Behaviorally Correct Learning from Informants

Verhaltenskorrektes Lernen von Informanten

Niklas Mohrin

Universitätsbachelorarbeit
zur Erlangung des akademischen Grades

Bachelor of Science
(B. Sc.)

im Studiengang
IT-Systems Engineering

eingereicht am 30. Juni 2022 am
Fachgebiet Algorithm Engineering der
Digital-Engineering-Fakultät
der Universität Potsdam

Gutachter	Prof. Dr. Tobias Friedrich
Betreuer	Dr. Timo Kötzing
	Vanja Doskoč

Abstract

In inductive inference, we investigate the learnability of classes of formal languages. We are interested in what classes of languages are learnable in certain learning settings. A class of languages is learnable, if there is a learner that can identify all of its languages and satisfies the constraints of the learning setting. To identify a language, a learner is presented with information about this very language. When learning from informants, this information consists of examples for numbers that are, and numbers that are not included in the target language. As more and more examples are presented, the learner outputs a hypothesis sequence. To satisfy behaviorally correct identification, this hypothesis sequence must eventually only list correct labels for the target language. In this thesis, we compare the effects of a number of semantic learning restrictions on the learning capabilities for behaviorally correct learning from informants.

To start, we collect and combine some known theorems to show that we can assume learners to be set-driven and total. Additionally, learners only need to identify languages by their canonical informant, an informant that is particularly well-formed. We then investigate the effects of a number of learning restrictions on the inference capabilities of learners. Most importantly, we specify all relations between monotonic and strong monotonic restrictions, including dual and combined versions. Monotonicity restrictions require that a learner outputs better generalizations (or, for the dual variants, specializations) over time. Similarly to what has been found in the setting of learning indexed families, these restrictions form a strict hierarchy. We show that weak monotonicity does not restrict learners. We reprove that cautiousness is a proper restriction by investigating three weaker variants. Interestingly, while infinite cautiousness does not lessen learning power in full-information text-learning, the setting they were introduced in, we show that all three decrease the set of learnable classes in our setting. Finally, we show that we can assume global consistency for learners satisfying any of the monotonic restrictions we investigate, with the exception of combined weak monotonicity.

Zusammenfassung

In der induktiven Inferenz untersuchen wir die Lernbarkeit von Klassen formaler Sprachen. Wir sind daran interessiert, welche Klassen von Sprachen in bestimmten Lernumgebungen lernbar sind. Eine Klasse von Sprachen ist lernbar, wenn es einen Lerner gibt, der alle enthaltenen Sprachen identifizieren kann und die Bedingungen der Lernumgebung erfüllt. Um eine Sprache zu identifizieren, erhält der Lerner Informationen über ebendiese Sprache. Beim Lernen von Informanten bestehen diese Informationen aus Beispielen für Zahlen, die in der Zielsprache enthalten sind, und für Zahlen, die in der Zielsprache nicht enthalten sind. Während mehr und mehr Beispiele präsentiert werden, gibt der Lerner eine Hypothesenfolge aus. Für verhaltenskorrekte Identifikation muss diese Hypothesenfolge irgendwann nur noch die Zielsprache beschreiben. In dieser Arbeit vergleichen wir die Auswirkungen einer Reihe semantischer Lernrestriktionen auf die Lernfähigkeit für verhaltenskorrektes Lernen von Informanten.

Zu Beginn sammeln und kombinieren wir einige bekannte Theoreme, um zu zeigen, dass wir davon ausgehen können, dass Lerner mengengetrieben und total sind. Außerdem müssen Lerner Sprachen nur anhand ihres kanonischen Informanten identifizieren, also eines Informanten, der besonders wohlgeformt ist. Anschließend untersuchen wir die Auswirkungen einer Reihe von Lernrestriktionen auf die Inferenzfähigkeiten des Lerners. Vor allem spezifizieren wir alle Beziehungen zwischen monotonen und stark monotonen Restriktionen, einschließlich dualer und kombinierter Varianten. Monotonitätsrestriktionen setzen voraus, dass ein Lerner im Laufe der Zeit bessere Verallgemeinerungen (oder, bei den dualen Varianten, Spezialisierungen) liefert. Ähnlich wie beim Lernen indizierter Familien bilden diese Einschränkungen eine strenge Hierarchie. Wir zeigen, dass schwache Monotonität Lerner nicht einschränkt. Wir beweisen, dass Behutsamkeit eine tatsächliche Einschränkung ist, indem wir drei schwächere Varianten untersuchen. Obwohl unendliche Behutsamkeit die Lernfähigkeit beim Text-Lernen mit vollständiger Information nicht verringert, zeigen wir, dass alle drei die Menge der lernbaren Klassen in unserer Konfiguration verringern. Schließlich zeigen wir, dass wir globale Konsistenz für Lerner annehmen können, die alle von uns untersuchten monotonen Restriktionen erfüllen, mit Ausnahme der kombinierten schwachen Monotonität.

Acknowledgments

I want to thank Timo and Vanja for their continuous feedback and help in finding new ideas. Without their insights and explanations, much of this work would have been out of scope for me. Their guidance improved this thesis, not only content-wise, but also concerning phrasing, style and LaTeX difficulties.

Thanks to Paula for sharing the effort on grasping the most difficult must-know definitions inductive inference has to offer, for being great company in the office and for verifying my thoughts and results. I definitely have to thank our "extended bachelor's project", consisting of Ben, Jonas, Niko and Sebastian for their regular input on all of my work and for supplying the office with coffee and graph theory brainteasers. Thanks to Kathrin and Lukas for being generally good mates in the office. Finally, I want to thank my friends Adrian and Tobias for having a look at the thesis.

Contents

1   Introduction

Inductive inference, also called *language learning in the limit*, is a branch of *recursion theory* and was introduced by Gold [Gol67]. The model was initially designed to draw parallels to how humans learn natural languages. Nowadays we are more interested into its implications for computability theory, machine learning and binary classification [Sei21]. Inductive inference is about the learnability of classes of recursively enumerable *formal languages*, i.e. subsets of the natural numbers. A class of languages is *learnable*, if there is an algorithmic *learner* that can correctly identify all of its languages. To identify a language, the learner has to recognize it when given some hints about it. We study what classes of languages are learnable when varying our requirements for admissible learners. For example, the content of the given information depends on the *presentation system*. The two presentation systems for language learning are *text* (**Txt**) and *informant* (**Inf**), both of which have been introduced by Gold [Gol67]. A text contains elements of the target language, whereas an informant also gives counter-examples.

Convergence criteria define what it means to correctly identify a language. The default convergence criterion is *explanatory learning* (**Ex**) and was introduced by Gold [Gol67]. **Ex** requires that the learner converges to one output that correctly explains the target language. A relaxed version of **Ex** is *behaviorally correct* (**Bc**) identification [CS83]. **Bc**-learners are allowed to output different explanations as long as they all explain the correct language. Out of the four combinations of text- and informant-learning and explanatory and behaviorally correct identification, **Bc**-learning from informants has been studied the least. In this thesis we collect known theorems in this area and fill some of the gaps.

Fixing **Inf** and **Bc**, we compare the effect of *learning restrictions* on the inference capabilities of learners. As they are given more and more information, learners output a *hypothesis sequence*. Learning restrictions are predicates over the hypothesis sequence and give a formal model for requiring learners to, for example, come up with better generalizations over time, come up with better specializations over time or just not contradict the input data. Learning restrictions can be grouped by some properties. A learning restriction is called *semantic*, if it is only concerned about a learner's conjectured languages, not the labels used to encode those [Köt17]. Another property of learning restrictions is *delayability*, which requires that the restriction is preserved when shifting or skipping hypotheses [KP16].

1.1 Contributions

We arrange our results in four chapters: The first for general observations and the other three for cautiousness, monotonicity and consistency.

The general observations are collected and combined theorems for groups of restrictions.

- For delayable restrictions, it is sufficient that learners identify languages by their canonical (a particularly well-formed) informant and that we can assume them to be set-driven and total (see Theorem 3.2).
- For restrictions that are both delayable and semantic, we can additionally limit the interaction with a learner to be iterative (see Theorem 3.4).

Using these findings, we show that the relaxation of **Bc** from **Ex** actually increases inference capabilities of learners (see Theorem 3.6).

We reprove that cautiousness is a proper restriction by investigating three weaker variants. We find that all three properly restrict learners in our setting (see Theorem 4.1). This is in contrast to explanatory learning form text, where infinite cautiousness does not weaken learners [KP16].

We provide an initial map for behaviorally correct learning from informants that covers monotonic and strong monotonic restrictions and their dual and combined counterparts (see Figure 1.1). We find that these restrictions form a strict hierarchy, analogous to what has been observed in the adjacent setting of learning *indexed families* by Lange and Zeugmann [LZ94]. This is due to the fact that most of their separations can be transferred while the inclusions hold by definition. We observe that weak monotonicity, like with **Ex**, does not pose a proper restriction on learners. However, our approach is very different from constructions for similar theorems in explanatory learning or learning indexed families. Additionally, our constructed learner behaves globally consistent (see Theorem 5.1).

In Theorem 6.7, Theorem 6.9 and Theorem 6.11, we construct globally consistent learners that preserve

- behaviorally correct identification,
- classic, dual and combined monotonicity,
- classic, dual and combined strong monotonicity and
- dual weak monotonicity.

Throughout the thesis, we employ established proof techniques such as separation using the *Operator Recursion Theorem* [Cas94] (see Theorem 3.6) and *poisoning* (see Theorem 5.1).
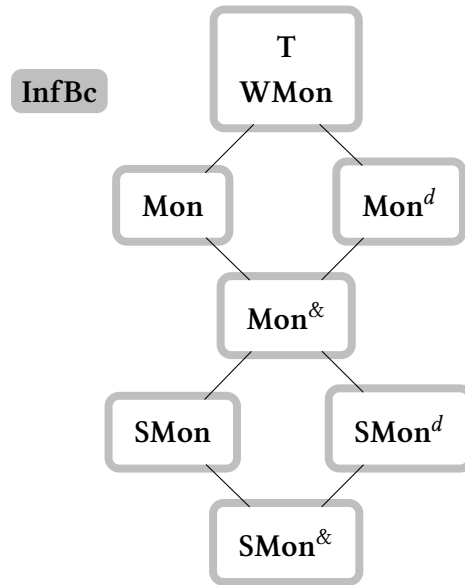
Figure 1.1: Relations between monotonicity constraints in **InfBc**-learning. Black lines indicate inclusions. Two learning restrictions are equivalent if and only if they lie in the same gray box. For all displayed restrictions, learners can additionally be assumed to be set-driven, total and globally consistent.

1.2 Related work

The most exhaustive work on learning from informants is by Aschenbach et al. [AKS18], where a map for a lot of common learning restrictions in the **Ex** setting is provided. Although they did not explicitly consider **Bc**, some of their separations transfer directly into our setting. Interestingly, they found that only **Mon**, **Caut** and **SMon** impose a proper restriction on learners. Furthermore, they observed that learners may be assumed to be set-driven. This is in contrast to text learning, where set-drivenness is a strong restriction. Their map does not include dual monotonic learning or the weak and strong counterparts. For learning indexed families, Lange et al. [LZK96] built a map for all monotonic restrictions in an **InfEx** learning scenario. Monotonic learning has also been studied in other research, for example by Doskoč and Kötzing [DK21a].

