whether, on any previous sequence $\sigma' \subsetneq \sigma$, it is consistent with the currently given information $\mathrm{content}(\sigma)$. If so, the learner only enumerates the same as such hypotheses (lines 3 to 5). While no such hypothesis is found, h' does a forward search with regard to h (lines 6 to 11). Then, h' only enumerates elements which are witnessed by *all* visible future hypotheses. Checking all visible future hypotheses is possible to check as the learner h is set-driven.

Note that this is a similar approach as when making **TxtGSemConvBc**-learners *everywhere* semantically conservative, compare with the proof of Theorem 3.6. We maintain the monitoring of the time of enumeration for each element (lines 6 to 11) and checking for previous consistent hypotheses (lines 3 to 5) to prevent non-conservative behaviour. A main observation is that for the learner h' to converge correctly, the initial learner h needs not be semantically conservative. It suffices that h is target-cautious (so that wrong hypotheses lack elements from the target

Algorithm 5: The **TxtGSemConvBc**-learner $h' \in \mathcal{R}$.

Parameter: A consistent, non-U-shaped and target-cautious **Sd**-learner $h \in \mathcal{R}$.
Input: A finite sequence $\sigma \subseteq \mathbb{N}^*$.
Initialisation: $t' \leftarrow 0$, $E_0 \leftarrow \mathrm{content}(\sigma)$ and, for all $t > 0$, $E_t \leftarrow \emptyset$.

1 **if** $\sigma = \epsilon$ *or* $\mathrm{content}(\sigma^-) \subsetneq \mathrm{content}(\sigma)$ **then**
2   **for** $t = 0$ **to** ∞ **do**
3   **if** $\exists \sigma' \subsetneq \sigma \colon \mathrm{content}(\sigma) \subseteq W^t_{h'(\sigma')}$ **then**
4             $\Sigma'_t \leftarrow \left\{\sigma' \subsetneq \sigma \mid \mathrm{content}(\sigma) \subseteq W^t_{h'(\sigma')}\right\}$
5             $E_{t+1} \leftarrow E_t \cup \bigcup_{\sigma' \in \Sigma'_t} W^t_{h'(\sigma')}$
6   **else**
7             $C_\sigma \leftarrow \mathrm{content}(\sigma)$
8             $F_{\sigma,t'} \leftarrow W^{t'}_{h(C_\sigma)}$
9   **if** $\forall D \subseteq F_{\sigma,t'} \colon \bigcup_{D' \subseteq F_{\sigma,t'}} W^{t'}_{h(C_\sigma \cup D')} \subseteq W^t_{h(C_\sigma \cup D)}$ **then**
10                 $E_{t+1} \leftarrow E_t \cup F_{\sigma,t'}$
11                 $t' \leftarrow t' + 1$
12 **else**
13     $W_{h'(\sigma)} \leftarrow W_{h'(\sigma^-)}$

language which then can be used for mind-changes) and non-U-shaped (so that we do not "unintentionally" output a correct guess prematurely).

We first show that h' is semantically conservative on arbitrary text $T \in \mathbf{Txt}$, that is, h' is $\tau(\mathbf{SemConv})$. The problem is that when a previous hypothesis becomes consistent with information currently given, the learner may have already enumerated incomparable data in its current hypothesis. This is prevented by closely monitoring the time of enumeration, namely by waiting until the enumerated data will certainly not cause such problems. We prove that h' is $\tau(\mathbf{SemConv})$ formally. Let $n, n' \in \mathbb{N}$ be (the lexicographically first pair) such that $n < n'$ and $\mathrm{content}(T[n']) \subseteq W_{h'(T[n])}$. We show that $W_{h'(T[n])} = W_{h'(T[n'])}$ by separately looking at each inclusion.

$\subseteq$: We obtain the inclusion $W_{h'(T[n])} \subseteq W_{h'(T[n'])}$ as follows. By assumption, we have $\mathrm{content}(T[n']) \subseteq W_{h'(T[n])}$. This implies that at some point during the enumeration of $W_{h'(T[n'])}$ the first if-clause (lines 3 to 5) will find $T[n]$ as a candidate and then $W_{h'(T[n'])}$ will enumerate $W_{h'(T[n])}$.

$\supseteq$: Assume there exists $x \in W_{h'(T[n'])} \setminus W_{h'(T[n])}$. Let x be the first such element enumerated and let $t_x \in \mathbb{N}$ be the step of enumeration with respect to $h(\mathrm{content}(T[n']))$, that is, $x \in W^{t_x}_{h(\mathrm{content}(T[n']))}$ but $x \notin W^{t_x-1}_{h(\mathrm{content}(T[n']))}$. Similarly, let $t_{\mathrm{content}} \in \mathbb{N}$ be the step where $\mathrm{content}(T[n']) \subseteq W_{h'(T[n])}$ is witnessed for the first time. By the definition of h', as $W_{h'(T[n'])}$ sees the consistent prior hypothesis exactly at step t_{content}, we get

$$W_{h'(T[n'])} \subseteq W^{t_{\mathrm{content}}-1}_{h(\mathrm{content}(T[n']))} \cup W_{h'(T[n])}.$$

That is, $W_{h'(T[n'])}$ enumerates (at most) elements from $W^{t_{\mathrm{content}}-1}_{h(\mathrm{content}(T[n']))}$ before, at step t_{content}, it continues to follow $W_{h'(T[n])}$. Now, we have $x \in W_{h'(T[n'])}$ but $x \notin W_{h'(T[n])}$ and, therefore, $x \in W^{t_{\mathrm{content}}-1}_{h(\mathrm{content}(T[n']))}$. Thus, $t_x < t_{\mathrm{content}}$. In particular, x is enumerated via the second if-clause (lines 6 to 11). Furthermore, since $W_{h'(T[n])}$ also enumerates $\mathrm{content}(T[n'])$ via the second if-clause (lines 6 to 11), we have that

$$\bigcup_{D' \subseteq W^{t_{\mathrm{content}}}_{h(\mathrm{content}(T[n]))}} W^{t_{\mathrm{content}}}_{h(\mathrm{content}(T[n]) \cup D')} \subseteq W_{h'(T[n])}.$$

As $D' = \mathrm{content}(T[n']) \subseteq W^{t_{\mathrm{content}}}_{h(\mathrm{content}(T[n]))}$ is a candidate in the big union, we get that

$$x \in W^{t_{\mathrm{content}}-1}_{h(\mathrm{content}(T[n']))} \subseteq \bigcup_{D' \subseteq W^{t_{\mathrm{content}}}_{h(\mathrm{content}(T[n]))}} W^{t_{\mathrm{content}}}_{h(\mathrm{content}(T[n]) \cup D')} \subseteq W_{h'(T[n])},$$

contradicting $x \notin W_{h'(T[n])}$. This concludes this part of the proof.

Now that h' is shown to be semantically conservative, we show that for any $L \in \mathcal{L}$ and any $T \in \mathbf{Txt}(L)$ we have, for $n \in \mathbb{N}$,

$$W_{h'(T[n])} \subseteq W_{h(\mathrm{content}(T[n]))}. \tag{3.6}$$

We show Equation (3.6) by induction on $n \in \mathbb{N}$. The case $n = 0$ follows immediately. Assume Equation (3.6) holds up to n. Note that, by definition of h', we have

$$W_{h'(T[n+1])} \subseteq \bigcup_{\substack{n' \leq n, \\ \mathrm{content}(T[n+1]) \subseteq W_{h'(T[n'])}}} W_{h'(T[n'])} \cup W_{h(\mathrm{content}(T[n+1]))}. \tag{3.7}$$

Let $n_{\mathrm{m}} \in \mathbb{N}$ be the minimal $n' \in \mathbb{N}$ such that $\mathrm{content}(T[n+1]) \subseteq W_{h'(T[n'])}$ if such n_{m} exists. If no such n_{m} exists, this part of the proof is concluded. As h'

is **SemConv**, for all $n'' \in \mathbb{N}$ with $n_{\mathrm{m}} \leq n'' \leq n$, we have

$$W_{h'(T[n_{\mathrm{m}}])} = W_{h'(T[n''])}.$$

Furthermore, for $n'' < n_{\mathrm{m}}$, no previous guess $W_{h'(T[n''])}$ contains content$(T[n_{\mathrm{m}}])$, as otherwise, by **SemConv** of h', if content$(T[n_{\mathrm{m}}]) \subseteq W_{h'(T[n''])}$ we obtain $W_{h'(T[n''])} = W_{h'(T[n_{\mathrm{m}}])} \supseteq$ content$(T[n+1])$, a contradiction to the minimality of n_{m}. Hence,

$$\bigcup_{\substack{n' \leq n, \\ \mathrm{content}(T[n+1]) \subseteq W_{h'(T[n'])}}} W_{h'(T[n'])} = W_{h'(T[n_{\mathrm{m}}])}. \tag{3.8}$$

In particular, h' does only a forward search on input $T[n_{\mathrm{m}}]$ (lines 6 to 11). As, by doing so, it eventually witnesses content$(T[n+1])$, we get by definition of h' that

$$W_{h'(T[n_{\mathrm{m}}])} \subseteq W_{h(\mathrm{content}(T[n+1]))}. \tag{3.9}$$

Combining Equations (3.8) and (3.9) with Equation (3.7), we observe

$$\begin{aligned} W_{h'(T[n+1])} &\subseteq \bigcup_{\substack{n' \leq n, \\ \mathrm{content}(T[n+1]) \subseteq W_{h'(T[n'])}}} W_{h'(T[n'])} \cup W_{h(\mathrm{content}(T[n+1]))} = \\ &= W_{h'(T[n_{\mathrm{m}}])} \cup W_{h(\mathrm{content}(T[n+1]))} = W_{h(\mathrm{content}(T[n+1]))}. \end{aligned}$$

So, we have that Equation (3.6) holds for the induction step and, therefore, for all $n \in \mathbb{N}$.

We close the proof by showing that h' **TxtGBc**-learns $\mathcal{L}$. To that end, let $L \in \mathcal{L}$ and $T \in \mathbf{Txt}(L)$. As h' is semantically conservative, it suffices to show that there exists $n \in \mathbb{N}$ such that $W_{h'(T[n])} = L$. We provide such n by case distinction.

1.C.: L is finite. Then there exists $n_0 \in \mathbb{N}$ with content$(T[n_0]) = L$. As h is consistent and target-cautious, we have $L = W_{h(\mathrm{content}(T[n_0]))}$. By Equation (3.6), we have $L = W_{h(\mathrm{content}(T[n_0]))} \supseteq W_{h'(T[n_0])}$ and, by consistency of h', $W_{h'(T[n_0])} \supseteq$ content$(T[n]) = L$. Altogether we have $W_{h'(T[n_0])} = L$ as required.

2.C.: L is infinite. Let $n_0 \in \mathbb{N}$ be minimal such that $W_{h(\mathrm{content}(T[n_0]))} = L$. Then, as h is non-U-shaped, content$(T[n_0])$ is a **Bc**-locking set for h on L. Let $n_1 \geq n_0$ be minimal such that

$$\forall i < n_0 \colon \mathrm{content}(T[n_1]) \not\subseteq W_{h'(T[i])}. \tag{3.10}$$

Such n_1 exists due to Equation (3.6) and h being target-cautious, that is, wrong hypotheses prior to $h(\mathrm{content}(T[n_0]))$ are either proper subsets of the target language or incomparable to it. In either case, some elements of the target language are not contained in these guesses.

