Isomorphisms and Embeddings between Limit Learning Settings

Isomorphien und Einbettungen zwischen Limeslernszenarien

Paula Marten

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, there exist a number of possible hypothesis spaces and learning criteria to determine whether an algorithmic learner can successfully learn an object. The object to be learned can either be a function or a formal language, and in both cases there are differing forms of presenting the object. A function can be represented by its graph, which could be given to the learner either in a canonical or arbitrary order, while a language can be learned from negative and positive information or just from positive information. For all these learning settings, there also exist a variety of learning restrictions, limiting when a learner is successful. The definition of these restrictions is usually bound to the objects to be learned and how they are labeled, making the transfer of knowledge about the relationship between learning criteria and the learnability of sets of objects nearly impossible.

In this work, we formalize learning settings and criteria to be universally applicable to the different forms of inductive inference. From there, we define isomorphisms and embeddings between learning settings and criteria. These mappings between different settings and criteria now allow for a knowledge transfer between different areas of inductive inference.

We prove an isomorphism between function learning and learning infinite decidable language with **C**-indices under semantic and delayable learning restrictions. For these learning restrictions also show various embeddings between learning settings, among them between learning functions and decidable languages with **C**-indices and vice versa. We also show that there cannot exist an isomorphism between language learning under either **C**- or **W**-indices for all semantic learning restrictions.

Zusammenfassung

In der inductive inference gibt es eine Reihe möglicher Hypothesenräume und Lernkriterien, die bestimmen, ob ein algorithmischer Lerner ein Objekt erfolgreich erkennen kann. Das zu lernende Objekt kann entweder eine Funktion oder eine formale Sprache sein, und in beiden Fällen gibt es unterschiedliche Formen der Darstellung des Objekts. Eine Funktion kann durch ihren Graphen dargestellt werden, der dem Lernenden entweder in einer kanonischen oder beliebigen Reihenfolge präsentiert werden kann, während eine Sprache aus negativen und positiven Informationen oder nur aus positiven Informationen gelernt werden kann. Für all diese Lernsituationen gibt es auch eine Reihe von Lernkriterien, die einschränken, wann ein Lerner erfolgreich ist. Die Definition dieser Kriterien ist in der Regel an die zu lernenden Objekte und deren Label gebunden, was den Übertrag von Wissen über die Beziehung zwischen Lernkriterien und der Lernfähigkeit von Objektmengen nahezu unmöglich macht.

In dieser Arbeit formalisieren wir die Lernbedingungen und -kriterien so, dass sie universell auf die verschiedenen Formen der induktiven Inferenz anwendbar sind. Darauf basierend definieren wir Isomorphismen und Einbettungen zwischen Lernszenarie und -kriterien. Diese Mappings zwischen verschiedenen Szenarien und Kriterien ermöglichen nun einen Wissenstransfer zwischen verschiedenen Bereichen der induktiven Inferenz.

Wir beweisen einen Isomorphismus zwischen Funktionslernen und dem Lernen unendlicher entscheidbarer Sprache mit **C**-Indizes unter semantischen und delayable Lernrestriktionen. Für diese Lernrestriktionen zeigen wir auch verschiedene Einbettungen zwischen Lernumgebungen, unter anderem zwischen Funktionslernen und entscheidbaren Sprachen mit **C**-Indizes und umgekehrt. Wir zeigen auch, dass es für alle semantischen Lernrestriktionen keinen Isomorphismus zwischen Sprachenlernen unter **C**- oder **W**-Indizes geben kann.

Acknowledgments

I want to thank Timo and Vanja for all the work, time and patience they put into supervising this thesis, I highly appreciated all of it. Timo laid the theoretical groundwork for this thesis, always had an open ear and office door to answer my questions, saw the story of this thesis and helped me work it out with the right theorems and proofs. Vanja of course was the one who then had to had to deal with my proofs, in many iterations. Only because of his feedback the process during these iterations slowly moved from just proofreading my ideas to actual proof reading. He provided helpful feedback, correcting my thesis on the weekends and answering questions even early in the morning and late at night, and worked some serious ORT magic, making the last theorem of this thesis possible. But most importantly Timo and Vanja gave me the courage and confidence to choose a rather abstract and mathematical topic, for which I am deeply grateful.