Concerning behaviorally correct learning, most findings are included in more general theorems about delayable or semantic restrictions. As an example, Aschenbach et al. [AKS18] showed that totality is not a restrictive assumption for delayable learning restrictions. Totality can be assumed for all semantic restrictions too, as shown by Kötzing et al. [KSS17].

Since most of the learning restrictions were built with **Ex**-learning in mind, there has been work to develop equivalent restrictions that are semantic. Kötzing et al. [KSS17] introduced the concept of a *semantic closure* to better classify restrictions and to find semantic equivalents of known restrictions. Following that, Doskoč and Kötzing [DK21b] investigated a number of semantic restrictions.

2 Preliminaries

We refer the reader to the textbook by Rogers Jr [Rog87] for a more thorough introduction to computability theory. We denote the set of natural numbers by $\mathbb{N} = \{0, 1, 2, \ldots\}$ and the empty set by $\emptyset$. For some set S, the cardinality of S is written as $|S|$. By $\subseteq, \supseteq, \subsetneq$ and $\supsetneq$ we denote the relations subset, superset, proper subset and proper superset. If two sets A, B are incomparable, i.e. there are $x \in A \setminus B$ and $y \in B \setminus A$, we write $A \# B$. By $\min(A)$ and $\max(A)$, we denote the minimum and maximum of A. By $\sup(A)$, we refer to the supremum of the set A, i.e. the smallest upper bound of A. For convenience, we suppose that $\sup(\emptyset) = -\infty$. For some set S, we denote the set of all sequences over S by $\mathbb{S}\mathrm{eq}(S)$. By $\sqsubseteq$ and $\sqsubsetneq$ we denote the relations subsequence and proper subsequence. The notation $\forall^{\infty} n \in \mathbb{N}$ means "for all but finitely many $n \in \mathbb{N}$".

If a partial function f is undefined for some input x, we write $f(x)\uparrow$ or $f(x) = \perp$ and say f diverges on input x. Otherwise, f converges on x, denoted by $f(x)\downarrow$. If a function f converges on all inputs $x \in \mathbb{N}$, we say that f is total. By $\mathfrak{P}, \mathcal{P}, \mathfrak{R}$ and $\mathcal{R}$, we denote the set of all functions, all computable functions, all total functions and all total computable functions respectively. For some function f and some number $n \in \mathbb{N}$, we denote the finite sequence of the first n outputs of f by $f[n]$.

We assume a numbering system φ for the set of computable functions. For each $f \in \mathcal{P}$, there is $e \in \mathbb{N}$ with $\varphi_e = f$. By ϕ, we denote a Blum complexity measure for φ [Blu67]. For example, for $e, x \in \mathbb{N}$, $\phi_e(x)$ could equal the number of steps the program coded by e takes on input x. By $\mathrm{dom}(f)$ and $\mathrm{range}(f)$, we refer to the domain and range of a function f. A *language* is a set of natural numbers. By $\mathcal{E}$, we denote the set of all recursively enumerable languages, i.e. all languages $L \subseteq \mathbb{N}$ for which there exists some $f \in \mathcal{P}$ with $L = \mathrm{dom}(f)$. To describe languages using natural numbers we use the W-indices system as our hypothesis space. For all $e \in \mathbb{N}$, we define $W_e = \mathrm{dom}(\varphi_e)$, e is called a *label* for the language W_e. Furthermore, for $e, t \in \mathbb{N}$, we define $W_e^t = \{x \in \mathbb{N} \mid x \le t \wedge \phi_e(x) \le t\}$. Notably, W_e^t is finite and its construction is total and computable. We define two numbers $a, b \in \mathbb{N}$ to be *semantically equivalent*, denoted by $a \equiv_W b$, if and only if $W_a = W_b$. We use established theorems from computability theory, such as s-m-n, Kleene's Recursion Theorem (**KRT**) [Kle52] and the Operator Recursion Theorem (**ORT**) [Cas94].

2.1 Informants

Before we define what an informant is, we fix an interpretation for its output and introduce some useful vocabulary. For any set $D \subseteq \mathbb{N} \times \{0, 1\}$ we define *positive information*, *negative information* and *outline* of D as

$$
\begin{aligned}
\mathrm{pos}(D) &= \{x \in \mathbb{N} \mid (x, 1) \in D\}, \\
\mathrm{neg}(D) &= \{x \in \mathbb{N} \mid (x, 0) \in D\}, \\
\mathrm{outline}(D) &= \mathrm{pos}(D) \cup \mathrm{neg}(D).
\end{aligned}
$$

For any finite or infinite sequence σ, we define

$$
\mathrm{content}(\sigma) = \mathrm{range}(\sigma).
$$

We define $\mathrm{pos}(\sigma)$, $\mathrm{neg}(\sigma)$ and $\mathrm{outline}(\sigma)$ analogously to the definition on sets.

A function $I \colon \mathbb{N} \to \mathbb{N} \times \{0, 1\}$ is an *informant* for a language $L \subseteq \mathbb{N}$ if $\mathrm{pos}(I) = L$ and $\mathrm{neg}(I) = \mathbb{N} \setminus L$. By **Inf**, we denote the set of all informants and for some language L we define $\mathbf{Inf}(L)$ to be the set of informants for L. An informant I is *canonical*, if and only if for all $x \in \mathbb{N}$, we have $I(x) = (x, 0)$ or $I(x) = (x, 1)$. For all languages L, we define $\hat{I}_L$ to be the canonical informant for L. Furthermore, we define $\mathbf{Inf}_{\mathrm{can}}$ to be set of all canonical informants.

2.2 Interaction operators

According to an *interaction operator*, the output of an informant I may be presented to a learner h in different formats. For a learner $h \in \mathfrak{P}$ and informant $I \in \mathbf{Inf}$ we define the interaction operators **G** (Gold / full-information, [Gol67]), **Psd** (partially set-driven, [Sch84]), **Sd** (set-driven, [WC80]) and **It** (iterative, [WC80]) that generate the *hypothesis sequence* of a learner h such that for all $i \in \mathbb{N}$ we have

$$
\begin{aligned}
\mathbf{G}(h, I)(i) &= h(I[i]), \\
\mathbf{Psd}(h, I)(i) &= h(\mathrm{content}(I[i]), i), \\
\mathbf{Sd}(h, I)(i) &= h(\mathrm{content}(I[i])), \\
\mathbf{It}(h, I)(i) &= h(\mathbf{It}(h, I)(i - 1), I[i]).
\end{aligned}
$$

Furthermore, we define the interaction operator **CflIt** (confluently iterative, [KS16]), which is equivalent to **It**, but also requires that the order and quantity in which the content is presented does not matter.

It is immediate that some operators pass strictly more information to the learner. Case and Kötzing [CK10] manifest this notion by introducing a partial ordering of the interaction operators. For two interaction operators β, β' we define the relation $\leqslant$ such that

$$\beta \leqslant \beta' \iff [\forall h \in \mathfrak{P} \exists h' \in \mathfrak{P} \forall I \in \mathbf{Inf}: \beta(h, I) = \beta'(h', I)].$$

Intuitively, β-learners can be translated into β'-learners. A slightly weakened version of this relation has been introduced by Kötzing et al. [KSS17] that only requires semantic equivalence of the hypothesis sequences. For two interaction operators β, β' we define the relation $\leqslant_{\text{sem}}$ such that

$$\beta \leqslant_{\text{sem}} \beta' \iff [\forall h \in \mathfrak{P} \exists h' \in \mathfrak{P} \forall I \in \mathbf{Inf}, n \in \mathbb{N}: \beta(h, I)(n) \equiv_W \beta'(h', I)(n)].$$

Intuitively, β-learners can be translated into semantically equivalent β'-learners. Additionally, if the relation goes both ways, we say $\beta \cong_{\text{sem}} \beta'$.

While **CflIt** is standing out of line in our definition above, it is of great use for generalizing over interaction operators. This is because it is a lower bound for all other interaction operators in the $\leqslant$ relation. We have

$$\mathbf{CflIt} \leqslant \mathbf{Sd} \leqslant \mathbf{Psd} \leqslant \mathbf{G},$$
$$\mathbf{CflIt} \leqslant \mathbf{It} \leqslant \mathbf{G},$$
$$\mathbf{It} \cong_{\text{sem}} \mathbf{G},$$
$$\mathbf{CflIt} \cong_{\text{sem}} \mathbf{Sd}.$$

2.3 Learning restrictions

Now that we know how hypothesis sequences are generated, we define some restrictions on them. Our restrictions are defined as predicates over the hypothesis sequence and the informant that generated that sequence. When using the restrictions, we view them as predicates or sets containing all the elements satisfying the predicate interchangeably. Let **T** be the predicate that is always true. To start, we define two convergence criteria.

For a hypothesis sequence $p \in \mathfrak{P}$ generated from an informant $I \in \mathbf{Inf}$, we define the convergence criteria explanatory and behaviorally correct such that

$$\mathbf{Ex}(p, I) \iff [\exists q \forall^\infty n: p(n) = q \wedge W_q = \text{pos}(I)],$$
$$\mathbf{Bc}(p, I) \iff [\forall^\infty n: W_{p(n)} = \text{pos}(I)].$$

Of course, $\mathbf{Ex} \subseteq \mathbf{Bc}$. Next, we define the concept of consistency as first introduced by Angluin [Ang80]. Consistency of a hypothesis $e \in \mathbb{N}$ with information $D \subseteq \mathbb{N} \times \{0, 1\}$ is defined such that

$$\mathbf{Cons}(e, D) \iff [\mathrm{pos}(D) \subseteq W_e \wedge \mathrm{neg}(D) \cap W_e = \emptyset].$$

We define this analogously for languages described by their label e and for other information formats, such as sequences. Furthermore, we define consistency of a hypothesis sequence $p \in \mathfrak{P}$ with an informant $I \in \mathbf{Inf}$ such that

$$\mathbf{Cons}(p, I) \iff [\forall n \in \mathbb{N} \colon \mathbf{Cons}(p(n), I[n])].$$

Monotonicity restrictions are organized into three groups: strong monotonicity, monotonicity and weak monotonicity. In each group, we have one variant that requires better generalizations, one that requires better specializations and one that is the conjunction of the two. The generalization variants were all introduced first. The dual (and combined) variants were all introduced by Lange et al. [LZK96]. Strong monotonicity [Jan91] is the most strict of the there groups. For all hypothesis sequences $p \colon \mathbb{N} \to \mathbb{N}$ and informants $I \in \mathbf{Inf}$, we define the restrictions strongly monotone, dual strongly monotone and combined strongly monotone, such that

$$\mathbf{SMon}(p, I) \iff [\forall s, t \colon s \leq t \Rightarrow W_{p(s)} \subseteq W_{p(t)}],$$
$$\mathbf{SMon}^d(p, I) \iff [\forall s, t \colon s \leq t \Rightarrow W_{p(s)} \supseteq W_{p(t)}],$$
$$\mathbf{SMon}^{\&}(p, I) \iff [\mathbf{SMon}(p, I) \wedge \mathbf{SMon}^d(p, I)].$$