Now, we show that Condition (3.10) actually holds for all $i < n_1$. If $n_0 = n_1$, this is immediately given. Otherwise, note that there exists some $i_0 < n_0$ such that $\mathrm{content}(T[n_1 - 1]) \subseteq W_{h'(T[i_0])}$ (by the minimal choice of n_1 for Condition (3.10)) and thus, for all n with $i_0 \leq n \leq n_1 - 1$, $W_{h'(T[i_0])} = W_{h'(T[n])}$ (as h' is **SemConv**). In particular, we have

$$\exists i_0 < n_0 \, \forall n \in \mathbb{N}, n_0 \leq n < n_1 : W_{h'(T[i_0])} = W_{h'(T[n])}.$$

As $\mathrm{content}(T[n_1]) \not\subseteq W_{h'(T[i_0])}$, we have

$$\forall i < n_1 \colon \mathrm{content}(T[n_1]) \not\subseteq W_{h'(T[i])}. \tag{3.10$'$}$$

Hence, elements enumerated by $W_{h'(T[n_1])}$ cannot be enumerated by the first if-clause (lines 3 to 5) but only by the second one (lines 6 to 11). Next, we show $W_{h'(T[n_1])} = L$. As $W_{h'(T[n_1])} \subseteq W_{h(\mathrm{content}(T[n_1]))}$ (see Equation (3.6)) and $\mathrm{content}(T[n_1])$ is a **Bc**-locking set for h on L, we get $W_{h'(T[n_1])} \subseteq L$. For the other direction, let $t' \in \mathbb{N}$ be the current step of enumeration. Observe that Condition (3.10′) implies that $W_{h'(T[n_1])}$ enumerates elements only via the second if-clause (see lines 6 to 11). As $\mathrm{content}(T[n_1])$ is a **Bc**-locking set for h on L, we have, for all $D \subseteq W^{t'}_{h(\mathrm{content}(T[n_1]))}$,

$$\bigcup_{D' \subseteq W^{t'}_{h(\mathrm{content}(T[n_1]))}} W^{t'}_{h(\mathrm{content}(T[n_1]) \cup D')} \subseteq W_{h(\mathrm{content}(T[n_1]) \cup D)} = L.$$

Thus, at some step $t \in \mathbb{N}$, E_{t+1} is set to $W^{t'}_{h(\mathrm{content}(T[n_1]))}$ and then the enumeration continues with $t' + 1$. In the end we have $L \subseteq W_{h'(T[n_1])}$ and, altogether, $L = W_{h'(T[n_1])}$. This concludes the proof. ■

4 Partially Set-Driven Learning

In this chapter, we focus on partially set-driven behaviourally correct learning. It is based on three different papers which are all joint work with Timo Kötzing. In Section 4.1, we discuss results from Doskoč and Kötzing [DK20] on variations of cautious learning. In particular, we show which of the considered learning paradigms are restrictive and which are not. We furthermore observe that the cautious variants allow for consistent **Bc***-learning. In Section 4.2, we extend this result to hold for* all *restrictions considered in this work by showing that non-U-shaped learning also allows for consistent* **Bc***-learning. Please note that this is the only result which has not yet been published under peer-review. We furthermore observe that for learning under the considered cautious restrictions the order in which the data is presented is not important. While such a behaviour is often observed in behaviourally correct learning, we provide a* first natural *example where it is in Section 4.3. This section is based on results from Doskoč and Kötzing [DK21a]. Lastly, in Section 4.4, we complete the partially set-driven behaviourally correct map. This section is based on results from Doskoč and Kötzing [DK22].*

When it comes to partially set-driven behaviourally correct learning, initial results regarding the pairwise relations between the various restrictions are known. It is known that monotone learning is a restriction [Jai+99], which is more powerful than strongly monotone learning [Jai+99]. Furthermore, it is also known that non-U-shaped learning separates from monotone learning [Jai+99; KSS17]. As unrestricted partially set-driven learning is as powerful as unrestricted Gold-style learning [Car+06; Ful90; KR88] and since non-U-shaped learning is known to restrict the latter [Bal+08; FJO94], it can be deduced that non-U-shaped learning is a restriction in partially set-driven learning as well.

We complete these initial results to obtain a full picture regarding the pairwise relation of the learning restrictions in Section 4.1. We start by studying the various cautious learning paradigms and show which ones are restrictive and which are not. We observe that hypotheses for finite languages are troublesome for the learners. At the same time, we provide a general result with which we show that all of the considered variants follow a nice general behaviour. Particularly, all of the cautious variants allow for consistent **Bc**-learning and the learning can be done with strongly

Bc-locking learners. The former result, in particular, solves an open question stated by Kötzing et al. [KSS17].

In Section 4.2, we study which restrictions allow for consistent **Bc**-learning. Considering the literature [Car+06; KSS17] and the results obtained in Section 4.1, it only remains to be shown whether non-U-shaped learning allows for consistent **Bc**-learning or not. We note that for set-driven and Gold-style learning the answer is already provided by Lemma 3.2 and Carlucci et al. [Car+06], respectively. We complete this result by showing that non-U-shaped partially set-driven learning can be done with consistent learners. Altogether, we show that *all* considered restrictions allow for consistent **Bc**-learning, see Corollary 4.11.

In the literature [Car+06; Ful90; KR88; KSS17], in Chapter 3 and in Section 4.1, we also observe that partially set-driven learning under some restriction often coincides with its Gold-style counterpart. While many of the studied restrictions seem to follow this behaviour, we find a *first natural* example where this does not hold true. For monotone learning, we provide a separating class of languages where we use prematurely conjectured elements to trick the partially set-driven monotone learner. This way, we show that monotone learners do rely on the order the data is given to them in Section 4.3.

Finally, we complete the partially set-driven behaviourally correct map in Section 4.4. In particular, the only remaining question to answer is whether monotone learning implies non-U-shaped learning. We provide a simulation argument where we take all possible future hypotheses and prove that, in fact, the implication holds.

4.1 Cautious Restrictions

We study the variations of cautious learning in the partially set-driven case and also provide results for Gold-style learning. We are interested in various aspects thereof. On one hand, we study which of the variations are restrictive and which are not. On the other hand, we provide a general result (see Theorem 4.1) for generalisations to target-cautious learning, showing that they coincide with their Gold-style counterpart. Using Theorem 4.1, we are able to deduce that *all* considered variations of cautious learning do not rely on the order the information is presented to them. This comes in handy when comparing the pairwise relations, as we can deal with both partially set-driven and Gold-style learning at once.

We start by showing when Gold-style and partially set-driven learners may be assumed equally powerful, just as Doskoč and Kötzing [DK20] do in the explanatory case. Unfortunately, the same approach does not bear fruits, as, although performing a search for **Bc**-locking sequences, we do not mimic the learner. Rather, we enumerate the learner's output on possible **Bc**-locking sequences, as discussed in private communication with Jain [Jai17]. If, for certain languages, the Gold-style learner refrains from overgeneralising the target language, our enumeration can maintain this behaviour.

▶ **Theorem 4.1.** Let P be a predicate on languages. Let δ be a learning restriction such that, for any sequence of hypotheses p and any text $T \in \mathbf{Txt}$,

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

Then,

1. the restriction δ allows for consistent **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. we have $[\mathbf{TxtPsd}\delta\mathbf{Bc}] = [\mathbf{TxtG}\delta\mathbf{Bc}]$. ◀

Proof. We provide a proof for each of the statements.

1. We show that δ allows for consistent **Bc**-learning. We follow the proof of Kötzing et al. [KSS17]. For a total learner $h \in \mathcal{R}$ let $L \in \mathbf{Txt}\beta\delta\mathbf{Bc}(h)$. We define $g \in \mathcal{R}$ in its starred form on finite sequences $\sigma \in \mathbb{N}^*$ as

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

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

2. To show that $[\mathbf{TxtPsd}\delta\mathbf{Bc}] = [\mathbf{TxtG}\delta\mathbf{Bc}]$, first, 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 **TxtGBc**-learning can be done partially set-driven, as discussed in private communication with Jain [Jai17]. We expand this idea so that the restriction δ is also preserved. To that end, let $h \in \mathcal{R}$ be a learner and $L \in \mathbf{TxtG}\delta\mathbf{Bc}(h)$. Now, define the **Psd**-learner $h' \in \mathcal{R}$ as follows. With the

S-m-n Theorem, we get a total computable function $p \in \mathcal{R}$ such that, for finite $D \subseteq \mathbb{N}$ and $t \in \mathbb{N}_{\geq |D|}$,

$$A_{D,t} \coloneqq W_{p(D,t)} = \bigcup_{\sigma \in D_\#^*} \left(\bigcap_{\tau \in D_\#^{\leq t}} W_{h(\sigma\tau)} \cap \bigcap_{\sigma' < \sigma, \sigma' \in D_\#^*} \bigcup_{\tau' \in D_\#^*} W_{h(\sigma'\tau')} \right), \tag{4.1}$$

$$W_{h'(D,t)} = \bigcup_{s \in \mathbb{N}} \begin{cases} A_{D,t}^s, & \text{if } \exists \rho \in D_\#^{\leq t} : A_{D,t}^s \subseteq W_{h(\rho)}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Intuitively, $A_{D,t}$ checks whether the information given is enough to witness a (minimal) **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 ensures that the restriction δ is maintained.

We start by proving $L \in$ **TxtPsdBc**(h'). To that end, let $T \in$ **Txt**(L). By Blum and Blum [BB75], there exists a **Bc**-locking sequence for h on L. Let $\alpha \in \mathbb{N}^*$ be the least such **Bc**-locking sequence with respect to $<$. By Osherson et al. [OSW86], for each $\alpha' < \alpha$ such that $\mathrm{content}(\alpha') \subseteq L$, there exists $\tau_{\alpha'} \in L_\#^*$ such that $\alpha'\tau_{\alpha'}$ is a **Bc**-locking sequence for h on L. Now, let $n_0 \in \mathbb{N}$ 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 show $A_{D,t} = L$.

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

$\sigma \leq \alpha$: As x is an element of the left hand intersection of Equation (4.1) and as, for $\sigma \leq \alpha$, we have $\tau_\sigma \in D_\#^{\leq t}$, we get $x \in W_{h(\sigma\tau_\sigma)} = L$.

$\sigma > \alpha$: Here, we exploit that x is an element of the right hand intersection of Equation (4.1). As $\alpha < \sigma$ and $\alpha \in D_\#^{\leq t}$, we have, for any $\tau \in D_\#^*$, that $x \in W_{h(\alpha\tau)} = L$.

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 **Bc**-locking sequence, we witness, for every $\tau \in D_{\#}^{\leq t}$, that $x \in W_{h(\alpha\tau)} = L$. Thus, the left hand intersection of Equation (4.1) contains x.

For the right hand intersection of Equation (4.1), 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 be shown 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 $s \in \mathbb{N}$ be the next step in the enumeration of $W_{h'(D,t)}$. We check whether we can enumerate $A_{D,t}^{s}$. As $\alpha \in D_{\#}^{\leq t}$ and $A_{D,t}^{s} \subseteq L = W_{h(\alpha)}$, we have a witness that we can enumerate $A_{D,t}^{s}$. Thus, for all $s \in \mathbb{N}$ 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 a finite $D \subseteq L$ and $t \in \mathbb{N}$ 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. ■

In Theorem 4.1, choosing P as the always-true predicate $\top$ results in target cautious learning. Together with $[\mathbf{TxtSdCaut_{Tar}Bc}] = [\mathbf{TxtPsdCaut_{Tar}Bc}]$ [KSS17], we immediately obtain the following corollary.

► **Corollary 4.2.** We have

$$[\mathbf{TxtSdCaut_{Tar}Bc}] = [\mathbf{TxtPsdCaut_{Tar}Bc}] = [\mathbf{TxtGCaut_{Tar}Bc}].$$

◄

To deal with $\mathbf{Caut_{Fin}}$, we introduce a slightly less restrictive version on which we can apply results established throughout this section. In its core, this is a similar approach to Kötzing and Palenta [KP16] introducing target-cautious learning ($\mathbf{Caut_{Tar}}$) in order to deal with cautious learning (**Caut**).

