Normal Forms for Semantically Witness-Based Learners in Inductive Inference*

Vanja Doskoč and Timo Kötzing

Hasso Plattner Institute, University of Potsdam, Germany
{vanja.doskoc, timo.koetzing}@hpi.de

Abstract. In *inductive inference*, we study learners (computable devices) inferring formal languages. In particular, we consider *semantically witness-based* learners, that is, learners which are required to *justify* each of their semantic mind changes. This natural requirement deserves special attention as it is a specialization of various important learning paradigms. As such, it has already proven to be fruitful for gaining knowledge about other types of restrictions.
In this paper, we provide a thorough analysis of semantically converging, semantically witness-based learners, obtaining *normal forms* for them. Most notably, we show that *set-driven globally* semantically witness-based learners are equally powerful as their *Gold-style semantically conservative* counterpart. Such results are key to understanding the, yet undiscovered, mutual relation between various important learning paradigms of semantically converging learners.

1 Introduction

Computably learning formal languages from a growing but finite amount of information thereof is referred to as *inductive inference* or *language learning in the limit* [5], a branch of (algorithmic) learning theory. Here, a learner h (a computable device) is successively presented all and only the information from a formal language L (a computably enumerable subset of the natural numbers). We call such a list of elements of L a *text* of L. With every new datum, the learner h makes a guess (a description for a c.e. set) about which language it believes to be presented. Once these guesses converge to a single, correct hypothesis explaining the language, the learner successfully *learned* the language L on this text. We say that h *learns* L if it learns L on every text of L.

We refer to this as *explanatory learning* as the learner, in the limit, provides an explanation of the presented language. If we drop the requirement to converge to a *single* correct hypothesis and allow the learner to oscillate between arbitrarily many correct ones, we refer to this as *behaviourally correct* learning [3,13] and denote it as[1] $\mathbf{TxtGBc}$. Since a learner which always guesses the same language can learn this very language, we study classes of languages which can be $\mathbf{TxtGBc}$-learned by a single learner. We denote the set of all such classes with $[\mathbf{TxtGBc}]$, which we refer to as the *learning power* of $\mathbf{TxtGBc}$-learners.

* This work was supported by DFG Grant Number KO 4635/1-1.

[1] Here, $\mathbf{Txt}$ indicates that the information is given from text, $\mathbf{G}$ stands for *Gold-style* learning, where the learner has *full information* on the elements presented to make its guess, and, lastly, $\mathbf{Bc}$ refers to behaviourally correct learning.

Additional restrictions, modelling various learning strategies, may be imposed on the learner. By studying these we discover how seemingly intuitive strategies affect the learning power. For example, it may seem evident to only make semantic mind changes when seeing a new datum *justifying* this mind change. However, it is known that following such a strategy, referred to as *semantically witness-based* learning ($\mathbf{SemWb}$, [9,10]), severely lessens the obtainable learning power.

Besides being intuitive yet restrictive, this restriction proved to be important in the literature. Together with *target-cautious* learning [8], where learners may not overgeneralize the target language, this paradigm encloses various important learning restrictions. Exemplary for explanatory learning, in settings where (syntactically) witness-based learning, as *specialization* or lower bound, and target-cautious learning, as *generalization* or upper bound, permit equivalent learning power, the three enclosed but seemingly incomparable restrictions, namely *conservativeness* [1], *weak monotonicity* [7,17] and *cautiousness* [12], are equivalent as well [9].

The still undiscovered mutual relation between the mentioned restrictions in the behaviourally correct setting makes it worthwhile to study semantically witness-based learning in this setting as well. The previous literature indicates analogous equalities to be possible. Particularly, semantically witness-based learners and, a generalization thereof, *semantically conservative* learners ($\mathbf{SemConv}$, [10]), which keep their guesses while they are consistent with the data given, are shown to be equally powerful [10]. This equality holds true regardless of the amount of information given, particularly, it holds true for both Gold-style and *set-driven* learners ($\mathbf{Sd}$, [16]), which base their hypotheses solely on the set of elements given. We enhance the analogy by showing that the learners perform equally well regardless of the amount of information given, drawing parallels to target-cautious learning, where Gold-style and set-driven learners are also equally powerful [4].