Shortly after that, Wiehagen [Wie90] introduced the concept of (what we today call) monotonicity, which requires this behavior only on correct elements. For all hypothesis sequences $p \colon \mathbb{N} \to \mathbb{N}$ and informants $I \in \mathbf{Inf}$, we define the restrictions monotone, dual monotone and combined monotone, such that

$$\mathbf{Mon}(p, I) \iff [\forall s, t \colon s \leq t \Rightarrow W_{p(s)} \cap \mathrm{pos}(I) \subseteq W_{p(t)} \cap \mathrm{pos}(I)],$$
$$\mathbf{Mon}^d(p, I) \iff [\forall s, t \colon s \leq t \Rightarrow \overline{W_{p(s)}} \cap \mathrm{neg}(I) \subseteq \overline{W_{p(t)}} \cap \mathrm{neg}(I)],$$
$$\mathbf{Mon}^{\&}(p, I) \iff [\mathbf{Mon}(p, I) \wedge \mathbf{Mon}^d(p, I)].$$

Finally, Jantke [Jan91] also introduced weak monotonicity where better generalizations are only required as long as new information does not contradict old hypotheses. For all hypothesis sequences $p \colon \mathbb{N} \to \mathbb{N}$ and informants $I \in \mathbf{Inf}$, we define the restrictions weakly monotone, dual weakly monotone and combined

weakly monotone, such that

$$\mathbf{WMon}(p, I) \iff [\forall s, t \colon s \leq t \land \mathbf{Cons}(p(s), I[t]) \Rightarrow W_{p(s)} \subseteq W_{p(t)}],$$

$$\mathbf{WMon}^d(p, I) \iff [\forall s, t \colon s \leq t \land \mathbf{Cons}(p(s), I[t]) \Rightarrow W_{p(s)} \supseteq W_{p(t)}],$$

$$\mathbf{WMon}^{\&}(p, I) \iff [\mathbf{WMon}(p, I) \land \mathbf{WMon}^d(p, I)].$$

Note that, $\mathbf{SMon} \subseteq \mathbf{Mon} \cap \mathbf{WMon}$. This holds analogously for $\mathbf{SMon}^d$ and $\mathbf{SMon}^{\&}$. Furthermore, combined weak monotonicity is equivalent to *semantic conservativeness* [KSS17].

The restriction of cautiousness was introduced by Osherson et al. [OSW82]. For a learner to be cautious, we require it to never conjecture a proper subset of a previous hypothesis. For explanatory learning, this poses a proper restriction as shown for example by Kötzing and Palenta [KP16] for text-learning and by Aschenbach et al. [AKS18] for informant-learning. To investigate the restriction of cautiousness more closely, Kötzing and Palenta [KP16] introduced three new variants of cautiousness. For all hypothesis sequences $p \colon \mathbb{N} \to \mathbb{N}$ and informants $I \in \mathbf{Inf}$, we define the restrictions cautious, target cautious, finitely cautious and infinitely cautious, such that

$$\mathbf{Caut}(p, I) \iff [\forall s < t \colon \neg(W_{p(s)} \supsetneq W_{p(t)})],$$

$$\mathbf{Caut_{Tar}}(p, I) \iff [\forall s \colon \neg(W_{p(s)} \supsetneq \mathrm{pos}(I))],$$

$$\mathbf{Caut_{Fin}}(p, I) \iff [\forall s < t \colon W_{p(s)} \supsetneq W_{p(t)} \Rightarrow W_{p(t)} \text{ is infinite}],$$

$$\mathbf{Caut_{\infty}}(p, I) \iff [\forall s < t \colon W_{p(s)} \supsetneq W_{p(t)} \Rightarrow W_{p(t)} \text{ is finite}].$$

Note that, $\mathbf{Caut} = \mathbf{Caut_{Fin}} \cap \mathbf{Caut_{\infty}}$ and that $\mathbf{SMon} \subseteq \mathbf{Caut} \subseteq \mathbf{Caut_{Tar}}$.

Learning restrictions can be combined. For example, to require both consistency and behavioral correctness, we use the restriction $\mathbf{ConsBc}$. We can require learning restrictions *locally* or *globally*. A local restriction only has to be fulfilled when an informant for a target language is presented, global restriction must hold for any informant. By Δ, we denote the set of all learning restrictions (excluding convergence criteria) defined above, i.e.

$$\Delta = \begin{Bmatrix} \mathbf{Cons}, \mathbf{SMon}, \mathbf{SMon}^d, \mathbf{SMon}^{\&}, \mathbf{Mon}, \mathbf{Mon}^d, \mathbf{Mon}^{\&}, \\ \mathbf{WMon}, \mathbf{WMon}^d, \mathbf{WMon}^{\&}, \mathbf{Caut}, \mathbf{Caut_{Tar}}, \mathbf{Caut_{Fin}}, \mathbf{Caut_{\infty}} \end{Bmatrix}.$$

In total, a learning setting consists of five elements: The set of allowed learners C, the set of inputs used P, the interaction operator β, the global restriction α and the local restriction δ. A setting S is thus defined as $S = (\alpha, P, C, \beta, \delta)$. We refer to

it as $C\tau(\alpha)P\beta\delta$, for example $\mathcal{R}\tau(\mathbf{Cons})\mathbf{Inf GBc}$. If $C = \mathcal{P}$ or $\alpha = \mathbf{T}$, we omit writing the respective part.

Given some learning setting $S = (\alpha, P, C, \beta, \delta)$, we say that a learner h S-learns the empty set, if $h \notin C$ or if there is some $I \in P$ for which $\alpha(\beta(h, I), I)$ is false. Otherwise, it learns the set of languages

$$S(h) = \{L \in \mathcal{E} \mid \forall I \in P(L): \delta(\beta(h, I), I)\}.$$

By $[S]$, we denote the set of all S-learnable sets of languages (by any learner).

2.4 Properties of learning restrictions

First, we introduce the concept of delayability [KP16]. Let $\mathfrak{S}$ denote the set of all non-decreasing, unbounded functions $\mathbb{N} \to \mathbb{N}$. A learning restriction δ is called *delayable*, if and only if for all informants $I, I' \in \mathbf{Inf}$ with content$(I) =$ content(I'), hypothesis sequences p and functions $s \in \mathfrak{S}$ the following implication holds. If we have $(p, I) \in \delta$ and for all $n \in \mathbb{N}$ that content$(I[s(n)]) \subseteq$ content$(I'[n])$, then $(p \circ s, I') \in \delta$. Intuitively, a learning restriction is delayable, if we can skip or duplicate hypotheses in the hypothesis sequence, but the restriction still holds. The second grouping we investigate is the one of *semantic* restrictions. The concept was first defined in Kötzing [Köt17], we follow the definitions by Kötzing et al. [KSS17] though. For all $p \in \mathfrak{P}$, we fix the set

$$\mathbf{Sem}(p) = \{p' \in \mathfrak{P} \mid \forall i: (p(i){\downarrow} \Leftrightarrow p'(i){\downarrow}) \wedge (p(i){\downarrow} \Rightarrow p(i) \equiv_W p'(i))\}.$$

A learning restriction δ is said to be *semantic*, if for any sequence p and informant I, we can conclude from $(p, I) \in \delta$ and $p' \in \mathbf{Sem}(p)$ that $(p', I) \in \delta$. Intuitively, a semantic restriction is indifferent about semantically equivalent hypothesis sequences. Out of all restrictions defined above, only $\mathbf{Cons}$ is *not* delayable and only $\mathbf{Ex}$ is *not* semantic. Note that the combination of delayable restrictions is again delayable and the combination of semantic restrictions is again semantic.

3

General observations

In this chapter, we collect some theorems for delayable and for semantic restrictions. In particular, we find that learning power is not dependent on the interaction operator (see Theorem 3.4). At the end of the chapter, we use our findings about delayable restrictions to motivate the case for behaviorally correct learning by showing that it is strictly more powerful than explanatory learning (see Theorem 3.6).

3.1 Delayable restrictions

We first look at delayable restrictions and observe that correct identification from canonical informants suffices and that we can assume learners to be set-driven and total. To do so, we combine two theorems by Aschenbach et al. [AKS18]. As our new theorem builds on one of the proofs, we also include it here.

▶ **Theorem 3.1 ([AKS18]).** For any delayable δ, we have $[\mathbf{Inf}_{\mathrm{can}}\mathbf{G}\delta] = [\mathbf{Inf}\mathbf{Sd}\delta]$.
◀

Proof. By definition, $[\mathbf{Inf}_{\mathrm{can}}\mathbf{G}\delta] \supseteq [\mathbf{Inf}\mathbf{Sd}\delta]$. Let h be a $\mathbf{Inf}_{\mathrm{can}}\mathbf{G}\delta$-learner and $\mathcal{L} = \mathbf{Inf}_{\mathrm{can}}\mathbf{G}\delta(h)$. We construct a learner g that builds the longest possible prefix of a canonical informant from the information it has. It then passes this prefix to h. Recall that for any language L, $\hat{I}_L$ is the canonical informant for L.

For any finite set $D \subseteq \mathbb{N} \times \{0, 1\}$, we define the length of the prefix as $\ell(D) = \max\{n \in \mathbb{N} \mid \forall i < n : i \in \mathrm{outline}(D)\}$. Next, the prefix itself is defined as $c(D) = \hat{I}_{\mathrm{pos}(D)}[\ell(D)]$. Finally, our prediction is $g(D) = h(c(D))$.

Let $L \in \mathcal{L}$ and let $I \in \mathbf{Inf}(L)$ be any informant for L. In order to apply delayability we need a simulating function $s \in \mathfrak{S}$. We pick $s : n \mapsto \ell(\mathrm{content}(I[n]))$. By definition of ℓ, s is non-decreasing. Since I is an informant, s must also be unbounded. Therefore, $s \in \mathfrak{S}$.

Let $n \in \mathbb{N}$. By definition of c and s, we have $\hat{I}_L[s(n)] = \hat{I}_L[\ell(\mathrm{content}(I[n]))] = c(\mathrm{content}(I[n]))$. Thus, $\mathrm{content}(\hat{I}_L[s(n)]) \subseteq \mathrm{content}(I[n])$ and

$$g(\mathrm{content}(I[n])) = h(c(\mathrm{content}(I[n]))) = h(\hat{I}_L[s(n)]).$$

Therefore, $\mathbf{Sd}(g, I) = \mathbf{G}(h, \hat{I}_L) \circ s$. As δ is delayable and $(\mathbf{G}(h, \hat{I}_L), \hat{I}_L) \in \delta$, we can conclude that $(\mathbf{Sd}(g, I), I) = (\mathbf{G}(h, \hat{I}_L) \circ s, I) \in \delta$. Hence, $\mathcal{L}$ is $\mathbf{Inf}\mathbf{G}\delta$-learnable. ■

▶ **Theorem 3.2.** For any delayable δ, we have $[\mathbf{Inf_{can}G}\delta] = [\mathcal{R}\mathbf{InfSd}\delta]$.   ◀

Proof. Aschenbach et al. [AKS18] showed that $[\mathcal{R}\mathbf{InfG}\delta] = [\mathbf{InfG}\delta]$. By Theorem 3.1, we have $[\mathbf{Inf_{can}G}\delta] = [\mathbf{InfG}\delta]$ and thus $[\mathbf{Inf_{can}G}\delta] = [\mathcal{R}\mathbf{Inf_{can}G}\delta]$. This means, that we can assume the G-learner picked in the proof of Theorem 3.1 to be total, making the constructed **Sd**-learner total as well.   ■

3.2 Semantic restrictions

With delayable restrictions, we get equivalent learning capabilities for interaction operators until **Sd**. To include iterative learners, we need to look at semantic restrictions. Kötzing et al. [KSS17] showed that $\mathbf{G} \cong_{\text{sem}} \mathbf{It}$ and $\mathbf{Sd} \cong_{\text{sem}} \mathbf{CflIt}$, allowing us to fill the gap. Firstly, they showed that for semantic δ and for any interaction operator β, we have $[\mathcal{R}\mathbf{Txt}\beta\delta] = [\mathbf{Txt}\beta\delta]$. The same proof they employed may be used to show that this also holds for informants.