► **Theorem 4.3.** We have

$$[\mathbf{TxtSdCaut_{Fin}Bc}] = [\mathbf{TxtPsdCaut_{Fin}Bc}] = [\mathbf{TxtGCaut_{Fin}Bc}].$$

◄

Proof. To prove the theorem, we apply the same idea as Kötzing and Palenta [KP16] when dealing with **Caut**, that is, we introduce a weaker version of $\mathbf{Caut}_{\mathbf{Fin}}$. For any sequence of hypotheses p and any $T \in \mathbf{Txt}$, let

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

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 4.1, 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 $h \in \mathcal{R}$ be a learner and let $L \in \mathbf{TxtPsd}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}(h)$. We first observe that, by Theorem 4.1, we may assume h to be consistent. Now, we follow the idea from Kötzing et al. [KSS17] and introduce the learner $h' \in \mathcal{R}$ which, on any finite set $D \subseteq \mathbb{N}$, is defined as $h'(D) := h(D, |D|)$. First, we show that h' **TxtSdBc**-learns L. If L is infinite, then we get $L \in \mathbf{TxtSdBc}(h)$ by Kötzing et al. [KSS17]. For finite L, let $T \in \mathbf{Txt}(L)$ and $n_0 \in \mathbb{N}$ be such that $\mathrm{content}(T[n_0]) = L$. Now, for $n \geq n_0$ and $D := \mathrm{content}(T[n]) = L$, we 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, that is, there exist a finite target language $L \subseteq \mathbb{N}$ 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, with Theorem 3.9, we have the following chain of inclusions.

$$\begin{aligned}
[\mathbf{TxtSdCautBc}] &\subseteq [\mathbf{TxtSdCaut}_{\mathbf{Fin}}\mathbf{Bc}] \subseteq \\
&\subseteq [\mathbf{TxtPsdCaut}_{\mathbf{Fin}}\mathbf{Bc}] \subseteq \\
&\subseteq [\mathbf{TxtGCaut}_{\mathbf{Fin}}\mathbf{Bc}] \subseteq \\
&\subseteq [\mathbf{TxtG}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}] = \\
&= [\mathbf{TxtSd}(\mathbf{Caut}_{\mathbf{Tar}})_{\mathbf{Fin}}\mathbf{Bc}] = \\
&= [\mathbf{TxtSdCautBc}].
\end{aligned}$$

This closes the proof. ■

We combine Corollary 4.2 and Theorem 4.3 with results from Chapter 3 and observe the following result.

▶ **Corollary 4.4.** We have

$$\begin{aligned}[\mathbf{TxtSdBc}] &= [\mathbf{TxtSdCaut_{Tar}Bc}] = [\mathbf{TxtGCaut_{Tar}Bc}] = \\ &= [\mathbf{TxtSdCaut_{Fin}Bc}] = [\mathbf{TxtGCaut_{Fin}Bc}] = \\ &= [\mathbf{TxtSdCautBc}] = [\mathbf{TxtGCautBc}] = \\ &= [\mathbf{TxtSdSemWbBc}] = [\mathbf{TxtGSemWbBc}].\end{aligned}$$

◀

Proof. For $\delta \in \{\mathbf{Caut_{Tar}}, \mathbf{Caut_{Fin}}\}$, we have

$$\begin{aligned}[\mathbf{TxtG}\delta\mathbf{Bc}] &\overset{\text{Cor. 4.2 and Thm. 4.3, respectively}}{=} [\mathbf{TxtSd}\delta\mathbf{Bc}] = \\ &\overset{\text{Thm. 3.9}}{=} [\mathbf{TxtSdSemWbBc}] = \\ &\overset{\text{Thm. 3.5}}{=} [\mathbf{TxtGSemWbBc}] \subseteq \\ &\subseteq [\mathbf{TxtG}\delta\mathbf{Bc}].\end{aligned}$$

For cautious learning we then have

$$\begin{aligned}[\mathbf{TxtGCaut_{Tar}Bc}] &\overset{\text{Cor. 4.2}}{=} [\mathbf{TxtSdCaut_{Tar}Bc}] \overset{\text{Thm. 3.9}}{=} [\mathbf{TxtSdCautBc}] \subseteq \\ &\subseteq [\mathbf{TxtGCautBc}] \subseteq [\mathbf{TxtGCaut_{Tar}Bc}].\end{aligned}$$

By Theorem 3.5, we have $[\mathbf{TxtSdSemWbBc}] = [\mathbf{TxtSdBc}]$, providing also the equality to $[\mathbf{TxtSdBc}]$. ■

To conclude the study on the variations of cautious learning, it remains to be shown that infinitely cautious learning, that is, $\mathbf{Caut_\infty}$, does not restrict the learning power of Gold-style and partially set-driven learners. We use the same idea as in the explanatory case [DK20], namely by ensuring that infinite suggestions only occur when the underlying information is **Bc**-locking data. To that end, we employ the **SFV**, see Algorithm 2.

▶ **Lemma 4.5.** We have

$$[\tau(\mathbf{Cons})\mathbf{TxtPsdCaut_\infty Bc}] = [\mathbf{TxtPsdBc}].$$

◀

Proof. By definition, we get $[\tau(\mathbf{Cons})\mathbf{TxtPsdCaut_\infty Bc}] \subseteq [\mathbf{TxtPsdBc}]$. For the other direction, let $h \in \mathcal{R}$ be a learner and let $L \in \mathbf{TxtPsdBc}(h)$. For the **Psd**-learner $h_s \in \mathcal{R}$ from Algorithm 2, we show that $L \in \tau(\mathbf{Cons})\mathbf{TxtPsdCaut_\infty Bc}(h_s)$.

By Proposition 3.3 (i), h_s is consistent on arbitrary input. As in the proof of Proposition 3.3 (iii), we get $L \in \mathbf{TxtPsdBc}(h_s)$. To show that h_s is $\mathbf{Caut}_\infty$, assume the opposite, that is, there exists $D, D' \subseteq L$ and $t, t' \in \mathbb{N}$ with $(D,t) \preceq (D',t')$ 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 3.3 (iv), firstly, (D,t) is a $\mathbf{Bc}$-locking information for $W_{h_s(D,t)}$ and, secondly, (D',t') is a $\mathbf{Bc}$-locking information both for $W_{h_s(D',t')}$ and, since $(D,t) \preceq (D',t')$ and (D,t) is a $\mathbf{Bc}$-locking information for $W_{h_s(D,t)}$, for $W_{h_s(D,t)}$ as well. Since $W_{h_s(D,t)} \neq W_{h_s(D',t')}$, this yields a contradiction. ■

We sum up the results obtained so far.

▶ **Corollary 4.6.** 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

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

Furthermore, the two blocks do not coincide. ◀

In particular, Corollary 4.6 shows that cautious learning may be assumed consistent in general. This answers an open problem stated by Kötzing et al. [KSS17]. We answer the same question for *all* considered cautious restrictions.

▶ **Corollary 4.7.** Let $\delta \in \{\mathbf{Caut}_\infty, \mathbf{Caut}_{\mathbf{Tar}}, \mathbf{Caut}_{\mathbf{Fin}}, \mathbf{Caut}\}$. Then, δ allows for consistent $\mathbf{Bc}$-learning, that is, for $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$, we have

$$[\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}].$$

◀

Together with Theorem 3.9, we get that semantically witness-based, semantically conservative, weakly monotone and cautious learning all coincide. This is the case not only for set-driven but also for partially set-driven and Gold-style learning.

▶ **Corollary 4.8.** For $\delta \in \{\mathbf{Caut}_{\mathbf{Tar}}, \mathbf{Caut}_{\mathbf{Fin}}, \mathbf{Caut}, \mathbf{SemWb}, \mathbf{SemConv}, \mathbf{WMon}\}$ and $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$, we have

$$[\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{TxtSdBc}].$$

◀

Furthermore, as partially set-driven learners combined with restrictions which allow for simulation on equivalent text are strongly **Bc**-locking without loss of generality, see Kötzing et al. [KSS17], so are all the Gold-style versions. We sum this up in the following corollary.

▶ **Corollary 4.9.** For all $\delta \in \{\mathbf{Caut}, \mathbf{Caut_{Tar}}, \mathbf{Caut_{Fin}}, \mathbf{Caut_{Tar}}\}$, we have that every $\mathbf{TxtG}\delta\mathbf{Bc}$-learner may be assumed strongly **Bc**-locking. ◀

4.2 Consistent Non-U-Shaped Learning

Whether a certain learning paradigm may be done with consistent learners is a problem often studied in the literature [Bār77; Car+06; KSS17]. Especially in behaviourally correct learning, it seems to be a natural property many learners unveil. While there are initial results on which learners may be assumed consistent, they do not cover all restrictions studied in this work. In this section, we show that, indeed, all considered restrictions allow for consistent **Bc**-learning.

In particular, it only remains to be shown for partially set-driven non-U-shaped learning. We follow a similar strategy as in the Gold-style case [Car+06]. The idea is to output a canonical hypothesis if all hypotheses based on the same content are overgeneralising, that is, containing more than the content. Otherwise, one outputs the content only. We note that it is imperative to check for *proper* overgeneralisations to allow the learner to fall back to the content if needed. We show that, despite not knowing the order of the previous hypotheses, this strategy works out.

▶ **Theorem 4.10.** The restriction **NU** allows for consistent **Bc**-learning, that is, for $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$, we have

$$[\tau(\mathbf{Cons})\mathbf{Txt}\beta\mathbf{NUBc}] = [\mathbf{Txt}\beta\mathbf{NUBc}].$$

◀

Proof. The case $\beta = \mathbf{G}$ has been shown by Carlucci et al. [Car+06],[8] while the case $\beta = \mathbf{Sd}$ follows from Lemma 3.2. We prove the statement for $\beta = \mathbf{Psd}$. The inclusion $[\tau(\mathbf{Cons})\mathbf{TxtPsdNUBc}] \subseteq [\mathbf{TxtPsdNUBc}]$ follows immediately. For the other direction, let $h \in \mathcal{R}$ be a learner and let $\mathcal{L} = \mathbf{TxtPsdNUBc}(h)$.

8 Note that Carlucci et al. [Car+06] do not explicitly talk about arbitrary texts, however, the transition of their proof is immediate and, thus, omitted.

Define the learner $g \in \mathcal{R}$ for any finite set $D \subseteq \mathbb{N}$ and $t \in \mathbb{N}_{\geq |D|}$ as

$$W_{g(D,t)} = \begin{cases} W_{h(D,|D|)}, & \text{if } \forall t' \in \{|D|, |D|+1, \ldots, t\} : D \subsetneq W_{h(D,t')}, \\ D, & \text{otherwise.} \end{cases}$$

Intuitively, the learner g checks whether the learner h actually overgeneralises the given information D on all hypotheses built from the same information, but with possibly different counters. In that case, it outputs the same hypothesis as $h(D, |D|)$. If that is not the case, the learner outputs just the content given, that is, the set D. We note that it is important to check for *proper* overgeneralisation, that is, to check for $D \subsetneq W_{h(D,t')}$. Otherwise, the learner may not be able to return to the set D if all considered previous hypotheses are consistent. It is immediate to see that g is consistent on arbitrary text. Now let $L \in \mathcal{L}$. We first show that g is **NU** on L. To show **Bc**-convergence, it then suffices to show that on each text $T \in \mathbf{Txt}(L)$ there exists some $n \in \mathbb{N}$ such that $W_{g(\text{content}(T[n]),n)} = L$.

We show that g is **NU** when learning L. To that end, let $T \in \mathbf{Txt}(L)$, $t_1, t_3 \in \mathbb{N}$, $D_1 := \text{content}(T[t_1])$ and $D_3 := \text{content}(T[t_3])$ be such that we have

$$W_{g(D_1,t_1)} = W_{g(D_3,t_3)} = L.$$

Now, we show that for any $t_2 \in \{t_1, \ldots, t_3\}$ and $D_2 := \text{content}(T[t_2])$, we have $W_{g(D_2,t_2)} = L$. We conduct a case distinction.

1.C.: We have for all $t' \in \{|D_1|, |D_1|+1, \ldots, t_1\}$ that $D_1 \subsetneq W_{h(D_1,t')}$. Then, we have

$$W_{g(D_1,t_1)} = W_{h(D_1,|D_1|)} = L.$$

As $(D_1, |D_1|) \preceq (D_2, |D_2|)$ and h is **NU**, we have that $L = W_{h(D_2,|D_2|)}$ and, for all $t' \in \{|D_2|, |D_2|+1, \ldots, t_2\}$, that $L = W_{h(D_2,t')}$. Thus, in case $D_2 \subsetneq L$, we get $W_{g(D_2,t_2)} = W_{h(D_2,|D_2|)} = L$ by the first condition in the definition of g. Otherwise, we get $W_{g(D_2,t_2)} = D_2 = L$ by the second condition in the definition of g as no proper overgeneralisation is witnessed.