The latter result already provides a powerful *normal form*. It states that semantically witness-based learners do not need to know the order and amount of the information given. This significantly extends the result [10] that such set-driven learners are as powerful as *partially set-driven* ones [2,15]. Note that the latter learners base their hypotheses on the content and amount of information given, but have no access to the order in which the information came. Another normal form shows that witness-based learners display such behaviour also *globally*, that is, on arbitrary text. This means that the learners always display a "decent" behaviour regardless whether the information given belongs to a language they actually learn. Lastly, semantically witness-based and semantically conservative learning is interchangeable also when required globally.

This paper is structured as follows. In Section 2 we introduce all necessary notation and preliminary results. In Section 3, we show that three normal forms can be assumed *simultaneously*. Our main result (Theorem 1) states that semantically conservative $\mathbf{G}$-learners may be assumed (a) globally (b) semantically witness-based and (c) set-driven.

2 Language Learning in the Limit

In this section we introduce notation and preliminary results used throughout this paper. Thereby, we consider basic computability theory as known [14]. We start with the mathematical notation and use $\subsetneq$ and $\subseteq$ to denote the proper subset and subset relation between sets, respectively. We denote the set of all natural numbers as $\mathbb{N} = \{0, 1, 2, \dots\}$ and the empty set as $\emptyset$. Furthermore, we let $\mathcal{P}$ and $\mathcal{R}$ be the set of all partial and total computable functions $p \colon \mathbb{N} \to \mathbb{N}$. Next, we fix an effective numbering $\{\varphi_e\}_{e \in \mathbb{N}}$ of all partial computable functions and denote the e-th computably enumerable set as $W_e = \mathrm{dom}(\varphi_e)$ and interpret the number e as an *index* or *hypothesis* thereof.

We learn recursively enumerable sets $L \subseteq \mathbb{N}$, called *languages*, using *learners*, that is, partial computable functions. By $\#$ we denote the *pause symbol* and for any set S we denote $S_\# := S \cup \{\#\}$. Then, a *text* is a total function $T \colon \mathbb{N} \to \mathbb{N} \cup \{\#\}$ and the collection of all texts is denoted as $\mathbf{Txt}$. For any text (or sequence) T we define the *content* of T as $\mathrm{content}(T) := \mathrm{range}(T) \setminus \{\#\}$. Here, range denotes the image of a function. A text of a language L is such that $\mathrm{content}(T) = L$. We denote the collection of all texts of L as $\mathbf{Txt}(L)$. Additionally, for $n \in \mathbb{N}$, we denote by $T[n]$ the initial sequence of T of length n, that is, $T[0] := \varepsilon$ (the empty string) and $T[n] := (T(0), T(1), \dots, T(n-1))$. For a set S, we call the sequence (text) where all elements of S are presented in strictly increasing order without interruptions (followed by infinitely many pause symbols if S is finite) the *canonical sequence (text) of S*. On finite sequences we use $\subseteq$ to denote the *extension relation*. Given two sequences σ and τ we write $\sigma^\frown\tau$ or (if readability permits) $\sigma\tau$ to denote the concatenation of these.

We use the following system to formalize learning criteria [11]. An *interaction operator* β takes a learner $h \in \mathcal{P}$ and a text $T \in \mathbf{Txt}$ as argument and outputs a possibly partial function p. Intuitively, β provides the information for the learner to make its guesses. We consider the interaction operators $\mathbf{G}$ for *Gold-style* or *full-information* learning [5] and $\mathbf{Sd}$ for *set-driven* learning [16]. Define, for any $i \in \mathbb{N}$,

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

Intuitively, a Gold-style learner has full information on the elements presented, while a set-driven learner bases its guesses solely on the content, that is, set of elements, given.