▶ **Theorem 3.3.** For any interaction operator β with $\mathbf{CflIt} \preccurlyeq \beta \preccurlyeq \mathbf{G}$ and semantic restriction δ holds $[\mathcal{R}\mathbf{Inf}\beta\delta] = [\mathbf{Inf}\beta\delta]$.   ◀

We can combine this with our observations for delayable learning restrictions, to get the final result.

▶ **Theorem 3.4.** For all learning restrictions δ that are delayable and semantic, for all interaction operators β with $\mathbf{CflIt} \preccurlyeq \beta \preccurlyeq \mathbf{G}$ we have $[\mathcal{R}\mathbf{Inf}\beta\delta] = [\mathbf{Inf_{can}G}\delta]$.   ◀

Proof. One inclusion is trivial. Since δ is delayable, Theorem 3.1 gives us that $[\mathbf{Inf_{can}G}\delta] = [\mathbf{InfSd}\delta]$. We can use that δ is semantic and $\mathbf{CflIt} \cong_{\text{sem}} \mathbf{Sd}$ to get $[\mathbf{InfSd}\delta] = [\mathbf{InfCflIt}\delta]$. By $\mathbf{CflIt} \preccurlyeq \beta$ we get $[\mathbf{InfCflIt}\delta] \subseteq [\mathbf{Inf}\beta\delta]$. Finally, Theorem 3.3 gives us $[\mathbf{Inf}\beta\delta] = [\mathcal{R}\mathbf{Inf}\beta\delta]$.   ■

▶ **Corollary 3.5.** For all $\mathbf{CflIt} \preccurlyeq \beta \preccurlyeq \mathbf{G}$ and all learning restrictions $\delta \in \Delta \setminus \{\mathbf{Cons}\}$ we have $[\mathbf{Inf}\beta\delta\mathbf{Bc}] = [\mathbf{Inf_{can}G}\delta\mathbf{Bc}]$.   ◀

3.3 Separation from explanatory learning

Kötzing and Schirneck [KS16] observed that $[\mathbf{TxtSdSMonBc}] \setminus [\mathbf{TxtGEx}] \neq \emptyset$. We modify this proof to work on informants, giving us a non-constructive proof that is more direct than the one by Aschenbach et al. [AKS18].

▶ **Theorem 3.6 ([AKS18]).** We have $[\mathbf{InfSdSMonBc}] \setminus [\mathbf{InfGEx}] \neq \emptyset$.   ◀

Proof. We consider the following **Sd**-learner h, which maps a finite set $D \subseteq \mathbb{N} \times \{0, 1\}$ to the hypothesis

$$W_{h(D)} = \begin{cases} \emptyset, & \text{if pos}(D) = \emptyset, \\ W_{\max(\text{pos}(D))}, & \text{otherwise.} \end{cases}$$

Let $\mathcal{L} = \textbf{InfSdSMonBc}(h)$. By Theorem 3.2 we have $[\textbf{InfGEx}] = [\mathcal{R}\textbf{InfSdEx}]$. By way of contradiction, assume that $\mathcal{L} \in [\mathcal{R}\textbf{InfSdEx}]$ as witnessed by some learner $g \in \mathcal{R}$. Using **ORT**, we get a computable sequence $(D_i)_{i \in \mathbb{N}}$ of sets as well as a total recursive function $p \in \mathcal{R}$ strongly monotone satisfying for all finite $D \subseteq \mathbb{N} \times \{0, 1\}$ and $b \in \{0, 1\}, i, t \in \mathbb{N}$ [1]

$$p(D, b) > \sup(\text{outline}(D)),$$

$$W_{p(D,b)} = \{p(D, b)\} \cup \bigcup_{j \in \mathbb{N}; D_j \downarrow} \text{pos}(D_j),$$

$$\text{succ}(D, b, t) = \text{content}\Big(\hat{I}_{\text{pos}(D) \cup \{p(D,b)\}}[p(D, b) + t]\Big),$$

$$D_0 = \emptyset,$$

$$D_{i+1} = \begin{cases} \text{succ}(D_i, b, t), & \text{if } \exists b \in \{0, 1\}, t \in \mathbb{N} : g(D_i) \neq g(\text{succ}(D_i, 0, t)), \\ \bot, & \text{otherwise.} \end{cases}$$

In every iteration of the $(D_i)_{i \in \mathbb{N}}$ sequence, the canonical informant describes a finite language that is one element larger than in the previous iteration. By our first restriction, the previous set did not include any information about the newly added element and thus, the new set does not contradict the previous one. The t parameter modifies how much additional negative information is included.

As each D_i matches the content of a prefix of a canonical informant, D_i is included in the input sequence for an **Sd**-learner that would result from a canonical informant for a superset of D_i. Consider the language $L = \bigcup_{i \in \mathbb{N}; D_i \downarrow} \text{pos}(D_i)$.

Case 1: L is infinite.   We have for any $p(D, b) \in L$ that

$$W_{p(D,b)} = \{p(D, b)\} \cup \bigcup_{i \in \mathbb{N}} \text{pos}(D_i) = \{p(D, b)\} \cup L = L.$$

---

[1]   For convenience, we suppose that $\sup(\emptyset) = -\infty$.

As h only outputs numbers that are contained in L, $L \in \mathcal{L}$. On the other hand, g cannot learn L from the canonical informant for L as it makes infinitely many mind changes by the construction of $(D_i)_{i \in \mathbb{N}}$.

Case 2: L is finite. That means, from one point on the sets D_i are undefined. Let D_k be the last set defined. As we have for $b \in \{0, 1\}$ that

$$\sup(L) = \sup(\mathrm{pos}(D_k)) \leq \sup(\mathrm{outline}(D_k)) < p(D_k, b),$$

we know $p(D_k, b) \notin L$. Consider the following two proper supersets of L:

$$L_0 = \{p(D_k, 0)\} \cup L,$$
$$L_1 = \{p(D_k, 1)\} \cup L.$$

For $b \in \{0, 1\}$, let $I \in \mathbf{Inf}(L_b)$ be an informant for L_b. When I shows its first positive information, h conjectures the set L until the positive information about $\max(L_b) = p(D_k, b)$ is presented. Now h switches to its final and correct guess $W_{p(D_k, b)} = L_b$. This sequence of hypotheses fulfills **SMon**, so $L_0, L_1 \in \mathcal{L}$.

From the definition of the sequence $(D_i)_{i \in \mathbb{N}}$, we know that g converges on the canonical informants for L_0 and L_1 to the same hypothesis, namely $g(D_k)$. Hence, g cannot learn both languages L_0 and L_1 as they are different, a contradiction. ∎

4   Cautiousness

In this section, we observe that, for learning from informants, requiring learners to be cautious poses a proper restriction. In particular, we find that all three reduced variants of cautiousness properly restrict learners. This is in contrast to what has been found for text-learning by Kötzing and Palenta [KP16] where $\mathbf{Caut}_\infty$ does not lessen learning power.

▶ **Theorem 4.1.** We have $[\mathbf{InfGMonBc}] \setminus [\mathbf{InfGCaut_{Tar}Bc}] \neq \emptyset$ and $[\mathbf{InfGMonBc}] \setminus [\mathbf{InfGCaut_\infty Bc}] \neq \emptyset$. ◀

Proof. This proof is analogous to the separation of **Caut** from **Mon** by Aschenbach et al. [AKS18]. Consider $\mathcal{L} = \{\mathbb{N} \setminus D \mid D \subseteq \mathbb{N} \text{ and } D \text{ is finite}\}$. It is easy to see that $\mathcal{L}$ is $\mathbf{InfGMonBc}$-learnable by the learner h that maps a finite sequence $\sigma \in \mathbb{Seq}(\mathbb{N} \times \{0,1\})$ to the hypothesis $W_{h(\sigma)} = \mathbb{N} \setminus \text{neg}(\sigma)$.

Let g be a $\mathbf{InfGBc}$-learner for $\mathcal{L}$. As $\mathbb{N} \in \mathcal{L}$, there is n_0 such that for all $n \geq n_0$ we have $W_{g(\hat{I}_\mathbb{N}[n_0])} = \mathbb{N}$. Let $L = \mathbb{N} \setminus \{n_0 + 1\}$. Then, $\hat{I}_L[n_0] = \hat{I}_\mathbb{N}[n_0]$ and thus $W_{g(\hat{I}_L[n_0])} = \mathbb{N} \supsetneq L$. As $L = \text{pos}(\hat{I}_L)$ and as L is infinite, g cannot learn $\mathcal{L}$ while preserving $\mathbf{Caut_{Tar}}$ or $\mathbf{Caut_\infty}$. ∎

▶ **Corollary 4.2.** We have $[\mathbf{InfGMonBc}] \setminus [\mathbf{InfGCautBc}] \neq \emptyset$. ◀

▶ **Theorem 4.3.** We have $[\mathbf{InfGBc}] \setminus [\mathbf{InfGCaut_{Fin}Bc}] \neq \emptyset$. ◀

Proof. Consider $\mathcal{L} = \{\mathbb{N}\} \cup \{D \subseteq \mathbb{N} \mid D \text{ is finite}\}$. $\mathcal{L}$ is $\mathbf{InfGBc}$-learnable by the leaner h that maps a finite sequence $\sigma \in \mathbb{Seq}(\mathbb{N} \times \{0,1\})$ to the hypothesis

$$W_{h(\sigma)} = \begin{cases} \mathbb{N}, & \text{if neg}(\sigma) = \emptyset, \\ \text{pos}(\sigma), & \text{otherwise.} \end{cases}$$

The proof that $\mathcal{L} \notin [\mathbf{InfGCaut_{Fin}Bc}]$ is analogous to Theorem 4.1, with the modification that we return to the finite set of all the positive information shown so far after the learner conjectures $\mathbb{N}$. ∎

5

<div align="right">

Monotonicity

</div>

In this section, we investigate the nine variants of monotonicity. We start by finding that weak monotonicity does not properly restrict learners. This is in line with what has been found for explanatory learning. For example, Aschenbach et al. [AKS18] showed that [**InfGEx**] = [**InfGWMonEx**]. In Theorem 5.1 we present a new construction for a weakly monotonic learner that is also globally consistent.

Afterwards, we compare the classic and dual versions of monotonicity and strong monotonicity. We find that in both cases, the sets of learnable language classes are incomparable (see Theorem 5.4 and Theorem 5.7). Then, we separate the classic versions from one another (see Corollary 5.8 and Corollary 5.12) and show that strong monotonicity and dual strong monotonicity imply combined monotonicity (see Theorem 5.11). For a map of all relations, see Figure 1.1.