2.C.: There exists $t' \in \{|D_1|, |D_1|+1, \ldots, t_1\}$ such that $\neg(D_1 \subsetneq W_{h(D_1,t')})$. Let $t'_0 \in \mathbb{N}$ be the minimal such t'. Then, by the second condition in the definition of g, we have

$$L = W_{g(D_1,t_1)} = D_1.$$

It follows that $D_1 = D_2 = D_3 = L$. Now we have that $t'_0 \in \{|D_1|, |D_1|+1, \ldots, t_2\}$ is such that $\neg(D_2 \subsetneq W_{h(D_2,t'_0)})$. Thus, $W_{g(D_2,t_2)} = D_2 = L$.

It remains to be shown that there exists $n \in \mathbb{N}$ such that $W_{g(\mathrm{content}(T[n]),n)} = L$. We distinguish the following cases.

1.C.: The language L is infinite. Consider the canonical text $T_c \in \mathbf{Txt}(L)$ of L and let $n_0 \in \mathbb{N}$ be such that $(T_c[n_0], n_0)$ is a **Bc**-locking information for h on L [KSS17]. Define $D_0 := \mathrm{content}(T_c[n_0])$. Then, for each finite set $D \subseteq L$ and $t \in \mathbb{N}$ with $(D_0, n_0) \preceq (D, t)$ we have $W_{h(D,t)} = L$. Now, let $n \in \mathbb{N}$ be such that $D_0 \subseteq \mathrm{content}(T[n])$. Then, by the first condition of the definition of g we have

$$W_{g(\mathrm{content}(T[n]),n)} = W_{h(\mathrm{content}(T[n]),|\mathrm{content}(T[n])|)} = L.$$

2.C.: The language L is finite. Let $n_0 \in \mathbb{N}$ be minimal such that $\mathrm{content}(T[n_0]) = L$. Furthermore, let $n \geq n_0$ be such that $W_{h(L,n)} = L$. Note that $\mathrm{content}(T[n]) = L$. Then, by the *second* condition in the definition of g we have $W_{g(L,n)} = L$. ■

With this result, in particular, we have that *all* restrictions considered in this work allow for consistent **Bc**-learning. We state this as a corollary.

▶ **Corollary 4.11.** All restrictions $\delta \in \{\mathbf{T}, \mathbf{NU}, \mathbf{Caut_{Tar}}, \mathbf{Caut_\infty}, \mathbf{Caut_{Fin}}, \mathbf{Caut}, \mathbf{WMon}, \mathbf{Mon}, \mathbf{SMon}, \mathbf{SemConv}, \mathbf{SemWb}\}$ allow for consistent **Bc**-learning, that is, for $\beta \in \{\mathbf{G}, \mathbf{Psd}, \mathbf{Sd}\}$, we have

$$[\tau(\mathbf{Cons})\mathbf{Txt}\beta\delta\mathbf{Bc}] = [\mathbf{Txt}\beta\delta\mathbf{Bc}].$$

◀

4.3 The Importance of the Order

For **Bc**-learning, we have seen plenty cases where the order, which the information comes in, is not important. In this section we provide a natural example where it is important! In particular, we show this for monotone behaviourally correct learning. The idea is to construct a class of languages where the learner must keep track of the order the elements were presented in, in order to deduce possible previous hypotheses and to be able to safely discard particular elements at a later point in learning-time. To obtain this result, we apply the technique of self-learning classes [CK16a] using the Operator Recursion Theorem [Cas74].

▶ **Theorem 4.12.** We have

$$[\mathbf{TxtGMonBc}] \setminus [\mathbf{TxtPsdMonBc}] \neq \emptyset.$$

◀

Proof. We provide a class witnessing the separation using self-learning classes, as seen in Case and Kötzing [CK16a, Thm. 3.6]. Consider the learner which for a finite sequence $\sigma \in \mathbb{N}^*$ is defined as

$$h(\sigma) = \begin{cases} \mathrm{ind}(\emptyset), & \text{if } \mathrm{content}(\sigma) = \emptyset, \\ \varphi_{\max(\mathrm{content}(\sigma))}(\sigma), & \text{otherwise.} \end{cases}$$

We show a stronger version, namely the separation from monotone explanatory learners. Let $\mathcal{L} = \mathbf{TxtGMonEx}(h)$. Assume there exists a **TxtPsdMonBc**-learner $h' \in \mathcal{R}$ which learns $\mathcal{L}$, that is, $\mathcal{L} \subseteq \mathbf{TxtPsdMonBc}(h')$. By the Operator Recursion Theorem (**ORT**, Case [Cas74]), there exists a family of strictly monotone increasing, total computable functions $(a_j)_{j\in\mathbb{N}}$ with pairwise disjoint range, a total computable function $f \in \mathcal{R}$, an index $e \in \mathbb{N}$ and two families of indices $(e_j)_{j\in\mathbb{N}}$, $(\hat{e}_k)_{k\in\mathbb{N}}$ such that for all finite sequences $\sigma \in \mathbb{N}^*$, where $\mathrm{first}(\sigma)$ is the first non-pause element in the sequence σ, we have

$$\varphi_{a_j(i)}(\sigma) = \begin{cases} e_j, & \text{if } \mathrm{content}(\sigma) \subseteq \mathrm{range}(a_j), \\ \hat{e}_k, & \text{else, if } \exists k\colon a_k(f(k)) \in \mathrm{content}(\sigma) \vee \\ & \exists k\colon \mathrm{first}(\sigma) \in \mathrm{range}(a_k) \wedge \\ & \qquad \wedge \max\{j \mid \mathrm{content}(\sigma) \cap \mathrm{range}(a_j) \neq \emptyset\} = k, \\ e, & \text{otherwise.} \end{cases}$$

$$\begin{aligned} f(j) &\ldots \text{returns the first } i \text{ found such that } a_j(i) \in W_{h'(\mathrm{content}(a_j[i]),i)}, \\ W_{e_j} &= \mathrm{range}(a_j), \\ W_{\hat{e}_k} &= \bigcup_{j' \leq k} \mathrm{content}(a_{j'}[f(j')]) \cup \{a_k(f(k))\}, \\ W_e &= \bigcup_j \mathrm{content}(a_j[f(j)]). \end{aligned}$$

Let $\mathcal{L}' = \{W_{e_j} \mid j \in \mathbb{N}\} \cup \{W_{\hat{e}_k} \mid k \in \mathbb{N}_{>0}\} \cup \{W_e\}$. A depiction of the class $\mathcal{L}'$ can be seen in Figure 4.1. We show that $\mathcal{L}'$ can be learned by h, but not by h', that is, $\mathcal{L}' \subseteq \mathcal{L} = \mathbf{TxtGMonEx}(h)$ but also $\mathcal{L}' \not\subseteq \mathbf{TxtPsdMonBc}(h')$. The intuition is the following. For some $j \in \mathbb{N}$, as long as only elements from W_{e_j} are presented, h will suggest e_j as its hypothesis. Thus, h' needs to learn W_{e_j} as well and eventually overgeneralise, that is, at some point $i \in \mathbb{N}$ we have $\mathrm{content}(a_j[i]) \subsetneq W_{h'(\mathrm{content}(a_j[i]),i)}$. The function $f(j)$ finds such i. Once the

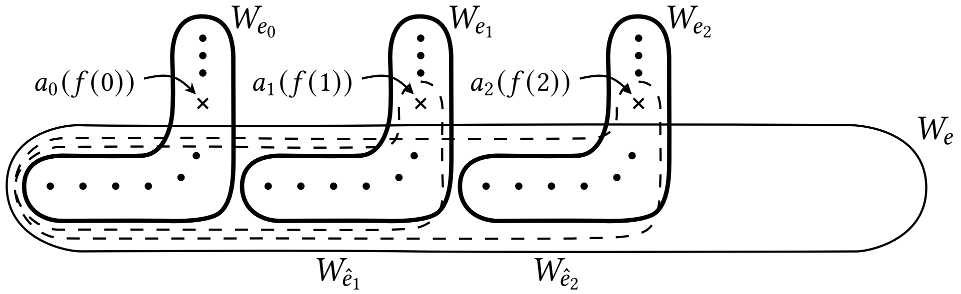


Figure 4.1: A depiction of the class $\mathcal{L}'$. Given $j \in \mathbb{N}_{>0}$, the dashed line depicts the set $W_{\hat{e}_j}$ and the cross indicates the element $a_j(f(j))$.

overgeneralisation happens, elements from, for $j' \neq j$, $\text{range}(a_{j'})$ may be presented. Knowing the order in which the elements were presented, the learner h now either keeps or discards the element $a_j(f(j))$ in its next hypothesis depending whether $j' < j$ or $j < j'$, respectively. If $j' < j$, h needs to keep $a_j(f(j))$ in its hypothesis as it still may be presented the set $W_{\hat{e}_j}$. Otherwise, it suggests the set W_e, only changing its mind if it sees, for appropriate $i \in \mathbb{N}$, an element of the form $a_i(f(i))$. Then, h is certain to be presented $W_{\hat{e}_i}$. So the Gold-style learner h can deal with this new information and preserve monotonicity, while h' cannot, as it does not know which information came first.

We proceed with the formal proof that h **TxtGMonEx**-learns $\mathcal{L}'$. Let $L' \in \mathcal{L}'$ and $T' \in \mathbf{Txt}(L')$. We first show the **Ex**-convergence and the monotonicity afterwards. For the former, we distinguish the following cases.

1.C.: For some $j \in \mathbb{N}$, we have $L' = W_{e_j}$. Let $n_0 \in \mathbb{N}$ be such that we have $\text{content}(T'[n_0]) \neq \emptyset$. Then, for $n \geq n_0$, there exists some $i \in \mathbb{N}$ such that $a_j(i) = \max(\text{content}(T'[n]))$. Thus,

$$h(T'[n]) = \varphi_{\max(\text{content}(T'[n]))}(T'[n]) = \varphi_{a_j(i)}(T'[n]) = e_j.$$

Hence, h learns W_{e_j} correctly.

2.C.: We have $L' = W_e$. Let $n_0 \in \mathbb{N}$ be the minimal and let $k_0 \in \mathbb{N}$ be such that $\text{content}(T'[n_0]) \neq \emptyset$ and $\text{first}(T'[n_0]) \in \text{range}(a_{k_0})$. Let $n_1 \geq n_0$ be minimal such that there exists $k > k_0$ such that $\text{content}(T'[n_1])$ also contains elements from $\text{content}(a_k)$. Then, for $n > n_1$ we have that $h(T'[n]) = e$, as there exists no $j \in \mathbb{N}$ with $a_j(f(j)) \in \text{content}(T')$ and also

$$\max\{j \mid \text{content}(T'[n]) \cap \text{range}(a_j) \neq \emptyset\} \neq k_0.$$

Thus, h learns W_e correctly.

3.C.: For some $k \in \mathbb{N}_{>0}$ we have $L' = W_{\hat{e}_k}$. In this case, there exists $n_0 \in \mathbb{N}$ such that, for some $k' < k$, $\mathrm{range}(a_{k'}) \cap \mathrm{content}(T'[n_0]) \neq \emptyset$ and $a_k(f(k)) \in \mathrm{content}(T'[n_0])$. Then, for $n \geq n_0$, we have $h(T'[n]) = \hat{e}_k$. Therefore, h learns $W_{\hat{e}_k}$ correctly.