Given a learning task, we can distinguish between various criteria for successful learning. A first such criterion is *explanatory* learning ($\mathbf{Ex}$, [5]), where the learner is expected to converge to a single, correct hypothesis in order to learn a language. Allowing the learner to oscillate between arbitrarily many semantically correct, but possibly syntactically different hypotheses we get *behaviourally correct* learning ($\mathbf{Bc}$, [3,13]). Formally, a *learning restriction* δ is a predicate on a total learning sequence p, that is, a total function, and a text $T \in \mathbf{Txt}$. For the mentioned criteria we have

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

We impose restrictions on the learners. In particular, we focus on *semantically witness-based* learning ($\mathbf{SemWb}$, [9,10]), where the learners need to *justify* each of their semantic mind changes. A generalization thereof is *semantically conservative* learning ($\mathbf{SemConv}$, [1]). Here, the learners may not change their mind while their hypotheses are consistent with the information given. A hypothesis is consistent if it contains all the information it is based on and if we require the learners to output consistent hypotheses we speak of *consistent* learning ($\mathbf{Cons}$, [1]). Formally, we define the restrictions as

$$\begin{aligned}
\mathbf{SemWb}(p,T) &:\Leftrightarrow \forall n, m\colon (\exists k\colon n \leq k \leq m \wedge W_{p(n)} \neq W_{p(k)}) \Rightarrow \\
&\qquad\qquad\qquad \Rightarrow (\mathrm{content}(T[m]) \cap W_{p(m)}) \backslash W_{p(n)} \neq \emptyset, \\
\mathbf{Cons}(p,T) &:\Leftrightarrow \forall n\colon \mathrm{content}(T[n]) \subseteq W_{h(T[n])}, \\
\mathbf{SemConv}(p,T) &:\Leftrightarrow \forall n, m\colon \big(n < m \wedge \mathrm{content}(T[m]) \subseteq W_{p(n)}\big) \Rightarrow \\
&\qquad\qquad\qquad \Rightarrow W_{p(n)} = W_{p(m)}.
\end{aligned}$$

Given two restrictions δ and δ', the juxtaposition $\delta\delta'$ denotes their combination, that is, intersection. Finally, the always true predicate $\mathbf{T}$ denotes the absence of a restriction.

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

Lastly, we discuss $\mathbf{Bc}$-locking sequences, the semantic counterpart to locking sequences [2]. Intuitively, a $\mathbf{Bc}$-locking sequence is a sequence where the learner correctly identifies the target language and does not make a semantic mind change anymore regardless what information of the language it is presented. Formally, given a language L and a $\mathbf{G}$-learner h, a sequence $\sigma \in L_{\#}^{*}$ is called a $\mathbf{Bc}$-*locking sequence* for h on L if and only if for every sequence $\tau \in L_{\#}^{*}$ we have that $W_{h(\sigma\tau)} = L$ [6]. When talking about $\mathbf{Sd}$-learners, we call a finite set D a $\mathbf{Bc}$-*locking set* of L if and only if for all D', with $D \subseteq D' \subseteq L$, we have $W_{h(D')} = L$.

While for each $\mathbf{G}$-learner h there exists a $\mathbf{Bc}$-locking sequence on every language it learns [2], not every text may contain an initial segment which is a $\mathbf{Bc}$-locking sequence. Learners which do have a $\mathbf{Bc}$-locking sequence on every text of a language they learn are called *strongly* $\mathbf{Bc}$-*locking* [8]. Formally, a learner is *strongly* $\mathbf{Bc}$-*locking* if on every language L it learns and on every text $T \in \mathbf{Txt}(L)$ there exists n such that $T[n]$ is a $\mathbf{Bc}$-locking sequence for h on L. The transition to set-driven learners is immediate.

3 Semantic Witness-based Learning

In this section, we provide a normal form for semantically witness-based learners, namely that $\tau(\mathbf{SemWb})\mathbf{TxtSdBc}$-learners are as powerful as $\mathbf{TxtGSemConvBc}$ ones (Theorem 1). We prove this normal form stepwise. We start by showing that each $\mathbf{TxtGSemConvBc}$-learner may be assumed semantically conservative on arbitrary text (Theorem 2). Afterwards, we prove that such learners base their guesses solely on the content given (Theorem 3). Lastly, we observe that they remain equally powerful when being globally semantically witness-based (Theorem 4).

Theorem 1. *We have that* $[\tau(\mathbf{SemWb})\mathbf{TxtSdBc}] = [\mathbf{TxtGSemConvBc}]$.

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

Theorem 2. *We have that* $[\tau(\mathbf{SemConv})\mathbf{TxtGBc}] = [\mathbf{TxtGSemConvBc}]$.