▶ **Theorem 5.1.** We have [**InfGBc**] = [τ(**Cons**)**InfGWMonBc**]. ◀

Proof. One inclusion holds by definition. For the other, let h be a **G**-learner and $\mathcal{L} \subseteq$ **InfGBc**(h). By Theorem 3.2, we can assume that h is total. Consider the learner g that maps a finite sequence $\sigma \in \mathbb{S}\text{eq}(\mathbb{N} \times \{0, 1\})$ to the hypothesis

$$E(\sigma) = \{h(\sigma)\} \cup \{g(\tau) \mid \tau \subsetneq \sigma\},$$

$$W_{g(\sigma)} = \text{pos}(\sigma) \cup \bigcup_{e \in E(\sigma)} \bigcup_{t \in \mathbb{N}} \begin{cases} W_e^t, & \text{if } \mathbf{Cons}(W_e^t, \sigma), \\ \emptyset, & \text{otherwise.} \end{cases}$$

Intuitively, we use h's hypothesis and, in order to remain weakly monotonic, collect all our previous hypotheses which are consistent with our current information.

Clearly, g is globally consistent. Let $L \in \mathcal{L}$ and $I \in \mathbf{Inf}(L)$. To show weak monotonicity, let $i, j \in \mathbb{N}$ with $i < j$ and suppose $\mathbf{Cons}(g(I[i]), I[j])$. Then, for all but finitely many $t \in \mathbb{N}$, we have $\mathbf{Cons}(W_{g(I[i])}^t, I[j])$ and, as $g(I[i]) \in E(I[j])$, we have $W_{g(I[i])} \subseteq W_{g(I[j])}$.

Finally, we show that g **Bc**-identifies L. Let $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $W_{h(I[n])} = L$. As we include h's hypotheses, we have for all $n \geq n_0$ that $L \subseteq W_{g(I[n])}$. It remains to show that the possibly infinitely many wrong elements included in $W_{g(I[n_0])}$ are sorted out at some point. To do so, we first observe that there is a point n_2 at which all wrong hypotheses among the first n_0 hypotheses

are sorted out. Then, assuming the hypothesis is still incorrect, we show that when new negative information is shown, there is some wrong element that is included in all remaining wrong hypotheses. As the enumeration of the wrong previous hypotheses stops once this element is found, only finitely many wrong elements can remain. Those are removed once they appear in the negative information of I. We proceed with the formal proof.

Since we have, for all $n \geq n_0$, that $W_{h(I[n])} = L$, no new wrong elements are introduced after the first n_0 hypotheses. Thus for all $n \geq n_0$ we have $W_{g(I[n])} \supseteq W_{g(I[n+1])}$. Firstly, let $n_1 \geq n_0$ be such that for all $n < n_0$ with $W_{g(I[n])} \not\subseteq L$ there is an element contradicting the consistency of $g(I[n])$ in $\text{neg}(I[n_1])$. As enumeration of these hypotheses stops once the wrong elements are witnessed, only finitely many wrong elements from guesses before n_0 remain in hypotheses $W_{g(I[n])}$ for $n \geq n_1$. Let n_2 be such that all of those are contained in $\text{neg}(I[n_2])$. If $W_{g(I[n_2])} = L$, we are done. Otherwise, it remains to show that the wrong elements from $W_{g(I[n])}$ for $n_1 \leq n < n_2$ that are still included in $W_{g(I[n_2])}$ are eventually removed.

Since I presents each element from $\overline{L}$ at some point, there is $n_3 > n_2$ with $W_{g(I[n_3-1])} \supsetneq W_{g(I[n_3])}$. As g is weakly monotonic, this implies that $g(I[n_3 - 1])$ is not consistent with $I[n_3]$. Since $\text{pos}(I[n_3]) \subseteq L \subseteq W_{g(I[n_3-1])}$, there is $x \in \text{neg}(I[n_3]) \cap W_{g(I[n_3-1])}$. Since we have for all n with $n_0 \leq n$ that $W_{g(I[n])} \supseteq W_{g(I[n+1])}$, we have for all n with $n_0 \leq n < n_3$ that $x \in W_{g(I[n])}$. Let

$$t_x = \max\{\min\{t \in \mathbb{N} \mid x \in W^t_{g(I[n])}\} \mid n_0 \leq n < n_3\}.$$

Then, for $t \geq t_x$ and $\tau \sqsubseteq I[n_3]$, all $W^t_{g(\tau)}$ are inconsistent with $I[n_3]$. Hence, $W_{g(I[n_3])}$ consists of only $W_{h(I[n_3])} = L$ and at most t_x many wrong elements from previous guesses. Let $n_4 > n_3$ such that all of those wrong elements are included in $\text{neg}(I[n_4])$, then for all $n \geq n_4$ we have $W_{g(I[n])} = L$. Therefore, g identifies L and **Bc**-learns $\mathcal{L}$. ∎

5.1 Separation of classic and dual variants

In this section, we show that for both monotonicity and strong monotonicity, the set of learnable languages classes by the classic and dual variants are incomparable. This is a key observation, as it suffices to conclude the complete map. All separations presented in this section are topological and transferred from analogous proofs by Lange and Zeugmann [LZ94] where they occurred in the setting of explanatory learning of indexed families.

▶ **Lemma 5.2.** We have $[\textbf{InfGSMonBc}] \setminus [\textbf{InfGSMon}^d\textbf{Bc}] \neq \emptyset$.   ◀

Proof. Consider $\mathcal{L} = \{D \subseteq \mathbb{N} \mid D \text{ finite}\}$. The set can be learned by the strong monotonic learner $h = \text{pos}$. It cannot be learned by a dual strong monotonic learner. Intuitively, to infer a language $L \in \mathcal{L}$, a learner has to conjecture a label for L at some point. When making this guess, there is some element $x \notin L$ for which no information has been given so far. Under dual strong monotonicity, the learner cannot include x in its later hypotheses. Thus, it cannot learn $L \cup \{x\} \in \mathcal{L}$. ∎

▶ **Lemma 5.3.** We have $[\mathbf{InfGSMon}^d\mathbf{Bc}] \setminus [\mathbf{InfGSMonBc}] \neq \emptyset$. ◀

Proof. Consider $\mathcal{L} = \{\mathbb{N}\} \cup \{\{0, 1, \ldots, n\} \mid n \in \mathbb{N}\}$. The learner h, that maps a finite sequence $\sigma \in \mathbb{Seq}(\mathbb{N} \times \{0, 1\})$ to the hypothesis

$$
W_{h(\sigma)} = \begin{cases} \mathbb{N}, & \text{if } \text{neg}(\sigma) = \emptyset, \\ \{0, 1, \ldots, \min(\text{neg}(\sigma)) - 1\}, & \text{otherwise} \end{cases}
$$

learns $\mathcal{L}$ dual strong monotonically.

$\mathcal{L}$ cannot be learned strong monotonically though. Suppose there is a learner g that $\mathbf{InfSMonBc}$-learns $\mathcal{L}$. Let $n \in \mathbb{N}$ be such that $W_{g(\hat{I}_{\mathbb{N}}[n])} = \mathbb{N}$. Then g cannot infer $\{0, 1, \ldots, n + 1\}$ from its canonical informant, because it guesses a program for $\mathbb{N}$ after seeing the first n elements. Hence, $\mathcal{L} \notin [\mathbf{InfSMonBc}]$. ∎

▶ **Theorem 5.4.** We have $[\mathbf{InfSMonBc}] \# [\mathbf{InfSMon}^d\mathbf{Bc}]$. ◀

Proof. By *Lemma 5.2* and *Lemma 5.3*. ∎

▶ **Lemma 5.5.** We have $[\mathbf{InfGMonBc}] \setminus [\mathbf{InfGMon}^d\mathbf{Bc}] \neq \emptyset$. ◀

Proof. For all $i \in \mathbb{N}$, let $a_i = 3i$, $b_i = 3i + 1$ and $c_i = 3i + 2$. For all $n, m \in \mathbb{N}$ with $n < m$, consider the languages

$$
\begin{aligned}
X &= \{a_i \mid i \in \mathbb{N}\}, \\
Y_n &= \{a_i \mid i \leq n\} \cup \{b_i \mid n < i\}, \\
Z_{n,m} &= \{a_i \mid i \leq n\} \cup \{b_i \mid n < i \leq m\} \cup \{c_m\}.
\end{aligned}
$$

Let $\mathcal{L} = \{X\} \cup \{Y_n \mid n \in \mathbb{N}\} \cup \{Z_{n,m} \mid n < m\}$. Intuitively, the languages in $\mathcal{L}$ describe the contents of streams that start out listing as, then switch to bs and then end with a c. However, these streams may also continue listing as or bs indefinitely.

Consider the learner h that maps a finite sequence $\sigma \in \mathbb{Seq}(\mathbb{N} \times \{0, 1\})$ to the

hypothesis

$$
W_{h(\sigma)} = \begin{cases} Z_{n,m}, & \text{if } \exists n, m \in \mathbb{N}\colon \{a_n, b_{n+1}, c_m\} \subseteq \text{pos}(\sigma), \\ Y_n, & \text{else if } \exists n \in \mathbb{N}\colon \{a_n, b_{n+1}\} \subseteq \text{pos}(\sigma), \\ X, & \text{otherwise.} \end{cases}
$$

Intuitively, h conjectures X until the boundary between the as and bs is included in the information. Then, it conjectures the according Y_n, until a c_m is presented. This hypothesis sequence is monotonic, $\mathcal{L} \in [\mathbf{InfGMonBc}]$.

Suppose that $\mathcal{L}$ is $\mathbf{InfGMon}^d\mathbf{Bc}$-learnable as witnessed by some learner g. For some $n, m \in \mathbb{N}$, we let g infer $Z_{n,m}$, but force it to conjecture X and Y_n before $Z_{n,m}$. To break dual monotonicity, we show that there is an element $b \notin Z_{n,m}$, which will be included in Y_n, but not in X.