We show that the learning is monotone. Let $n \in \mathbb{N}$. As long as $\mathrm{content}(T'[n])$ is empty, h returns $\mathrm{ind}(\emptyset)$. Once $\mathrm{content}(T'[n])$ is not empty anymore and as long as $\mathrm{content}(T'[n])$ only contains elements from, for some $j \in \mathbb{N}$, $\mathrm{range}(a_j)$, the learner h outputs (a code for) the set W_{e_j}. Note that j is the index of the element $\mathrm{first}(T'[n])$, that is, $\mathrm{first}(T'[n]) \in \mathrm{range}(a_j)$. If ever, for some later n, $\mathrm{content}(T'[n]) \setminus \mathrm{range}(a_j) \neq \emptyset$, then h only changes its mind if there exists $k > j$ such that $\mathrm{content}(T'[n]) \cap \mathrm{range}(a_k) \neq \emptyset$ (note that in case $j < k$, h does not change its mind). Depending on whether $a_k(f(k)) \in \mathrm{content}(T'[n])$ or not, h changes its mind to (a code of) either $W_{\hat{e}_k}$ or W_e, respectively. In the former case, the learner h is surely presented the set $W_{\hat{e}_k}$, making this mind change monotone. In the latter case, no element of $W_{e_j} \setminus \mathrm{content}(a_j[f(j)])$ is contained the target language. These are exactly the elements h discards from its hypothesis, keeping a monotone behaviour. The learner only changes its mind again if it witnesses, for some $k' \geq k$, the element $a_{k'}(f(k'))$. It will then output (a code of) the set $W_{\hat{e}_{k'}}$. This is, again, monotonic behaviour, as h is sure to be presented the set $W_{\hat{e}_{k'}}$. Altogether, h is monotone on any text of L'.

Thus, h identifies all languages in $\mathcal{L}'$ correctly. Now, we show that h' cannot do so too. We do so by providing a text of W_e where h' makes infinitely many wrong guesses. To that end, consider the text T of W_e given as

$$a_0[f(0)]a_1[f(1)]a_2[f(2)]\ldots$$

For $j \in \mathbb{N}_{>0}$, since $a_j(f(j)) \in W_{h'(\mathrm{content}(a_j[f(j)]), f(j))}$, we have

$$a_j(f(j)) \in W_{h'(\mathrm{content}(T[\sum_{m \leq j} f(m)]), \sum_{m \leq j} f(m))},$$

as $T[\sum_{m \leq j} f(m)]$ is an initial sequence for a text for $W_{\hat{e}_j}$. But, as $a_j(f(j)) \notin W_e$, the learner h' makes infinitely many incorrect conjectures and thus does not identify W_e on the text T correctly, a contradiction. ■

4.4 Completing the Partially Set-Driven Map

We recall the current situation. Corollary 4.8 already shows that semantically conservative, semantically witness-based, weakly monotone and (various variations of) cautious partially set-driven learning coincide with unrestricted set-driven learning. However, these restrictions are known to limit partially set-driven learning [Ful90; Jai+99; KSS17]. Furthermore, they are known to be incomparable to monotone learning [Jai+99], while both are more powerful than strongly monotone learning [Jai+99]. Non-U-shaped learning separates from the mentioned restrictions [Jai+99; KSS17]. However, non-U-shaped learning is a restriction for Gold-style learners [Bal+08; FJO94] and, equivalently, for partially set-driven learners [Car+06; Ful90; KR88].

We complete this map by showing that monotone learning implies non-U-shaped learning, see Theorem 4.13. In the simulation argument, we create a new hypothesis by adding all information obtainable by some future hypothesis generated from the seen elements. If this generates a correct hypothesis, no future hypotheses may be wrong, as otherwise the current hypothesis must contain further elements which are not in the language as well.

▶ **Theorem 4.13.** We have

$$[\mathbf{TxtPsdMonBc}] \subseteq [\mathbf{TxtPsdNUBc}].$$

◀

Proof. Let $h \in \mathcal{R}$ be a **TxtPsdMonBc**-learner. Note that h is, without loss of generality, strongly **Bc**-locking [KSS17]. Let furthermore $\mathcal{L} = \mathbf{TxtPsdMonBc}(h)$. We provide a **TxtPsdNUBc**-learner $h' \in \mathcal{R}$ which learns $\mathcal{L}$. For all finite sets $D \subseteq \mathbb{N}$, all $t \in \mathbb{N}_{\geq |D|}$ and all $s \in \mathbb{N}_{>0}$, define

$$W^0_{h'(D,t)} = D,$$
$$W^s_{h'(D,t)} = \bigcup_{\substack{(D',t') \text{ with} \\ (D,t) \preceq (D',t') \preceq (W^{s-1}_{h'(D,t)}, t+s)}} W^{s-1}_{h(D',t')}.$$

Finally, $W_{h'(D,t)} = \bigcup_{s \in \mathbb{N}} W^s_{h'(D,t)}$. Intuitively, the learner h' produces its hypothesis on (D, t) iteratively. At stage $s \in \mathbb{N}$, $W^s_{h'(D,t)}$ enumerates all elements witnessed by the learner h on some hypothesis extending (D, t) using elements witnessed so far, that is, elements in $W^{s-1}_{h'(D,t)}$.

We show that h' **TxtPsdNUBc**-learns $\mathcal{L}$. Let $L \in \mathcal{L}$ and $T \in \mathbf{Txt}(L)$. We provide a proof in two parts.

1. We first show that there exists an $n_0 \in \mathbb{N}$ such that $W_{h'(\mathrm{content}(T[n_0]),n_0)} = L$.
2. Afterwards, we show that, for all $n \in \mathbb{N}$, whenever $W_{h'(\mathrm{content}(T[n]),n)} = L$ we have, for all $n' > n$, also $W_{h'(\mathrm{content}(T[n']),n')} = L$.

For the first part, let $n_0 \in \mathbb{N}$ be such that $(D, t) := (\mathrm{content}(T[n_0]), n_0)$ is a **Bc**-locking information for h on L. Then, by definition of h', we have $W_{h'(D,t)} \supseteq W_{h(D,t)} = L$. For the other direction, we show that for all $s \in \mathbb{N}$ we have $W^s_{h'(D,t)} \subseteq L$ by induction on s. We get the statement for $s = 0$ immediately. Assuming it holds for $s \in \mathbb{N}$, we show it for $s + 1$. Since $W^s_{h'(D,t)} \subseteq L$, we have that (D', t') with $(D, t) \leq (D', t') \leq (W^s_{h'(D,t)}, t + s + 1)$ is also a **Bc**-locking information for h on L. In particular, we have $W_{h(D',t')} = L$. This results in

$$W^{s+1}_{h'(D,t)} = \bigcup_{\substack{(D',t') \text{ with} \\ (D,t)\leq(D',t')\leq(W^s_{h'(D,t)},t+s+1)}} W^s_{h(D',t')} \subseteq L.$$

This concludes the induction.

For the second part, let $n \in \mathbb{N}$ and $(D, t) := (\mathrm{content}(T[n]), n)$ be such that $W_{h'(D,t)} = L$, let $n' \geq n$ and $(D'', t'') := (\mathrm{content}(T[n']), n')$. Note that $D \subseteq D'' \subseteq L$ and $t'' \geq t$. We show that $W_{h'(D'',t'')} = L$. First, note that (D'', t'') will eventually be considered when enumerating $W_{h'(D,t)}$, that is, there exists an $s \in \mathbb{N}$ such that $(D'', t'') \leq (W^{s-1}_{h'(D,t)}, t + s)$. Hence,

$$\begin{aligned} W_{h'(D'',t'')} &= \bigcup_{s\in\mathbb{N}} \bigcup_{\substack{(D',t') \text{ with} \\ (D'',t'')\leq(D',t')\leq(W^{s-1}_{h'(D'',t'')},t+s)}} W^{s-1}_{h(D',t')} \subseteq \\ &\subseteq \bigcup_{s\in\mathbb{N}} \bigcup_{\substack{(D',t') \text{ with} \\ (D,t)\leq(D',t')\leq(W^{s-1}_{h'(D,t)},t+s)}} W^{s-1}_{h(D',t')} = W_{h'(D,t)} = L. \end{aligned}$$

Secondly, we show that for each $x \in L = W_{h'(D,t)}$ we also have $x \in W_{h'(D'',t'')}$. We show (by induction) that for every $s \in \mathbb{N}$

$$W^s_{h'(D,t)} \subseteq W_{h'(D'',t'')}.$$

For $s = 0$ we have $W^s_{h'(D,t)} = D \subseteq D'' = W^0_{h'(D'',t'')} \subseteq W_{h'(D'',t'')}$. Let the statement be fulfilled until s. At step $s + 1$, we distinguish the following cases.

1.C.: If $W^s_{h'(D,t)} = W^{s+1}_{h'(D,t)}$, that is, no new element is enumerated, the statement of the induction step is true immediately.

2.C.: If $W^s_{h'(D,t)} \subsetneq W^{s+1}_{h'(D,t)}$, let $x \in W^{s+1}_{h'(D,t)} \setminus W^s_{h'(D,t)}$. Note that $x \in L$. Let $(\tilde{D}, \tilde{t})$, with $(D,t) \preceq (\tilde{D}, \tilde{t}) \preceq (W^s_{h'(D,t)}, t{+}s)$, be the information on which x was witnessed, that is, $x \in W_{h(\tilde{D},\tilde{t})}$. By $W^s_{h'(D,t)} \subseteq W_{h'(D'',t'')}$ (the induction assumption), there exists $s'' \in \mathbb{N}$ such that $(\tilde{D}, \tilde{t}) \preceq (W^{s''}_{h'(D'',t'')}, t'' + s'')$. Since h is monotone and $x \in L$, we have

$$x \in W_{h(W^{s''}_{h'(D'',t'')}, t''+s'')} \overset{\text{Def. of } h'}{\subseteq} W_{h'(D'',t'')}.$$

Altogether, we get the desired result. ■

5 Gold-Style Learning

What remains to be studied is Gold-style behaviourally correct learning. In particular, it remains to be shown how monotone and non-U-shaped learning interact. In Section 5.1, we first show that monotone learners may be assumed strongly **Bc***-locking. In particular, with this result, learning under* any *considered restriction may be done with strongly* **Bc***-locking learners. We use this result then in Section 5.2 to make monotone learners non-U-shaped. This concludes all three main maps. We note that both sections are based on Doskoč and Kötzing [DK22].*

The overall situation for Gold-style learning is basically analogous to the initial situation for partially set-driven learning as discussed in Section 4.4, see Corollary 4.8 as well as Baliga et al. [Bal+08], Fulk et al. [FJO94], Jain et al. [Jai+99] and Kötzing et al. [KSS17]. We complete the map by showing that monotone learning implies non-U-shaped learning, see Theorem 5.3.

We aim to employ a similar approach as for the partially set-driven case. To that end, we have to overcome two obstacles. Firstly, we show that monotone Gold-style learners are strongly **Bc**-locking, see Theorem 5.1. In particular, this shows that *all* restrictions studied in this thesis allow for strongly **Bc**-locking learning. Secondly, Gold-style learners infer from sequences, meaning that extensions considered at a certain step do not necessarily have to be considered in later steps (as opposed to partially set-driven learning). We circumvent this by also enumerating elements from previous guesses on which the learner shows a monotone behaviour, as they are likely part of the target language.

5.1 Strongly Bc-Locking Monotone Learning

To show that any Gold-style monotone behaviourally correct learner may be assumed strongly **Bc**-locking, we expand the strategy used in the proof of Theorem 4.1. The idea is to extend the search for **Bc**-locking sequences. However, this extended search may cause correct elements to be temporarily discarded. We carefully add such elements.