Proof. The inclusion $[\tau(\mathbf{SemConv})\mathbf{TxtGBc}] \subseteq [\mathbf{TxtGSemConvBc}]$ is immediate. For the other, let h be a consistent learner [10] and $\mathcal{L} = \mathbf{TxtGSemConvBc}(h)$. We provide a learner h' which $\tau(\mathbf{SemConv})\mathbf{TxtGBc}$-learns $\mathcal{L}$.

We do so with the help of an auxiliary $\tau(\mathbf{SemConv})\mathbf{TxtGBc}$-learner $\hat{h}$, 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 $\hat{h}$.

Algorithm 1: The auxiliary $\tau(\mathbf{SemConv})$-learner $\hat{h}$.

Parameter: $\mathbf{TxtGSemConv}$-learner h.
Input: Finite sequence $\sigma \in \mathbb{S}(\mathbb{N})$.
Semantic Output: $W_{\hat{h}(\sigma)} = \bigcup_{t \in \mathbb{N}} E_t$.
Initialization: $t' \leftarrow 0$, $E_0 \leftarrow \mathrm{content}(\sigma)$ and, for all $t > 0$, $E_t \leftarrow \emptyset$.

for $t = 0$ **to** ∞ **do**
  if $\exists \sigma' \subsetneq \sigma \colon \mathrm{content}(\sigma) \subseteq W^t_{\hat{h}(\sigma')}$ **then**
    $\Sigma'_t \leftarrow \{\sigma' \subsetneq \sigma \mid \mathrm{content}(\sigma) \subseteq W^t_{\hat{h}(\sigma')}\}$
    $E_{t+1} \leftarrow E_t \cup \bigcup_{\sigma' \in \Sigma'_t} W^t_{\hat{h}(\sigma')}$
  else if $\forall \sigma' \subsetneq \sigma \colon \mathrm{content}(\sigma) \not\subseteq W^t_{h(\sigma')}$ **then**
    $S(\sigma, t') \leftarrow \mathbb{S}\left(W^{t'}_{h(\sigma)} \setminus \mathrm{content}(\sigma)\right)$
    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**
      $E_{t+1} \leftarrow E_t \cup W^{t'}_{h(\sigma)}$
      $t' \leftarrow t' + 1$
  else
    $E_{t+1} \leftarrow E_t$

Consider the learner $\hat{h}$ as in Algorithm 1 with parameter h. Given some input σ, the intuition is the following. Once $\hat{h}$, on any previous sequence σ', 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 $n \in \mathbb{N}$,

$$W_{\hat{h}(T[n])} \subseteq W_{h(T[n])}. \tag{1}$$

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

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

1. Case: L is finite. Then there exists n_0 with $\mathrm{content}(T[n_0]) = L$. Let $n \geq n_0$. By $\mathbf{SemConv}$ and consistency of h, we have $L = W_{h(T[n])}$. By Equation (1), we have $W_{h(T[n])} \supseteq W_{\hat{h}(T[n])}$ and, by consistency of $\hat{h}$, $W_{\hat{h}(T[n])} \supseteq \mathrm{content}(T[n]) = L$. Altogether we have $W_{\hat{h}(T[n])} = L$ as required.
2. Case: L is infinite. Let n_0 be minimal such that $W_{h(T[n_0])} = L$. Then, as h is semantically conservative, $T[n_0]$ is a $\mathbf{Bc}$-locking sequence for h on L and we have

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

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 (1). For the other direction, let t' be the current step of enumeration. As $T[n_0]$ is a $\mathbf{Bc}$-locking sequence, we have, for all $\tau \in S(T[n_0], t') = \mathbb{S}\left(W^{t'}_{h(T[n_0])} \setminus \mathrm{content}(T[n_0])\right)$,

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

Thus, at some step t, $E_{t+1} \leftarrow W^{t'}_{h(T[n_0])}$ and, then, the enumeration continues with $t' \leftarrow t' + 1$. 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 (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 $\mathbf{SemConv}$ on arbitrary text $T \in \mathbf{Txt}$. 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 $\tau(\mathbf{SemConv})$ formally. Let $n < n'$ be such that $\mathrm{content}(T[n']) \subseteq W_{\hat{h}(T[n])}$. 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 and 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 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_{content} be the step where $\mathrm{content}(T[n']) \subseteq W_{\hat{h}(T[n])}$ is witnessed for the first time. Now, by the definition of $\hat{h}$, we have

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

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

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

which must be witnessed in order for $W_{\hat{h}(T[n])}$ to enumerate $\mathrm{content}(T[n'])$ via the second if-clause (lines 5 to 9), that is, to get $\mathrm{content}(T[n']) \subseteq W_{\hat{h}(T[n])}$. 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. *We have that* $[\tau(\mathbf{SemConv})\mathbf{TxtSdBc}] = [\tau(\mathbf{SemConv})\mathbf{TxtGBc}]$.