Let $I_X \in \mathbf{Inf}(X)$. Then there is $n_X \in \mathbb{N}$ such that $W_{g(I_X[n_X])} = X$. Let $n \in \mathbb{N}$ be such that for all $i \geq n$ we have $a_i, b_i, c_i \notin \text{outline}(I_X[n_X])$. This means, that g conjectures X without knowing whether $\text{pos}(I_X)$ really describes an infinite stream of as, or whether it may change to list bs instead. Let I_Y be an informant for Y_n such that $I_X[n_X] = I_Y[n_X]$. Such an informant exists, because by definition of n, we know $\text{pos}(I_X[n_X]) \subseteq \{a_i \mid i < n\}$ and $\text{neg}(I_X[n_X]) \cap \{b_i \mid i \in \mathbb{N}\} \subseteq \{b_i \mid i < n\}$. Let $n_Y \in \mathbb{N}$ with $n_Y > n_X$ be such that $W_{g(I_Y[n_Y])} = Y_n$. Similarly to n, let $m \in \mathbb{N}$ be such that $m > n$ and for all $i \geq m$ we have $a_i, b_i, c_i \notin \text{outline}(I_Y[n_Y])$ and let $I_Z \in \mathbf{Inf}(Z_{n,m})$ with $I_Y[n_Y] = I_Z[n_Y]$. Given I_Z, g first conjectures X, then Y and finally Z. We have $b_{m+1} \notin X$ and $b \in Y$. As $b_{m+1} \notin Z$, dual monotonicity is violated, a contradiction. ∎

▶ **Lemma 5.6.** We have $[\mathbf{InfGMon}^d\mathbf{Bc}] \setminus [\mathbf{InfGMonBc}] \neq \emptyset$. ◀

Proof. For all $n, m \in \mathbb{N}$ with $n < m$, consider the languages

$$
\begin{aligned}
X &= 2\mathbb{N}, \\
Y_n &= \{2n + 1\} \cup \{2i \mid i \leq n\}, \\
Z_{n,m} &= Y_n \cup \{2m\}.
\end{aligned}
$$

Let $\mathcal{L} = \{X\} \cup \{Y_n \mid n \in \mathbb{N}\} \cup \{Z_{n,m} \mid n, m \in \mathbb{N}, n < m\}$. Consider the learner h that maps a finite sequence $\sigma \in \mathbb{Seq}(\mathbb{N} \times \{0, 1\})$ to the hypothesis

$$
W_{h(\sigma)} = \begin{cases} Z_{n,m}, & \text{if } \exists n, m \in \mathbb{N}\colon n < m \wedge \{2n, 2n + 1, 2m\} \subseteq \text{pos}(\sigma), \\ Y_n, & \text{else if } \exists n \in \mathbb{N}\colon \{2n, 2n + 1\} \subseteq \text{pos}(\sigma), \\ X, & \text{otherwise.} \end{cases}
$$

Intuitively, h conjectures X until for some $n \in \mathbb{N}$, we find $2n+1$ in the positive data, suggesting that the target language is Y_n. Then, if another even number $2m > 2n+1$ is found, h switches to $Z_{n,m}$. This hypothesis sequence is dual monotonic, so $\mathcal{L} \in [\textbf{InfGMon}^d\textbf{Bc}]$.

Suppose that $\mathcal{L}$ is **InfGMonBc**-learnable as witnessed by some learner g. For some $n, m \in \mathbb{N}$, we let g infer $Z_{n,m}$, but force it to conjecture X and Y_n before $Z_{n,m}$. To break monotonicity, we show that the number $2m \in Z_{n,m}$ will be included in X, but not in Y_n.

Let $n_X \in \mathbb{N}$ such that $W_{g(\hat{I}_X[n_X])} = X$. Let $n = n_X + 1$ and $Y = Y_n$. Now, let $n_Y > n_X$ such that $W_{g(\hat{I}_Y[n_Y])} = Y$. Note that $\hat{I}_X[n_X] = \hat{I}_Y[n_X]$, as both include all even numbers and exclude all odd numbers up to $2n_X$. Finally, let $m = n_Y + 1$ and $Z = Z_{n,m}$. Let $n_Z > n_Y$ such that $W_{g(\hat{I}_Z[n_Z])} = Z$. Again, $\hat{I}_Y[n_Y] = \hat{I}_Z[n_Y]$. This means when inferring $Z \in \mathcal{L}$ from its canonical informant, g conjectures X, then Y and then Z. We have $2m \in X$ and $2m \notin Y$. As $2m \in Z$, monotonicity is violated, a contradiction. ∎

▶ **Theorem 5.7.** We have $[\textbf{InfGMonBc}]\#[\textbf{InfGMon}^d\textbf{Bc}]$. ◀

Proof. By Lemma 5.5 and Lemma 5.6. ∎

5.2 Completing the picture of monotonic constraints

In this section, we collect all the other relations between monotonic learning restrictions. In particular, we show that both variants of strong monotonicity imply combined monotonicity. All other theorems are implied by our previous separations.

▶ **Corollary 5.8.** We have $[\textbf{InfGMonBc}] \subsetneq [\textbf{InfGBc}]$ as well as $[\textbf{InfGMon}^d\textbf{Bc}] \subsetneq [\textbf{InfGBc}]$. ◀

Proof. By definition, both $[\textbf{InfGMonBc}]$ and $[\textbf{InfGMon}^d\textbf{Bc}]$ are subsets of $[\textbf{InfGBc}]$. As they are incomparable by Theorem 5.7, neither of them can be equal to $[\textbf{InfGBc}]$ though. Therefore, both are proper subsets. ∎

▶ **Corollary 5.9.** We have $[\textbf{InfGMon}^\&\textbf{Bc}] \subsetneq [\textbf{InfGMonBc}]$ as well as $[\textbf{InfGMon}^\&\textbf{Bc}] \subsetneq [\textbf{InfGMon}^d\textbf{Bc}]$. ◀

Proof. The reasoning is the same as for Corollary 5.8. ∎

▶ **Corollary 5.10.** We have $[\mathbf{InfGSMon}^\&\mathbf{Bc}] \subsetneq [\mathbf{InfGSMonBc}]$ as well as $[\mathbf{InfGSMon}^\&\mathbf{Bc}] \subsetneq [\mathbf{InfGSMon}^d\mathbf{Bc}]$.   ◀

Proof. The reasoning is the same as for Corollary 5.8.   ■

▶ **Theorem 5.11.** We have $[\mathbf{InfGSMonBc}] \subsetneq [\mathbf{InfGMon}^\&\mathbf{Bc}]$ as well as $[\mathbf{InfGSMon}^d\mathbf{Bc}] \subsetneq [\mathbf{InfGMon}^\&\mathbf{Bc}]$.   ◀

Proof. Consider a **G**-learner h and $\mathcal{L} = \mathbf{InfGSMonBc}(h)$. Let $L \in \mathcal{L}$ and $I \in \mathbf{Inf}(L)$. Since h is strongly monotonic, we have for all $t \in \mathbb{N}$ that $W_{h(I[t])} \subseteq L$. We get $\mathrm{neg}(I[t]) \subseteq \overline{L} \subseteq \overline{W_{h(I[t])}}$ and thus $\overline{W_{h(I[t])}} \cap \mathrm{neg}(I[t]) = \mathrm{neg}(I[t])$. For $s, t \in \mathbb{N}$ with $s \leq t$ we have

$$\overline{W_{h(I[s])}} \cap \mathrm{neg}(I[s]) = \mathrm{neg}(I[s]) \subseteq \mathrm{neg}(I[t]) = \overline{W_{h(I[t])}} \cap \mathrm{neg}(I[t]).$$

Hence, **SMon** not only implies **Mon**, but also **Mon**^{d} and therefore **Mon**$^\&$. For a dual strong monotonic learner h, we can show that for all $t \in \mathbb{N}$ we have $W_{h(I[t])} \cap \mathrm{pos}(I[t]) = \mathrm{pos}(I[t])$ and consequently that h is **Mon**$^\&$. In conclusion, we know that both $[\mathbf{InfGSMonBc}]$ and $[\mathbf{InfGSMon}^d\mathbf{Bc}]$ are subsets of $[\mathbf{InfGMon}^\&\mathbf{Bc}]$. Since $[\mathbf{InfGSMonBc}]$ and $[\mathbf{InfGSMon}^d\mathbf{Bc}]$ are incomparable by Theorem 5.4, neither of them can be equal to $[\mathbf{InfGMon}^\&\mathbf{Bc}]$, so they are both proper subsets.   ■

▶ **Corollary 5.12.** We have $[\mathbf{InfMonBc}] \setminus [\mathbf{InfSMonBc}] \neq \emptyset$.   ◀

Proof. As **SMon** $\subseteq$ **Caut**, this is a direct consequence of Corollary 4.2.   ■

▶ **Corollary 5.13.** We have $[\mathbf{InfGSMon}^d\mathbf{Bc}] \subsetneq [\mathbf{InfGMon}^d\mathbf{Bc}]$.   ◀

Proof. This is a direct consequence of Theorem 5.11 and Corollary 5.10.   ■

6

Consistency

In this section, we observe that, on its own, consistency does not restrict **InfBc**-learners, as the seen data can easily be patched into hypotheses (see Theorem 6.4). Using this approach, we can also preserve all variants of monotonicity and strong monotonicity (see Theorem 6.9). For weak monotonicity, we already observed that global consistency can be assumed in Theorem 5.1. We conclude the chapter with Theorem 6.11 where we employ *poisoning* to add global consistency to a dual weakly monotonic learner.

6.1 Patching learners

When patching hypotheses with the information seen so far, we can achieve consistency, while maintaining correct hypotheses. First, we make some general observations about patched hypotheses and then use those to add consistency to a set-driven **Bc**-learner. Afterwards, we conclude that this suffices to show that all learners without further restrictions can be assumed to be globally consistent.

▶ **Definition 6.1.** We define the function patch $\in \mathcal{R}$ such that for all $e \in \mathbb{N}$ and finite sets $D \subseteq \mathbb{N} \times \{0, 1\}$ we have

$$W_{\mathrm{patch}(e,D)} = (W_e \cup \mathrm{pos}(D)) \setminus \mathrm{neg}(D).$$

◀

▶ **Lemma 6.2.** For all informants $I \in \mathbf{Inf}$ and numbers $e, t \in \mathbb{N}$, we have $\mathbf{Cons}(\mathrm{patch}(e, \mathrm{content}(I[t])), I[t])$. ◀

Proof. Let $D = I[t]$. As I is an informant, we have $\mathrm{pos}(D) \cap \mathrm{neg}(D) = \emptyset$. Hence,

$$\mathrm{pos}(D) \subseteq (W_e \setminus \mathrm{neg}(D)) \cup \mathrm{pos}(D)$$
$$= (W_e \cup \mathrm{pos}(D)) \setminus \mathrm{neg}(D)$$
$$= W_{\mathrm{patch}(e,D)}.$$

Furthermore, $W_{\mathrm{patch}(e,D)} \cap \mathrm{neg}(D) = ((W_e \cup \mathrm{pos}(D)) \setminus \mathrm{neg}(D)) \cap \mathrm{neg}(D) = \emptyset.$ ∎

▶ **Lemma 6.3.** For numbers $e \in \mathbb{N}$ and informants $I \in \mathbf{Inf}(W_e)$ we have for all t that $\mathrm{patch}(e, \mathrm{content}(I[t])) \equiv_W e$. ◀

Proof. Let $t \in \mathbb{N}$ and $D = I[t]$. As I is an informant, we have $\mathrm{pos}(D) \subseteq \mathrm{pos}(I)$ and $\mathrm{neg}(D) \cap \mathrm{pos}(I) = \emptyset$. Then we have

$$
\begin{aligned}
W_{\mathrm{patch}(e,D)} &= (W_e \cup \mathrm{pos}(D)) \setminus \mathrm{neg}(D) \\
&= (\mathrm{pos}(I) \cup \mathrm{pos}(D)) \setminus \mathrm{neg}(D) \\
&= \mathrm{pos}(I) \\
&= W_e.
\end{aligned}
$$

∎

▶ **Theorem 6.4.** We have $[\tau(\mathbf{Cons})\mathbf{InfSdBc}] = [\mathbf{InfSdBc}]$. ◀

Proof. By definition, $[\tau(\mathbf{Cons})\mathbf{InfSdBc}] \subseteq [\mathbf{InfSdBc}]$. Let h be a **Sd** learner and $\mathcal{L} = \mathbf{InfSdBc}(h)$. We use the **Sd**-learner $g \colon D \mapsto \mathrm{patch}(h(D), D)$. Lemma 6.2 yields that g is globally consistent.