▶ **Theorem 5.1.** Any **TxtGMonBc**-learner may be assumed strongly **Bc**-locking without loss of generality. ◀

Proof. This proof is inspired by the proof based on private communication with Jain [Jai17] where the equality of Gold-style and partially set-driven learning is shown for certain restrictions, see also Theorem 4.1. Let $h \in \mathcal{R}$ be a learner and let $\mathcal{L} = \mathbf{TxtGMonBc}(h)$. We provide a strongly **Bc**-locking **TxtGMonBc**-learner h' for $\mathcal{L}$ as follows. For two finite sequences $\sigma, \sigma' \in \mathbb{N}^*$, define the auxiliary function $g \in \mathcal{R}$ as

$$W_{g(\sigma',\sigma)} = \bigcap_{\tau \in \mathrm{content}(\sigma)_\#^{\leq|\sigma|}} W_{h(\sigma'\tau)} \cap \bigcap_{\substack{\sigma'' \leq \sigma',\\ \sigma'' \in \mathrm{content}(\sigma')_\#^*}} \bigcup_{\tau'' \in \mathrm{content}(\sigma')_\#^*} W_{h(\sigma''\tau'')}.$$

Then, define the learner $h' \in \mathcal{R}$ on finite sequences $\sigma \in \mathbb{N}^*$ as

$$W_{h'(\sigma)} = \bigcup_{\sigma' \subseteq \sigma} W_{g(\sigma',\sigma)}.$$

The intuition is the following. With the function g, we search for minimal **Bc**-locking sequences, see the proof of Theorem 4.1. To ensure that g eventually only contains elements from the target language, we extend the left hand intersection to be based on σ. However, as σ contains more and more information, additional sequences are also considered in the right hand intersection. This may lead to already enumerated elements being discarded (even if they belong to a target language). To prevent this, we take the union over all possible $W_{g(\sigma',\sigma)}$.

We formally show that h' has the desired properties. First, we show that h' is monotone. Let $L \in \mathcal{L}$ and $\sigma_1, \sigma_2 \in L_\#^*$ with $\sigma_1 \subseteq \sigma_2$. We show that, for all $x \in \mathbb{N}$, it holds that

$$x \in W_{h'(\sigma_1)} \cap L \Rightarrow x \in W_{h'(\sigma_2)} \cap L.$$

As $x \in W_{h'(\sigma_1)}$, there exists $\sigma_1' \subseteq \sigma_1$ such that $x \in W_{g(\sigma_1',\sigma_1)}$, that is,

$$x \in \bigcap_{\tau \in \mathrm{content}(\sigma_1)_\#^{\leq|\sigma_1|}} W_{h(\sigma_1'\tau)} \cap \bigcap_{\substack{\sigma'' \leq \sigma_1',\\ \sigma'' \in \mathrm{content}(\sigma_1')_\#^*}} \bigcup_{\tau'' \in \mathrm{content}(\sigma_1')_\#^*} W_{h(\sigma''\tau'')}. \tag{5.1}$$

In particular, $x \in W_{h(\sigma_1')}$. We show that $x \in W_{g(\sigma_1',\sigma_2)}$. By monotonicity of h, we have

$$x \in \bigcap_{\tau \in \mathrm{content}(\sigma_2)_\#^{\leq|\sigma_2|}} W_{h(\sigma_1'\tau)}.$$

As the right hand intersection in Equation (5.1) (of which x is an element) does not depend on σ_1, we have that

$$x \in \bigcap_{\tau \in \mathrm{content}(\sigma_2)_\#^{\leq|\sigma_2|}} W_{h(\sigma_1'\tau)} \cap \bigcap_{\substack{\sigma'' \leq \sigma_1', \\ \sigma'' \in \mathrm{content}(\sigma_1')_\#^*}} \bigcup_{\tau'' \in \mathrm{content}(\sigma_1')_\#^*} W_{h(\sigma''\tau'')} = W_{g(\sigma_1',\sigma_2)}.$$

By definition of h' and since $\sigma_1' \subseteq \sigma_1 \subseteq \sigma_2$, we have

$$W_{g(\sigma_1',\sigma_2)} \subseteq \bigcup_{\sigma' \subseteq \sigma_2} W_{g(\sigma',\sigma_2)} = W_{h'(\sigma_2)}.$$

Thus, $x \in W_{h'(\sigma_2)} \cap L$.

We now show that h' is strongly **Bc**-locking (and thus also **Bc**-learns $\mathcal{L}$). Let $L \in \mathcal{L}$ and $T \in \mathbf{Txt}(L)$. Let $\sigma_0 \in L_\#^*$ be the $\leq$-minimal **Bc**-locking sequence for h on L [BB75]. For each $\sigma' < \sigma_0$ with $\mathrm{content}(\sigma') \subseteq L$, let $\tau_{\sigma'} \in L_\#^*$ be such that $\sigma'\tau_{\sigma'}$ is a **Bc**-locking sequence for h on L [OSW86]. Let $n_0 \in \mathbb{N}$ be such that h converges on $T[n_0]$, that is, for all $n' \geq n_0$, $W_{h(T[n'])} = L$. Let $n_1 \geq n_0$ be such that

- $\sigma_0 \leq T[n_1]$,
- $\sigma_0 \in \mathrm{content}(T[n_1])_\#^*$, and
- for all $\sigma' < \sigma_0$ such that $\mathrm{content}(\sigma') \subseteq L$, we have that $\mathrm{content}(\sigma'\tau_{\sigma'}) \subseteq \mathrm{content}(T[n_1])$ and $|\tau_{\sigma'}| \leq n_1$.

To show that $\sigma_1 := T[n_1]$ is a **Bc**-locking sequence for h' on L, we show that, for any $\rho \in L_\#^*$, $h'(\sigma_1\rho)$ is a correct guess, that is, $W_{h'(\sigma_1\rho)} = L$. Let $\rho \in L_\#^*$. We prove each direction of $W_{h'(\sigma_1\rho)} = L$ separately.

1.C.: We show $W_{h'(\sigma_1\rho)} \subseteq L$. Let $x \in W_{h'(\sigma_1\rho)}$. Then there exists $\sigma' \subseteq \sigma_1\rho$ such that $x \in W_{g(\sigma',\sigma_1\rho)}$. In particular,

$$x \in \bigcap_{\tau \in \mathrm{content}(\sigma_1\rho)_\#^{\leq|\sigma_1\rho|}} W_{h(\sigma'\tau)} \cap \bigcap_{\substack{\sigma'' \leq \sigma', \\ \sigma'' \in \mathrm{content}(\sigma')_\#^*}} \bigcup_{\tau'' \in \mathrm{content}(\sigma')_\#^*} W_{h(\sigma''\tau'')}. \quad (5.2)$$

We distinguish based on the relation between σ' and σ_1.

1.1.C.: If $\sigma' \subseteq \sigma_1$, then there exists $\tau \in \mathrm{content}(\sigma_1\rho)_\#^{\leq|\sigma_1\rho|}$ such that $\sigma'\tau = \sigma_1$. As $h(\sigma_1)$ is a correct guess and $W_{h(\sigma_1)}$ is considered in the left hand intersection of Equation (5.2), we have that $x \in L$.

1.2.C.: If $\sigma' \supsetneq \sigma_1$, we have $\sigma_0 \leq \sigma_1 \subseteq \sigma'$ and $\sigma_0 \in \mathrm{content}(\sigma_1)^*_\# \subseteq \mathrm{content}(\sigma')^*_\#$. Thus, σ_0 is considered in the right hand intersection of Equation (5.2). Since, for any $\tau \in L^*_\#$, we have $W_{h(\sigma_0\tau)} = L$, we get $x \in W_{h(\sigma_0\tau)} = L$.

2.C.: We show $L \subseteq W_{h'(\sigma_1\rho)}$. Let $x \in L$. We show that $x \in W_{g(\sigma_1,\sigma_1\rho)}$. As h is monotone, $\sigma_1 \subseteq \sigma_1\rho$ and h converges on σ_1, we have

$$x \in \bigcap_{\tau \in \mathrm{content}(\sigma_1)^*_\#} W_{h(\sigma_1 ^\frown \tau)}.$$

Moreover, by choice of n_1, we have, for all $\sigma'' \leq \sigma_1$ with $\sigma'' \in \mathrm{content}(\sigma_1)^*_\#$, that $\tau''_{\sigma''} \in \mathrm{content}(\sigma_1)^*_\#$. As $\sigma''\tau''_{\sigma''}$ is a **Bc**-locking sequence for h on L, we get $x \in W_{h(\sigma''\tau''_{\sigma''})}$. Hence,

$$x \in \bigcap_{\substack{\sigma'' \leq \sigma_1, \\ \sigma'' \in \mathrm{content}(\sigma_1)^*_\#}} \bigcup_{\tau'' \in \mathrm{content}(\sigma_1)^*_\#} W_{h(\sigma''\tau'')}.$$

Altogether, $x \in W_{g(\sigma_1,\sigma_1\rho)} \subseteq W_{h'(\sigma_1\rho)}$.

In the end, we have $W_{h'(\sigma_1\rho)} = L$, which concludes the proof. ■

The literature [KSS17] and Corollary 4.9 imply the following corollary.

▶ **Corollary 5.2.** For any restriction $\delta \in \{$**T**, **NU**, **Caut**$_{\mathbf{Tar}}$, **Caut**$_\infty$, **Caut**$_{\mathbf{Fin}}$, **Caut**, **WMon**, **Mon**, **SMon**, **SemConv**, **SemWb**$\}$ and interaction operator $\beta \in \{$**G**, **Psd**, **Sd**$\}$, we have that any **Txt**$\beta\delta$**Bc**-learner may be assumed strongly **Bc**-locking without loss of generality. ◀

5.2 Completing the Gold-Style Map

Lastly, it remains to be shown that Gold-style monotone behaviourally correct learning implies non-U-shaped learning. With Theorem 5.1, we may assume the monotone learner to be strongly **Bc**-locking and, thus, we may employ a similar strategy as in the partially set-driven case, see Theorem 4.13. There, we enumerate the output of *all* possible future hypotheses. However, Gold-style learners infer their conjectures from sequences, meaning that extensions considered at a certain step do not necessarily have to be considered in later steps. To circumvent this issue, we also enumerate elements from previous guesses on which the learner shows a monotone behaviour. These elements are likely part of the target language and, if not, are safe to be discarded eventually.

▸ **Theorem 5.3.** We have

$$[\mathbf{TxtGMonBc}] \subseteq [\mathbf{TxtGNUBc}].$$

◂

Proof. Let $h \in \mathcal{R}$ be a **TxtGMonBc**-learner. Without loss of generality, h may be assumed strongly **Bc**-locking, see Theorem 5.1. Let $\mathcal{L} = \mathbf{TxtGMonBc}(h)$. We provide a learner $h' \in \mathcal{R}$ which **TxtGNUBc**-learns $\mathcal{L}$. Let a finite sequence $\sigma \in \mathbb{N}^*$ be given. In each step $s \in \mathbb{N}$, we employ both a *forward enumeration strategy* (via the sets $F_{\sigma,s}$) as well as a *backward search strategy* (via the sets $B_{\sigma,s}$). The learner h' is then defined as

$$W_{h'(\sigma)} = \bigcup_{s\in\mathbb{N}} B_{\sigma,s} \cup F_{\sigma,s}.$$

We proceed by defining $F_{\sigma,s}$ (forward enumeration sets) and $B_{\sigma,s}$ (backwards search sets) formally. Let $F_{\sigma,0} = B_{\sigma,0} = \mathrm{content}(\sigma)$. Furthermore, let

$$F_{\sigma,s+1} = F_{\sigma,s} \cup \bigcup_{\tau\in(F_{\sigma,s}\cup B_{\sigma,s})_{\#}^{\leq s}} W^s_{h(\sigma\tau)}.$$

Intuitively, $F_{\sigma,s+1}$ contains all elements enumerated by some possible future guess, that is, for $\tau \in (F_{\sigma,s} \cup B_{\sigma,s})_{\#}^{\leq s}$, elements enumerated by $W_{h(\sigma\tau)}$. Note that this is a similar approach as in the **Psd**-case, see the proof of Theorem 4.13. However, as opposed to partially set-driven learning, this alone does not suffice. In particular, $F_{\sigma,s}$ may consider $\sigma^\frown\tau$ and $\sigma^\frown\tau'$, where $\tau \neq \tau'$, in its enumeration, but, for a hypothesis building on later information $\sigma' \supseteq \sigma$, $F_{\sigma',s}$ cannot consider both, as σ' cannot extend both $\sigma^\frown\tau$ and $\sigma^\frown\tau'$ at the same time. To circumvent this, we need the backwards search set $B_{\sigma,s}$.