Proof. Let h be a learner and $\mathcal{L} = \tau(\mathbf{SemConv})\mathbf{TxtGBc}(h)$. We may assume h to be globally consistent [10]. We provide a learner h' 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 $d := \max(D)$, σ_D be the canonical sequence of D and $D_{<x} := \{y \in D \mid y < x\}$. 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' be such that, for each finite set D,

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

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

$$W_{h'(D)} \subseteq W_{h(\sigma_D)}. \tag{2}$$

We proceed by proving that h' $\tau(\mathbf{SemConv})\mathbf{TxtSdBc}$-learns $\mathcal{L}$. First, we show the $\mathbf{TxtSdBc}$-convergence. The idea here is to find a $\mathbf{Bc}$-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. Case: 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 (2), we have $W_{h'(L)} \subseteq W_{h(\sigma_L)} = L$, concluding this case.
2. Case: L is infinite. Let T_c be the canonical text of L, and let σ_0 be a $\mathbf{Bc}$-locking sequence for h on T_c. Such a $\mathbf{Bc}$-locking sequence exists, as h is strongly $\mathbf{Bc}$-locking [10, Thm. 7]. Let $D_0 := \mathrm{content}(\sigma_0)$. For any input $D \subseteq L$ such that $D \supseteq D_0$, we show that $W_{h'(D)} = L$. By Equation (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 $x > \max(\mathrm{content}(\sigma_0))$ 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 $\mathbf{TxtSdBc}$-convergence. It remains to be shown that h' is $\tau(\mathbf{SemConv})$. Let $D' \subseteq D''$ and $D'' \subseteq W_{h'(D')}$. 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 $\sigma' := \sigma_{D'}$, $d' := \max(D')$ and $\sigma'_{<x} := \sigma_{(D'_{<x})}$. 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

$$W_{h(\sigma'_{<y})} = W_{h(\sigma')}. \tag{3}$$

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

$$\sigma'_{<x} = \sigma''_{<x}. \tag{4}$$

By Equations (2) and (3), we have $D'' \subseteq W_{h'(D')} \subseteq W_{h(\sigma')} = W_{h(\sigma'_{<m})}$. As, by Equation (4), $\sigma'_{<m} = \sigma''_{<m} \subseteq \sigma''$ and h is $\tau(\mathbf{SemConv})$, we get

$$W_{h(\sigma')} = W_{h(\sigma'')}. \tag{5}$$

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 (2) and (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 and get

$$\begin{aligned} x < m\colon D' \cup \{x\} &\subseteq D'' \cup \{x\} \subseteq W_{h(\sigma''_{<x})} \overset{(4)}{=} W_{h(\sigma'_{<x})}, \\ m < x < d'\colon D' \cup \{x\} &\subseteq D'' \cup \{x\} \subseteq W_{h(\sigma''_{<x})} \overset{(*)}{=} W_{h(\sigma'')} \overset{(5)}{=} W_{h(\sigma')} \overset{(3)}{=} \\ &\overset{(3)}{=} W_{h(\sigma'_{<m})} \overset{(*)}{=} W_{h(\sigma'_{<x})}. \end{aligned}$$

We use h being $\tau(\mathbf{SemConv})$ in the steps marked by $(*)$. Thus, $x \in W_{h'(D')}$.