Let $L \in \mathcal{L}$, $I \in \mathbf{Inf}(L)$ and $n \in \mathbb{N}$ with $W_{g(\mathrm{content}(I[n]))} = L$. Using Lemma 6.3, we have $g(\mathrm{content}(I[n])) \equiv_W h(\mathrm{content}(I[n]))$ and thus $W_{g(\mathrm{content}(I[n]))} = L$, so g is **Bc**-learning $\mathcal{L}$. ∎

Note that Theorem 6.4 is also a corollary of the far more intricate Theorem 5.1. We can easily extend Theorem 6.4 to all other interaction operators.

▶ **Theorem 6.5.** For all β with $\mathbf{Sd} \preccurlyeq \beta \preccurlyeq \mathbf{G}$ we have $[\tau(\mathbf{Cons})\mathbf{Inf}\beta\mathbf{Bc}] = [\mathbf{InfGBc}]$. ◀

Proof. One inclusion holds by definition. Using Corollary 3.5 and $\mathbf{Sd} \preccurlyeq \beta$, we have $[\mathbf{InfGBc}] = [\mathbf{InfSdBc}] = [\tau(\mathbf{Cons})\mathbf{InfSdBc}] \subseteq [\tau(\mathbf{Cons})\mathbf{Inf}\beta\mathbf{Bc}]$. ∎

Using the approach to show $\mathbf{CflIt} \cong_{\mathrm{sem}} \mathbf{Sd}$ in [KSS17], a confluently iterative learner can fall back to a globally consistent set-driven learner, yielding the following result.

▶ **Theorem 6.6.** We have $[\tau(\mathbf{Cons})\mathbf{InfCflItBc}] = [\tau(\mathbf{Cons})\mathbf{InfSdBc}]$. ◀

▶ **Theorem 6.7.** For all interaction operators β with $\mathbf{CflIt} \preccurlyeq \beta \preccurlyeq \mathbf{G}$ we have $[\tau(\mathbf{Cons})\mathbf{Inf}\beta\mathbf{Bc}] = [\mathbf{InfGBc}]$. ◀

Proof. This follows from Theorem 6.5 and Theorem 6.6. ∎

As an aside, this is in contrast to two observations for explanatory learning made by Aschenbach et al. [AKS18] that both $[\tau(\mathbf{Cons})\mathbf{InfGEx}]$ and $[\mathbf{Inf}_{\mathrm{can}}\mathbf{GConsEx}]$ are proper subsets of $[\mathbf{InfGConsEx}]$. Their proofs use the fact that global consistency can force a learner to make syntactic mindchanges that it otherwise would not have.

▶ **Corollary 6.8.** We have $[\tau(\mathbf{Cons})\mathbf{InfGBc}] = [\mathbf{Inf}_{\mathrm{can}}\mathbf{GConsBc}] = [\mathbf{InfGBc}]$.

◀

Proof. The statement $[\tau(\mathbf{Cons})\mathbf{InfGBc}] = [\mathbf{InfGBc}]$ follows directly from Theorem 6.7. For the other equality, one inclusion holds by definition and Corollary 3.5 yields $[\mathbf{Inf}_{\mathrm{can}}\mathbf{GConsBc}] \subseteq [\mathbf{Inf}_{\mathrm{can}}\mathbf{GBc}] = [\mathbf{InfGBc}]$. ∎

6.2 Preserving monotonicity constraints

We observe that our method of making learners consistent preserves some variants of monotonicity. In particular, we patch the set of restrictions

$$\Delta_M = \{\mathbf{Mon}, \mathbf{Mon}^d, \mathbf{Mon}^{\&}, \mathbf{SMon}, \mathbf{SMon}^d, \mathbf{SMon}^{\&}\}.$$

For weak monotonicity, patching does not suffice, but we come up with another solution.

▶ **Theorem 6.9.** For all $\delta \in \Delta_M$, we have $[\tau(\mathbf{Cons})\mathbf{InfSd\delta Bc}] = [\mathbf{InfSd\delta Bc}]$. ◀

Proof. One inclusion holds by definition. Let h be a $\mathbf{InfSd\delta Bc}$-learner and $\mathcal{L} = \mathbf{InfSd\delta Bc}(h)$. We use $g\colon D \mapsto \mathrm{patch}(h(D), D)$ as our modified learner. By Lemma 6.2 and Lemma 6.3, g is globally consistent and $\mathbf{Bc}$-learns $\mathcal{L}$.

We proceed to show that the additions and subtractions of the patch-function to h's hypotheses do not violate δ. Let $L \in \mathcal{L}$, $I \in \mathbf{Inf}(L)$ and $s, t \in \mathbb{N}$ with $s \leq t$. We abbreviate $S = \mathrm{content}(I[s])$ and $T = \mathrm{content}(I[t])$.

Case 1: $\delta = \mathbf{Mon}$. We use the fact that the patched-in elements continue to be patched-in and the patched-out elements are not considered in the definition of **Mon**. Since $\mathrm{neg}(S) \cap L = \emptyset$ and $\mathrm{pos}(S) \subseteq L$, we get

$$\begin{aligned}
W_{g(S)} \cap L &= W_{\mathrm{patch}(h(S),S)} \cap L \\
&= ((W_{h(S)} \cup \mathrm{pos}(S)) \setminus \mathrm{neg}(S)) \cap L \\
&= (W_{h(S)} \cup \mathrm{pos}(S)) \cap L
\end{aligned}$$

$$= (W_{h(S)} \cap L) \cup \mathrm{pos}(S).$$

The same holds for T. Since h is **Mon**, we know $W_{h(S)} \cap L \subseteq W_{h(T)} \cap L$. Using $\mathrm{pos}(S) \subseteq \mathrm{pos}(T)$, we have

$$W_{g(S)} \cap L = (W_{h(S)} \cap L) \cup \mathrm{pos}(S) \subseteq (W_{h(T)} \cap L) \cup \mathrm{pos}(T) = W_{g(T)} \cap L.$$

Case 2: $\delta = \mathbf{Mon}^d$.   For this case, the idea is similar, but we need to do slightly more work. We use De Morgan's law and the fact that $L \cap \mathrm{neg}(S) = \emptyset$ to move the L term inside. Since $\mathrm{pos}(S) \cup L = L$, we can then remove the $\mathrm{pos}(S)$ term. Finally, we use the two inclusions, one from h being $\mathbf{Mon}^d$ and the other $\mathrm{neg}(S) \subseteq \mathrm{neg}(T)$. To get back, the same steps may be applied in reverse.

$$\begin{aligned}
\overline{W_{g(S)}} \cap \overline{L} &= \overline{(W_{h(S)} \cup \mathrm{pos}(S)) \setminus \mathrm{neg}(S)} \cap \overline{L} \\
&= \overline{((W_{h(S)} \cup \mathrm{pos}(S)) \setminus \mathrm{neg}(S)) \cup L} \\
&= \overline{(W_{h(S)} \cup L) \setminus \mathrm{neg}(S)} \\
&= (\overline{W_{h(S)}} \cap \overline{L}) \cup \mathrm{neg}(S) \\
&\subseteq (\overline{W_{h(T)}} \cap \overline{L}) \cup \mathrm{neg}(T) \\
&= \overline{W_{g(T)}} \cap \overline{L}.
\end{aligned}$$

Case 3: $\delta = \mathbf{Mon}^\&$.   Both restrictions are preserved as proven above.

Case 4: $\delta = \mathbf{SMon}$.   For **SMon**, the argument is similar to case 1. Still, elements that are patched-in stay patched-in. Since h is strongly monotonic, we have for all $n \in \mathbb{N}$ that $W_{h(\mathrm{content}(I[n]))} \subseteq L$. Given that for any $n \in \mathbb{N}$, $\mathrm{neg}(I[n])$ and L are disjoint, there are never any elements to patch out of the hypotheses of h. Formally, we have

$$W_{g(S)} = W_{h(S)} \cup \mathrm{pos}(S) \subseteq W_{h(T)} \cup \mathrm{pos}(T) = W_{g(T)}.$$

Case 5: $\delta = \mathbf{SMon}^d$.   This is analogous to the case $\delta = \mathbf{SMon}$, with only the modification that now the added hypotheses are not considered and that $\mathrm{neg}(S) \subseteq \mathrm{neg}(T)$. Furthermore, since h is dual strongly monotonic and identifies L at some point, we have for all n that $L \subseteq W_{h(\mathrm{content}(I[n]))}$. This means, that our patching never adds any more elements to the hypothesis, as they are all in already. We have

$$W_{g(S)} = W_{h(S)} \setminus \mathrm{neg}(S) \supseteq W_{h(T)} \setminus \mathrm{neg}(T) = W_{g(T)}.$$

Case 6: $\delta = \text{SMon}^{\&}$.   Both restrictions are preserved as proven above.   ∎

▶ **Corollary 6.10.** For all interaction operators $\text{Sd} \preccurlyeq \beta \preccurlyeq \text{G}$ and $\delta \in \Delta_M$, we have $[\tau(\text{Cons})\text{Inf}\beta\delta\text{Bc}] = [\text{Inf}_{\text{can}}\text{G}\delta\text{Bc}]$.   ◀

Proof. One inclusion is trivial. Since δBc is delayable, we can use Theorem 3.1. With Theorem 6.9, we get

$$[\text{Inf}_{\text{can}}\text{G}\delta\text{Bc}] = [\text{InfSd}\delta\text{Bc}] = [\tau(\text{Cons})\text{InfSd}\delta\text{Bc}] \subseteq [\tau(\text{Cons})\text{Inf}\beta\delta\text{Bc}].$$

∎

Finally, we also show that we can provide consistent versions of weakly monotonic learners. In Theorem 5.1 we have already seen that we consistency is no restriction for weakly monotonic learners, because neither poses a restriction at all. Making dual weakly monotonic learners consistent is more difficult than the other variants of monotonicity. Instead of just patching in the seen data, we use a poisoning approach.

▶ **Theorem 6.11.** We have $[\tau(\text{Cons})\text{InfGWMon}^d\text{Bc}] = [\text{InfGWMon}^d\text{Bc}]$.   ◀

Proof. We know by Theorem 3.2 that $[\textbf{InfGWMon}^d\textbf{Bc}] = [\mathcal{R}\textbf{InfSdWMon}^d\textbf{Bc}]$. Let $h \in \mathcal{R}$ be a **Sd**-learner and $\mathcal{L} = \mathcal{R}\textbf{InfSdWMon}^d\textbf{Bc}(h)$. Consider the **G**-learner g that maps a finite sequence $\sigma \in \mathbb{Seq}(\mathbb{N} \times \{0, 1\})$ to the hypothesis

$$W_{g(\sigma)} = \begin{cases} \text{pos}(\sigma), & \text{if } \exists \tau \sqsubseteq \sigma\colon \text{pos}(\tau) = \text{pos}(\sigma) \wedge \text{pos}(\sigma) \nsubseteq W_{h(\text{content}(\tau))}, \\ W_{h(\text{content}(\sigma))}, & \text{else if } \textbf{Cons}(h(\text{content}(\sigma)), \sigma), \\ \mathbb{N} \setminus \text{neg}(\sigma), & \text{otherwise.} \end{cases}$$

Intuitively, we use h's hypotheses, but poison them if we can prove that they are wrong. This means that, if we observe that they are inconsistent with our positive information, we conjecture exactly this positive data, because our hypothesis then becomes inconsistent when new positive information is presented. We must then stick to this poisoned conjecture as long as no new positive information is shown. If h includes wrong elements, we blow up our hypothesis to include everything but the negative data. This hypothesis becomes inconsistent once new negative information is presented. Notably, both poisoned hypotheses are already correct if no new positive or negative information is shown, respectively.