To define $B_{\sigma,s}$, we introduce the following auxiliary predicate and function. Given a finite sequence $\rho \in \mathbb{N}^*$ and an element $x \in \mathbb{N}$, we define

$$\mathbf{MonBeh}(x, \rho, s, \sigma) \Leftrightarrow \forall \tau \in \mathrm{content}(\sigma)^{\leq s+|\sigma|} : x \in W^s_{h(\rho^\frown\tau)}.$$

Intuitively, $\mathbf{MonBeh}(x, \rho, s, \sigma)$ checks whether the learner h, starting on information ρ, exhibits a monotonic behaviour regarding the element x. We further introduce a function which gives us the newly enumerated element by some hypothesis. In particular, for a finite sequence $\sigma' \subseteq \sigma$, let $\tilde{x} := \mathbf{nextEl}(\sigma', \sigma, s)$ be the element enumerated next by $F_{\sigma',s}$ (and has not yet been dealt with). Furthermore, let $\tilde{\sigma} \supseteq \sigma'$ be the (minimal) sequence on which $\tilde{x}$ has been seen for the first time inside $F_{\sigma',s}$. We define the backwards search set as, for finite sequences $\sigma, \sigma'' \in \mathbb{N}^*$

with $\sigma'' \subseteq \sigma$ and $s \in \mathbb{N}$,

$$\begin{aligned} B_{\sigma,0,\sigma''} &= \text{content}(\sigma''), \\ B_{\sigma,s+1,\sigma''} &= B_{\sigma,s,\sigma''} \cup \begin{cases} \{\tilde{x}\}, & \text{for } \tilde{x} = \mathbf{nextEl}(\sigma'', \sigma, s) \text{ via } \tilde{\sigma} \text{ if } \mathbf{MonBeh}(\tilde{x}, \tilde{\sigma}, s, \sigma), \\ \emptyset, & \text{otherwise.} \end{cases} \\ B_{\sigma,s+1} &= B_{\sigma,s} \cup \bigcup_{\sigma''' \subsetneq \sigma} B_{\sigma,s,\sigma'''}. \end{aligned}$$

Note that $\bigcup_{s\in\mathbb{N}} B_{\sigma,s,\sigma''} \subseteq \bigcup_{s\in\mathbb{N}} F_{\sigma'',s}$. The idea behind the backwards search is based on the following observation. Given two sequences $\sigma' \subseteq \sigma$, let x be the first element enumerated by $F_{\sigma',s}$ (which is not in content(σ')). If x is an element of the target language, it will eventually be enumerated in $F_{\sigma,s}$ as well (as it has to appear in $W_{h(\sigma)}$ by monotonicity of h). However, this may not hold true for another element enumerated by $F_{\sigma',s}$, as it may use the information $\sigma'^\frown x$, which in general is no subsequence of σ, to obtain the element. With the backwards search, we check for such elements and enumerate them in case the learner h shows a monotonic behaviour regarding them.

We show that h' **TxtGNUBc**-learns $\mathcal{L}$. Let $L \in \mathcal{L}$ and $T \in \mathbf{Txt}(L)$. First, we show that there exists $n_1 \in \mathbb{N}$ such that $W_{h'(T[n_1])} = L$ and afterwards that, for each $n, n' \in \mathbb{N}$ with $n \leq n'$, if $W_{h'(T[n])} = L$, then $W_{h'(T[n'])} = L$. To that end, let $n_0 \in \mathbb{N}$ be such that $T[n_0]$ is a **Bc**-locking sequence for h on L (this exists by Theorem 5.1). For each $n < n_0$, let $\tilde{x}_n = \mathbf{nextEl}(T[n], T[n_0], s)$ (via $\tilde{\sigma}_n$) be the first newly enumerated element not in L (if such exists). Then, let $n_1 \geq n_0$ be such that, for all $n < n_0$, for each associated $\tilde{\sigma}_n$ there exists $\tau \in \text{content}(T[n_1])_{\#}^{\leq |T[n_1]|}$ such that $h(\tilde{\sigma}_n^\frown \tau)$ is a correct guess. Then, for all $s \in \mathbb{N}$, $\mathbf{MonBeh}(\tilde{x}, \tilde{\sigma}, s, T[n_1])$ fails. It follows that no $B_{T[n_1],s,T[n]}$ contains elements which are not in L, that is, for all $n < n_0$ and $s \in \mathbb{N}$ we have

$$B_{T[n_1],s,T[n]} \subseteq L.$$

Also, for $n \in \mathbb{N}$ with $n_0 \leq n \leq n_1$, $B_{T[n_1],s,T[n]}$ only contains elements in L (as $T[n_0]$ is a **Bc**-locking sequence). Hence, we have

$$\bigcup_{s\in\mathbb{N}} B_{T[n_1],s} \subseteq L.$$

In particular, as $T[n_1]$ is also a **Bc**-locking sequence, we get

$$\bigcup_{s\in\mathbb{N}} F_{T[n_1],s} = L.$$

Altogether, we get

$$W_{h'(T[n_1])} = \bigcup_{s\in\mathbb{N}} B_{T[n_1],s} \cup F_{T[n_1],s} = L.$$

It remains to be shown that, for each $n, n' \in \mathbb{N}$ with $n \leq n'$, if $W_{h'(T[n])} = L$, then $W_{h'(T[n'])} = L$. Let $n \in \mathbb{N}$ be minimal such that $W_{h'(T[n])} = L$. We show that, for $n' \geq n$, we have $W_{h'(T[n'])} = L$ as well. Note that by definition of the backwards search sets, for $\tilde{n} \leq n$, we have

$$\bigcup_{s\in\mathbb{N}} B_{T[n],s,T[\tilde{n}]} \supseteq \bigcup_{s\in\mathbb{N}} B_{T[n'],s,T[\tilde{n}]}.$$

Furthermore, we have

$$\bigcup_{s\in\mathbb{N}} F_{T[n],s} \supseteq \bigcup_{s\in\mathbb{N}} F_{T[n'],s} \cup \bigcup_{\substack{\tilde{n}\in\mathbb{N},\\ n\leq\tilde{n}\leq n'}} \bigcup_{s\in\mathbb{N}} B_{T[n'],s,T[\tilde{n}]},$$

as, $T[n']$ is, for a sufficiently large $s \in \mathbb{N}$, a candidate within $F_{T[n],s}$ and the backwards search set $\bigcup_{s\in\mathbb{N}} B_{T[n'],s,T[\tilde{n}]}$ can only enumerate as much as the forward enumeration set $\bigcup_{s\in\mathbb{N}} F_{T[\tilde{n}],s}$. Thus, $W_{h'(T[n'])} \subseteq W_{h'(T[n])} = L$. Next we show that each element $x \in W_{h'(T[n])} = L$ will be enumerated in $W_{h'(T[n'])}$. We show this by case distinction depending how x is enumerated in $W_{h'(T[n])}$.

1.C.: For some $s' \in \mathbb{N}$, the element x is enumerated in $F_{T[n],s'}$. Then, we get $x \in \bigcup_{s\in\mathbb{N}} B_{T[n'],s,T[n]} \big(\subseteq W_{h'(T[n'])}\big)$ as the **MonBeh** check passes for elements in L.

2.C.: For some $s' \in \mathbb{N}$ and $\tilde{n} \leq n$, we have $x \in B_{T[n],s',T[\tilde{n}]}$. Then, x is enumerated in $\bigcup_{s\in\mathbb{N}} B_{T[n'],s,T[\tilde{n}]} \big(\subseteq W_{h'(T[n'])}\big)$ as the **MonBeh** check passes for elements in L.

Thus, $W_{h'(T[n'])} \supseteq L$ and, altogether, $W_{h'(T[n'])} = L$. ■

6 Conclusion and Outlook

In this thesis, we investigate behaviourally correct learning in the formalisation of Gold [Gol67]. We extend the literature by studying the restrictions mentioned in Chapters 1 and 2 with regards to their relation and properties.

Firstly, we fix a mode of data presentation and study the relations of various restrictions with each other. This continues the work of Jain et al. [Jai+16], Kötzing and Palenta [KP16] and Kötzing and Schirneck [KS16], where this is done in the case of explanatory learning. We provide a full picture for common interaction operators, namely for Gold-style, partially set-driven and set-driven learning. We see that many results are similar to the respective explanatory map. Most notably, we show that monotone learning does imply non-U-shaped learning in the Gold-style and partially set-driven case. In the case of set-driven learning it does as well, as non-U-shaped learning is no restriction there. Furthermore, we show that semantically conservative, weakly monotone and cautious learning coincide for all the studied modes of data presentation.

We furthermore study the mentioned restrictions with regards to their properties. In particular, we focus on consistent and strongly **Bc**-locking learning. The former models the seemingly natural desire to include all data seen during learning correctly. The famous *inconsistency phenomenon* states that in the case of explanatory learning consistency is a restriction [Bār77]. Often, the opposite is observed in behaviourally correct learning, where many restrictions allow for consistent **Bc**-learning. We complete these partial results by showing that *all* considered restrictions allow for consistent **Bc**-learning. This way, we not only answer open questions from the literature [KSS17] but also provide a full picture.

Furthermore, we study whether learning can be done with strongly **Bc**-locking learners. This is more of technical interest as many simulation arguments rely on the search for **Bc**-locking data. Thus, it is comfortable to know whether they exist for the learner at hand. While Kötzing et al. [KSS17] provide many general results when this behaviour can be expected, we complete this work by including *all* considered restrictions. This way, we provide a full picture regarding this property.

While the thesis gives a rounded picture regarding the pairwise relation or properties of the studied restrictions, there are ways to continue this work. We discuss possibilities next.

Decisive Learning. The provided maps can be extended to contain further important restrictions. In particular, *decisive* learning [OSW82] has been studied in the context of map charting in the case of explanatory learning [Jai+16; KP16; KS16]. In decisive learning, the learners are not allowed to return to a (semantically) abandoned hypothesis. The importance of this restriction is also displayed by the fact that it is a natural specialisation to non-U-shaped learning while it is a generalisation to semantically witness-based learning. As such, it is known to be restrictive to partially set-driven and Gold-style learning [Bal+08]. In these cases, it is also strictly less powerful than non-U-shaped learning [Bal+08].

First, note that we extend these results by implicitly showing that decisive learning is no restriction in the set-driven case, see Theorem 3.9. However, interactions to many other restrictions elude us. In particular, it is open whether monotone learning implies decisive learning, as it is the case in the explanatory counterpart [KP16; KS16]. We state the following open problem.

▶ **Open Problem 6.1.** In the case of Gold-style and partially set-driven learning, does monotone behaviourally correct learning imply decisive behaviourally correct learning? ◀

To get a better understanding of the problem, one could consider variants of decisive learning [Bal+08; Car+06], similarly to the variations of cautious learning, see Kötzing and Palenta [KP16] or Sections 3.1 and 4.1. Baliga et al. [Bal+08], for example, consider *second-time decisive* learning, where the learners may abandon a hypothesis at most once. They show that it is no restriction in explanatory learning. Similarly, Carlucci et al. [Car+06] consider decisiveness on *wrong*, *overinclusive* and *overgeneralising* conjectures, where the learner may not return to abandoned wrong, incomparable or overgeneralising conjectures, respectively. They investigate the relation between these and show that, in particular, neither of the latter two is restrictive. By studying the mentioned variants further, one could gain valuable insights in the behaviour of decisive learners.

The Presentation of the Data. Another interesting direction yet to be studied is the interplay between the maps. Especially in behaviourally correct learning, we have seen that Gold-style learning often coincides with partially set-driven learning. However, in Section 4.3, we provide a first natural example of where this is not the case. This raises the question whether there are similarly natural examples where partially set-driven learning does not suffice to obtain the full learning power. We state the question for non-U-shaped and decisive learning.

▶ **Open Problem 6.2.** Does it hold, for decisive and non-U-shaped learning, that Gold-style and partially set-driven learning are equally powerful? ◀

Once this question is solved, the next step is to consider the interplay between the restrictions with various forms of data presentation. This combines both the study on the impact of the data presentation and the impact of the restrictions.