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

$$x \in D'' \cup \{x\} \subseteq W_{h'(D')} \subseteq W_{h(\sigma')} \overset{(5)}{=} W_{h(\sigma'')}.$$

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

$$\begin{aligned} x < m : D'' \cup \{x\} &\subseteq W_{h(\sigma'_{<x})} \overset{(4)}{=} W_{h(\sigma''_{<x})}, \\ m < x < d' : D'' \cup \{x\} &\subseteq W_{h(\sigma'_{<x})} \overset{(*)}{=} W_{h(\sigma'_{<m})} \overset{(4)}{=} W_{h(\sigma''_{<m})} \overset{(*)}{=} W_{h(\sigma''_{<x})}, \\ d' < x < d'' : D'' \cup \{x\} &\subseteq W_{h(\sigma')} = W_{h(\sigma'_{<m})} = W_{h(\sigma''_{<m})} \overset{(4)}{=} W_{h(\sigma''_{<x})}. \end{aligned}$$

We use h being $\tau(\mathbf{SemConv})$ in the steps marked by $(*)$. In the end, $x \in W_{h'(D'')}$. $\square$

Hence, we may assume h to be $\tau(\mathbf{SemConv})\mathbf{TxtSdBc}$. Lastly, we observe that h may even be assumed globally semantically witness-based. This concludes the proof of Theorem 1 and, thus, also this section.

Theorem 4. *We have that* $[\tau(\mathbf{SemWb})\mathbf{TxtSdBc}] = [\tau(\mathbf{SemConv})\mathbf{TxtSdBc}]$.

Proof. Let $\delta \in \{\mathbf{SemWb}, \mathbf{SemConv}\}$. Since δ-learners may be assumed to be consistent [10, Thm. 8], which also holds true when the restrictions are required globally, we have $[\tau(\mathbf{Cons}\delta)\mathbf{TxtSdBc}] = [\tau(\delta)\mathbf{TxtSdBc}]$. Since $\mathbf{Cons} \cap \mathbf{SemWb} = \mathbf{Cons} \cap \mathbf{SemConv}$ [10, Lem. 11], the theorem holds. $\square$

References

1. Angluin, D.: Inductive inference of formal languages from positive data. Information and Control **45**, 117–135 (1980)
2. Blum, L., Blum, M.: Toward a mathematical theory of inductive inference. Information and Control **28**, 125–155 (1975)
3. Case, J., Lynes, C.: Machine inductive inference and language identification. In: Proc. of the International Colloquium on Automata, Languages and Programming (ICALP). pp. 107–115 (1982)
4. Doskoč, V., Kötzing, T.: Cautious limit learning. In: Proc. of the International Conference on Algorithmic Learning Theory (ALT) (2020)
5. Gold, E.M.: Language identification in the limit. Information and Control **10**, 447–474 (1967)
6. Jain, S., Osherson, D., Royer, J.S., Sharma, A.: Systems that Learn: An Introduction to Learning Theory. MIT Press, Cambridge (MA), Second Edition (1999)
7. Jantke, K.: Monotonic and non-monotonic inductive inference. New Generation Computing **8**, 349–360 (1991)
8. Kötzing, T., Palenta, R.: A map of update constraints in inductive inference. Theoretical Computer Science **650**, 4–24 (2016)
9. Kötzing, T., Schirneck, M.: Towards an atlas of computational learning theory. In: Proc. of the Symposium on Theoretical Aspects of Computer Science (STACS). pp. 47:1–47:13 (2016)
10. Kötzing, T., Schirneck, M., Seidel, K.: Normal forms in semantic language identification. In: Proc. of the International Conference on Algorithmic Learning Theory (ALT). pp. 76:493–76:516 (2017)
11. Kötzing, T.: Abstraction and Complexity in Computational Learning in the Limit. Ph.D. thesis, University of Delaware (2009)
12. Osherson, D.N., Stob, M., Weinstein, S.: Learning strategies. Information and Control **53**, 32–51 (1982)
13. Osherson, D.N., Weinstein, S.: Criteria of language learning. Information and Control **52**, 123–138 (1982)
14. Rogers Jr., H.: Theory of recursive functions and effective computability. Reprinted by MIT Press, Cambridge (MA) (1987)
15. Schäfer-Richter, G.: Über Eingabeabhängigkeit und Komplexität von Inferenzstrategien. Ph.D. thesis, RWTH Aachen University, Germany (1984)
16. Wexler, K., Culicover, P.W.: Formal principles of language acquisition. MIT Press, Cambridge (MA) (1980)
17. Wiehagen, R.: A thesis in inductive inference. In: Nonmonotonic and Inductive Logic. pp. 184–207 (1991)