We first show that g is actually computable. To enumerate $W_{g(\sigma)}$, the program $g(\sigma)$ does the following for some input $x \in \mathbb{N}$: Firstly, if $x \in \text{pos}(\sigma)$, it returns,

if $x \in \text{neg}(\sigma)$, it diverges. Then, it verifies that $\text{pos}(\sigma)$ is included in h's hypothesis for each $\tau \sqsubseteq \sigma$, diverging if this is not the case. Finally, it tries to find any $y \in \{x\} \cup \text{neg}(\sigma)$ that is also in $W_{h(\text{content}(\sigma))}$, diverging if there is none. Hence, g is computable by the s-m-n theorem.

First, we show that g **Bc**-learns $\mathcal{L}$. Let $L \in \mathcal{L}$ and $I \in \textbf{Inf}(L)$. Let $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $W_{h(\text{content}(I[n]))} = L$. This implies that $\textbf{Cons}(h(\text{content}(I[n])), I[n])$ and hence $W_{g(I[n])}$ is either $\text{pos}(I[n])$ or $W_{h(\text{content}(I[n]))}$. If L is finite, then there is $n_1 \geq n_0$ such that $\text{pos}(I[n_1]) = L$, so for all $n \geq n_1$ we have $W_{g(I[n])} = L$. If L is infinite, let $n_2 > n_0$ be minimal such that $\text{pos}(I[n_0]) \subsetneq \text{pos}(I[n_2])$. Then, $W_{g(I[n_2])} = W_{h(\text{content}(I[n_2]))} = L$ and we get by induction over all $n \geq n_2$ that $W_{g(I[n])} = L$.

Clearly, g is globally consistent. We proceed to show that g is dual weakly monotonic. Let $s, t \in \mathbb{N}$ with $s < t$. We abbreviate $S = I[s]$ and $T = I[t]$. Suppose $\textbf{Cons}(g(S), T)$.

Case 1: $W_{g(S)} = \text{pos}(S) \neq W_{h(\text{content}(S))}$. Therefore, there is $\tau \sqsubseteq S$ with $\text{pos}(\tau) = \text{pos}(S)$ and $\text{pos}(S) \not\subseteq W_{h(\text{content}(\tau))}$. Since $\textbf{Cons}(g(S), T)$, we have $\text{pos}(T) \subseteq W_{g(S)} = \text{pos}(S)$, so $\text{pos}(S) = \text{pos}(T)$. As $\tau \sqsubseteq S \sqsubseteq T$, we have $W_{g(T)} = \text{pos}(T) = W_{g(S)}$.

Case 2: $W_{g(S)} = W_{h(\text{content}(S))}$. If $W_{g(T)} = \text{pos}(T)$, then the assumption $\textbf{Cons}(g(S), T)$ gives us $W_{g(T)} = \text{pos}(T) \subseteq W_{g(S)}$. Suppose $W_{g(T)} \neq \text{pos}(T)$. The precondition $g(S) \equiv_W h(\text{content}(S))$ implies $\textbf{Cons}(h(\text{content}(S)), T)$. As h is dual weakly monotonic, we have $W_{h(\text{content}(S))} \supseteq W_{h(\text{content}(T))}$. Furthermore, we have $W_{h(\text{content}(S))} \cap \text{neg}(T) = \emptyset$ and thus $W_{h(\text{content}(T))} \cap \text{neg}(T) = \emptyset$. As $W_{g(T)} \neq \text{pos}(T)$, we have $\text{pos}(T) \subseteq W_{h(\text{content}(T))}$ and thus $\textbf{Cons}(h(\text{content}(T)), T)$. Therefore, $W_{g(T)} = W_{h(\text{content}(T))} \subseteq W_{h(\text{content}(S))} = W_{g(S)}$.

Case 3: $W_{g(S)} = \mathbb{N} \setminus \text{neg}(S)$. Since g is consistent and $\text{neg}(S) \subseteq \text{neg}(T)$, we have $W_{g(T)} \cap \text{neg}(S) = \emptyset$ and thus $W_{g(T)} \subseteq \mathbb{N} \setminus \text{neg}(S) = W_{g(S)}$. ∎

7 Further research

It is still unclear where dual weak monotonicity and combined weak monotonicity, also known as semantic conservativeness, should be located in our map. Making a learner dual weakly monotonic appears to require a complete consistency check with previous hypotheses before enumerating elements. This is why we do not believe that a modification similar to the one for classic weak monotonicity of learners can be achieved. The separation of $\mathbf{WMon}^d$ would also separate $\mathbf{WMon}^\&$ from $\mathbf{T}$.

▶ **Conjecture 7.1.** We have $[\mathbf{InfGWMon}^d\mathbf{Bc}] \subsetneq [\mathbf{InfGBc}]$. ◀

For $\mathbf{WMon}^\&$, we still think that the pattern of adding consistency holds. We have observed that poisoning approaches work for adding consistency to weakly monotonic[2] and dual weakly monotonic learners. To preserve combined monotonicity, it should be sufficient to add conditions to check whether either of the poisoned hypotheses has been given before.

▶ **Conjecture 7.2.** We have $[\mathbf{InfGWMon}^\&\mathbf{Bc}] = [\tau(\mathbf{Cons})\mathbf{InfGWMon}^\&\mathbf{Bc}]$. ◀

Although it appears that consistency does not further restrict any of the monotonicity constraints and we know that the implications are not as strict as for $\mathbf{Ex}$ (see Corollary 6.8), it is unclear whether an additional requirement for consistency narrows the learning power of learners under any other common restriction.

The restriction $\mathbf{SemWb}$ (semantically witness-based, [KSS17]) could reveal more about the relationship of monotonic and cautious learners, because it is designed as a common lower bound of the two. While we showed that all three reduced variants of cautiousness are properly restricting learners, we do not know how they relate to each other. Furthermore, we do not know how the variants of cautiousness relate to other restrictions. In particular, there may be an interesting relation to the monotonicity restrictions we mapped out in this work, as they all limit subset relations in the hypothesis sequence.

For semantic learning in general, there are several research directions. There is very recent work by Marten [Mar22] that investigates semantic learning restrictions in more abstract settings such as learning functions instead of languages. It would

---

2 not included, as it is implied by Theorem 5.1

also be interesting to see if stronger normal forms can be found, similar to what has been done by Kötzing et al. [KSS17] and Doskoč and Kötzing [DK21b] for text-learning.

Lastly, there are of course more learning restrictions that are still missing from our map. For example, the map for explanatory learning provided by Aschenbach et al. [AKS18] includes the semantic restrictions **Dec** (decisive) and **NU** (non-U-shaped).

Bibliography

[AKS18]   Martin Aschenbach, Timo Kötzing, and Karen Seidel. **Learning from infor-mant: relations between learning success criteria**. *arXiv preprint* (2018) (see pages 3, 9, 11, 12, 15, 17, 25, 30).

[Ang80]   Dana Angluin. **Inductive inference of formal languages from positive data**. *Information and control* 45:2 (1980), 117–135 (see page 8).

[Blu67]   Manuel Blum. **A machine-independent theory of the complexity of recursive functions**. *Journal of the ACM* 14:2 (1967), 322–336 (see page 5).

[Cas94]   John Case. **Infinitary self-reference in learning theory**. *Journal of Experimental & Theoretical Artificial Intelligence* 6:1 (1994), 3–16 (see pages 2, 5).

[CK10]   John Case and Timo Kötzing. **Strongly non-u-shaped learning results by general techniques**. In: *COLT*. Vol. 2010. Citeseer. 2010, 181–193 (see page 7).

[CS83]   John Case and Carl Smith. **Comparison of identification criteria for machine inductive inference**. *Theoretical Computer Science* 25:2 (1983), 193–220 (see page 1).

[DK21a]   Vanja Doskoč and Timo Kötzing. **Mapping monotonic restrictions in inductive inference**. In: *Conference on Computability in Europe*. Springer. 2021, 146–157 (see page 3).

[DK21b]   Vanja Doskoč and Timo Kötzing. **Normal forms for semantically witness-based learners in inductive inference**. In: *Conference on Computability in Europe*. Springer. 2021, 158–168 (see pages 4, 30).

[Gol67]   E Mark Gold. **Language identification in the limit**. *Information and control* 10:5 (1967), 447–474 (see pages 1, 6).

[Jan91]   Klaus P Jantke. **Monotonic and non-monotonic inductive inference**. *New Generation Computing* 8:4 (1991), 349–360 (see page 8).

[Kle52]   Stephen Cole Kleene. **Introduction to metamathematics** (1952) (see page 5).

[Köt17]   Timo Kötzing. **A solution to Wiehagen's thesis**. *Theory of Computing Systems* 60:3 (2017), 498–520 (see pages 1, 10).

[KP16]   Timo Kötzing and Raphaela Palenta. **A map of update constraints in inductive inference**. *Theoretical Computer Science* 650 (2016), 4–24 (see pages 1, 2, 9, 10, 15).

[KS16]   Timo Kötzing and Martin Schirneck. **Towards an atlas of computational learning theory**. In: *Symposium on Theoretical Aspects of Computer Science*. 2016 (see pages 6, 12).

[KSS17]   Timo Kötzing, Martin Schirneck, and Karen Seidel. **Normal forms in semantic language identification**. In: *International Conference on Algorithmic Learning Theory*. 2017, 493–516 (see pages 3, 4, 7, 9, 10, 12, 24, 29, 30).

[LZ94]   Steffen Lange and Thomas Zeugmann. **Characterization of language learning front informant under various monotonicity constraints**. *Journal of Experimental & Theoretical Artificial Intelligence* 6:1 (1994), 73–94 (see pages 2, 18).

[LZK96]   Steffen Lange, Thomas Zeugmann, and Shyam Kapur. **Monotonic and dual monotonic language learning**. *Theoretical Computer Science* 155:2 (1996), 365–410 (see pages 3, 8).

[Mar22]   Paula Marten. **Isomorphisms and embeddings between limit learning settings**. Bachelor's thesis. Universität Potsdam, 2022 (see page 29).

[OSW82]   Daniel N Osherson, Michael Stob, and Scott Weinstein. **Learning strategies**. *Information and Control* 53:1-2 (1982), 32–51 (see page 9).

[Rog87]   Hartley Rogers Jr. **Theory of recursive functions and effective computability**. MIT press, 1987 (see page 5).

[Sch84]   Gisela Schäfer-Richter. **Über Eingabeabhängigkeit und Komplexität von Inferenzstrategien**. PhD thesis. RWTH Aachen, 1984 (see page 6).

[Sei21]   Karen Seidel. **Modelling binary classification with computability theory**. PhD thesis. Universität Potsdam, 2021 (see page 1).

[WC80]   Kenneth Wexler and Peter W Culicover. **Formal principles of language acquisition**. MIT Press (MA), 1980 (see page 6).

[Wie90]   Rolf Wiehagen. **A thesis in inductive inference**. In: *International Workshop on Nonmonotonic and Inductive Logic*. Springer. 1990, 184–207 (see page 8).

Declaration of Authorship

I hereby declare that this thesis is my own unaided work. All direct or indirect sources used are acknowledged as references.

Potsdam, June 29, 2022   _____

                                                    Niklas Mohrin