Simulation Arguments. Another direction of research lies in the thorough analysis of different simulation arguments. This would nicely complement the rich literature on separation arguments, where Case and Kötzing [CK16a] provide general results using self-learning classes. In this work, we study the weak and strong forward verification, see Section 3.1, and obtain interesting properties of the learners and their output. However, such studies may be applied to other strategies used throughout the literature or this work. For example, the simulation strategy discussed with Sanjay Jain in a private communication [Jai17] occurs in multiple proofs, see Theorems 4.1 and 5.1. Thus, it may be fruitful to study such approaches and extensions thereof to obtain properties of learning.

Bibliography

[Ang80] Dana Angluin. **Inductive Inference of Formal Languages from Positive Data**. *Information and Control* 45 (1980), 117–135 (see pages 2, 3, 9).

[Bal+08] Ganesh Baliga, John Case, Wolfgang Merkle, Frank Stephan and Rolf Wiehagen. **When unlearning helps**. *Information and Computation* 206 (2008), 694–709 (see pages 2–4, 9, 11, 39, 53, 57, 66).

[Bār77] Jānis M. Bārzdiņš. **Inductive Inference of Automata, Functions and Programs**. In: *American Mathematical Society Translations*. 1977, 107–122 (see pages 3, 5, 47, 65).

[BB75] Lenore Blum and Manuel Blum. **Toward a Mathematical Theory of Inductive Inference**. *Information and Control* 28 (1975), 125–155 (see pages 2, 6, 8, 12, 42, 59).

[Car+06] Lorenzo Carlucci, Sanjay Jain, Efim B. Kinber and Frank Stephan. **Variations on U-shaped learning**. *Information and Computation* 204 (2006), 1264–1294 (see pages 5, 10, 11, 19, 39, 40, 47, 53, 66).

[Cas74] John Case. **Periodicity in Generations of Automata**. *Mathematical Systems Theory* 8 (1974), 15–32 (see pages 49, 50).

[CC13] Lorenzo Carlucci and John Case. **On the Necessity of U-Shaped Learning**. *Top. Cogn. Sci.* 5:1 (2013), 56–88 (see page 3).

[CK10] John Case and Timo Kötzing. **Solutions to Open Questions for Non-U-Shaped Learning with Memory Limitations**. In: *Algorithmic Learning Theory (ALT)*. Vol. 6331. Lecture Notes in Computer Science. Springer, 2010, 285–299 (see page 3).

[CK16a] John Case and Timo Kötzing. **Strongly non-U-shaped language learning results by general techniques**. *Information and Computation* 251 (2016), 1–15 (see pages 3, 6, 49, 50, 67).

[CK16b] John Case and Timo Kötzing. **Topological separations in inductive inference**. *Theoretical Computer Science* 620 (2016), 33–45 (see page 11).

[CL82] John Case and Christopher Lynes. **Machine Inductive Inference and Language Identification**. In: *International Colloquium on Automata, Languages and Programming (ICALP)*. 1982, 107–115 (see pages 1, 9).

[CM08] John Case and Samuel E. Moelius. **U-shaped, iterative, and iterative-with-counter learning**. *Machine Learning* 72:1-2 (2008), 63–88 (see page 3).

[DK20] Vanja Doskoč and Timo Kötzing. **Cautious Limit Learning**. In: *Algorithmic Learning Theory (ALT)*. 2020, 117:251–117:276 (see pages 15, 39, 40, 45).

[DK21a] Vanja Doskoč and Timo Kötzing. **Mapping Monotonic Restrictions in Inductive Inference**. In: *Computability in Europe (CiE)*. 2021, 12813:146–12813:157 (see page 39).

[DK21b] Vanja Doskoč and Timo Kötzing. **Normal Forms for Semantically Witness-Based Learners in Inductive Inference**. In: *Computability in Europe (CiE)*. 2021, 12813:158–12813:168 (see page 15).

[DK22] Vanja Doskoč and Timo Kötzing. **Maps of Restrictions for Behaviourally Correct Learning**. In: *Computability in Europe (CiE)*. 2022, 103–114 (see pages 15, 39, 57).

[FJO94] Mark A. Fulk, Sanjay Jain and Daniel N. Osherson. **Open Problems in "Systems That Learn"**. *J. Comput. Syst. Sci.* 49:3 (1994), 589–604 (see pages 3, 4, 39, 53, 57).

[Ful90] Mark A. Fulk. **Prudence and Other Conditions on Formal Language Learning**. *Information and Computation* 85 (1990), 1–11 (see pages 10, 39, 40, 53).

[Gol67] E. Mark Gold. **Language Identification in the Limit**. *Information and Control* 10 (1967), 447–474 (see pages iii, v, 1–3, 8, 9, 65).

[Jai+16] Sanjay Jain, Timo Kötzing, Junqi Ma and Frank Stephan. **On the role of update constraints and text-types in iterative learning**. *Information and Computation* 247 (2016), 152–168 (see pages 3, 4, 6, 65, 66).

[Jai+99] Sanjay Jain, Daniel N. Osherson, James S. Royer and Arun Sharma. **Systems that Learn: An Introduction to Learning Theory**. MIT Press, Cambridge (MA), Second Edition, 1999 (see pages 4, 6, 11, 12, 15, 39, 53, 57).

[Jai17] Sanjay Jain. *Personal Communication*. 2017 (see pages 41, 58, 67).

[Jan91] Klaus Jantke. **Monotonic and non-monotonic inductive inference**. *New Generation Computing* 8 (1991), 349–360 (see pages 2, 9).

[Köt09] Timo Kötzing. **Abstraction and Complexity in Computational Learning in the Limit**. PhD thesis. University of Delaware, 2009 (see page 8).

[Köt17] Timo Kötzing. **A Solution to Wiehagen's Thesis**. *Theory of Computing Systems* 60 (2017), 498–520 (see page 12).

[KP16] Timo Kötzing and Raphaela Palenta. **A map of update constraints in inductive inference**. *Theoretical Computer Science* 650 (2016), 4–24 (see pages 2–6, 9, 12, 43, 44, 65, 66).

[KR88] Stuart A. Kurtz and James S. Royer. **Prudence in Language Learning**. In: *Computational Learning Theory (COLT)*. 1988, 206–219 (see pages 10, 11, 39, 40, 53).

[KS16] Timo Kötzing and Martin Schirneck. **Towards an Atlas of Computational Learning Theory**. In: *Symposium on Theoretical Aspects of Computer Science (STACS)*. 2016, 47:1–47:13 (see pages 3, 4, 6, 12, 65, 66).

[KSS17] Timo Kötzing, Martin Schirneck and Karen Seidel. **Normal Forms in Semantic Language Identification**. In: *Algorithmic Learning Theory (ALT)*. 2017, 76:493–76:516 (see pages 3–6, 9–13, 15, 17, 18, 21, 24, 26, 29, 30, 33, 39–41, 43, 44, 46, 47, 49, 53, 57, 60, 65).

[LZ93] Steffen Lange and Thomas Zeugmann. **Monotonic versus non-monotonic language learning**. In: *Nonmonotonic and Inductive Logic*. 1993, 254–269 (see pages 2, 9).

[Mar+92] Gary F. Marcus, Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, Fei Xu and Harald Clahsen. **Overregularization in Language Acquisition**. *Monographs of the Society for Research in Child Development* 57:4 (1992), i–178 (see page 2).

[MM69] G.A. Miller and D. McNeill. **Psycholinguitics: handbook of social psychology**. Addisen, 1969 (see page 2).

[OSW82] Daniel N. Osherson, Michael Stob and Scott Weinstein. **Learning Strategies**. *Information and Control* 53 (1982), 32–51 (see pages 2, 9, 66).

[OSW86] Daniel N. Osherson, Michael Stob and Scott Weinstein. **Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists**. MIT Press, Cambridge (MA), 1986 (see pages 42, 59).

[OW82] Daniel N. Osherson and Scott Weinstein. **Criteria of Language Learning**. *Information and Control* 52 (1982), 123–138 (see pages 1, 9).

[RB99] W.C. Ritchie and T.K. Bhatia. **Handbook of Child Language Acquisition**. Academic Press, 1999 (see page 2).

[Rog87] Hartley Rogers Jr. **Theory of recursive functions and effective computability**. Reprinted by MIT Press, Cambridge (MA), 1987 (see page 7).

[Sch84] Gisela Schäfer-Richter. **Über Eingabeabhängigkeit und Komplexität von Inferenzstrategien**. PhD thesis. RWTH Aachen University, Germany, 1984 (see pages 2, 8).

[WC80] Kenneth Wexler and Peter W. Culicover. **Formal principles of language acquisition**. *MIT Press, Cambridge (MA)* (1980) (see pages 2, 8).

[Wie91] Rolf Wiehagen. **A thesis in inductive inference**. In: *Nonmonotonic and Inductive Logic*. 1991, 184–207 (see pages 2, 9).

List of Publications

Articles in Refereed Conference Proceedings

[1] **Confident Iterative Learning in Computational Learning Theory**. In: *Current Trends in Theory and Practice of Computer Science (SOFSEM)*. 2018, 30–42.

[2] **Towards Explainable Real Estate Valuation via Evolutionary Algorithms**. In: *Genetic and Evolutionary Computation Conference (GECCO)*. 2022, 1130–1138. Joint work with Sebastian Angrick, Ben Bals, Niko Hastrich, Maximilian Kleissl, Jonas Schmidt, Maximilian Katzmann, Louise Molitor and Tobias Friedrich.

[3] **Learning Languages with Decidable Hypotheses**. In: *Computability in Europe (CiE)*. Vol. 12813. 2021, 25–37. Joint work with Julian Berger, Maximilian Böther, Jonathan Gadea Harder, Nicolas Klodt, Timo Kötzing, Winfried Lötzsch, Jannik Peters, Leon Schiller, Lars Seifert, Armin Wells and Simon Wietheger.

[4] **Non-Monotone Submodular Maximization with Multiple Knapsacks in Static and Dynamic Settings**. In: *European Conference on Artificial Intelligence (ECAI)*. Vol. 325. 2020, 435–442. Joint work with Tobias Friedrich, Andreas Göbel, Aneta Neumann, Frank Neumann and Francesco Quinzan.

[5] **Optical Character Recognition Guided Image Super Resolution**. In: *Symposium on Document Engineering (DocEng)*. 2022. Joint work with Philipp Hildebrandt, Maximilian Schulze, Sarel Cohen, Raid Saabni and Tobias Friedrich.

[6] **Cautious Limit Learning**. In: *Algorithmic Learning Theory (ALT)*. Vol. 117. 2020, 251–276. Joint work with Timo Kötzing.

[7] **Mapping Monotonic Restrictions in Inductive Inference**. In: *Computability in Europe (CiE)*. Vol. 12813. 2021, 146–157. Joint work with Timo Kötzing.

[8] **Normal Forms for Semantically Witness-Based Learners in Inductive Inference**. In: *Computability in Europe (CiE)*. Vol. 12813. 2021, 158–168. Joint work with Timo Kötzing.

[9] **Maps of Restrictions for Behaviourally Correct Learning**. In: *Computability in Europe (CiE)*. Vol. 13359. 2022, 103–114. Joint work with Timo Kötzing.

[10] **Adaptive Sampling for Fast Constrained Maximization of Submodular Functions**. In: *Artificial Intelligence and Statistics (AISTATS)*. Vol. 130. 2021, 964–972. Joint work with Francesco Quinzan, Andreas Göbel and Tobias Friedrich.

Contributions to Refereed Workshops

[11] **Fine-Grained Localization, Classification and Segmentation of Lungs with Various Diseases**. In: *CVPR Workshop on Fine-Grained Visual Categorization (FGVC@CVPR)*. 2021. Joint work with Julian Berger, Tibor Bleidt, Martin Büßemeyer, Marcus Ding, Moritz Feldmann, Moritz Feuerpfeil, Janusch Jacoby, Valentin Schröter, Bjarne Sievers, Moritz Spranger, Simon Stadlinger, Paul Wullenweber, Sarel Cohen and Tobias Friedrich.

[12] **Drug Repurposing for Multiple COVID Strains using Collaborative Filtering**. In: *ICLR Workshop on Machine Learning for Preventing and Combating Pandemics (MLPCP@ICLR)*. 2021. Joint work with Otto Kißig, Martin Taraz, Sarel Cohen and Tobias Friedrich.