Moreover, I would like to thank Niklas for not only proofreading this thesis, but also for having just the right theorem up his sleeve to repair my proof, the many conversations about inductive inference and the shared journey into this topic. Without his help, parts of this theses would probably just be wrecked instead of correct.

Contents

1 Introduction

Inductive inference is a branch of algorithmic learning theory investigating which functions or formal languages, that is, subsets of the natural numbers, can be recognized by an algorithmic *learner*. Regardless of whether the learner should learn functions or formal languages, it successively gets presented information about these items. The learner should then recognize the presented item and output a description, ideally for the target itself [Gol67].

Gold gave a first criterion for the successful learning of an item. *Explanatory* learning short **Ex**, is successful if a learner, given information about an item, eventually outputs a correct label for the target item and stops changing the output. With this first definition of successful learning, it becomes apparent that we can always build a learner to learn a certain item, it just has to constantly output a label of the item. This is why we focus on learning classes of items.

Learning settings describe which kind of items should be learned, how information about these items is given and to which extent the learner can access said information. To describe **Ex**-learning of languages a possible setting would be **InfGEx**-learning. In this setting the learner has full access to all the information it has seen at this point. This is denoted by **G**, which stands for Gold-style learning. The learner gets presented information about a language by an *informant*, denoted by **Inf**. Informants are sequences of tuples containing positive and negative information about the language. Recently, some work has been done investigating the relationships between certain success criteria for learning from informants [AKS18; KKS21; Moh22].

Another setting for **Ex**-learning would be **ArbGEx**. Here we learn classes of computable functions. The function to be learned will be presented by an unordered sequence over the graph of the function, we call this an *arbitrary presentation* and denote it by **Arb** [Gol67].

There are numerous additional criteria for successful learning, and different settings to learn in. The "building-blocks"-approach introduced by Kötzing [Köt09] allows for a modular definition and combination of learning settings and criteria. In this thesis, we present a unified definition of learning settings and learning criteria, which is universally applicable to different forms of inductive inference. This allows for a further generalization of the maps used to portray the relationships between different settings and criteria, as seen in Kötzing and Schirneck [KS16].

A *learning task* is a set of objects, of which we aim to learn subsets using a single learner. The *presentation system* determines how a learner receives information about the item to be learned, while the *hypothesis space* fixes a mapping of natural numbers to the items of the learning task.

These abstract definitions of learning setting do not only allow us to generalize learning restrictions to be defined independently of the hypothesis space, in which the restriction is used, but also to investigate the structural relations between different learning settings and learning criteria. For this, we define embeddings and isomorphisms of learning settings and learning criteria. Now we can deduce relations between the learning restriction in one setting from their relations in another setting, allowing us to generalize knowledge about the restrictions gained in one setting.

To use the insights gained from embeddings and isomorphisms between different learning tasks and hypothesis spaces we generalize definitions for some commonly used learning restrictions. Traditionally, a learning restriction is defined separately for each task, presentation system and hypothesis space with equivalent intention, but differing syntax. Learning restrictions themselves can fulfil certain criteria, such as being *semantic* [Köt17; KSS17]. If a learner successfully learns a language under a certain semantic restriction, another learner outputting semantically equivalent labels also successfully learns the language under this restriction. We show that the new definitions preserve the semantic properties of learning restrictions. This allows to transfer assumptions about the success of learners even across different settings.

These properties for restrictions quickly become important, as we go on to show isomorphisms between different learning settings. We show an isomorphism between all total functions and total functions with a range limited to 1 and 0 that map to 1 infinitely often for all delayable and semantic learning restrictions. Moreover, we prove an isomorphism between all total functions with a range limited to 1 and 0 and decidable languages for all semantic learning restrictions. From this, we can also deduce an isomorphism between total functions with a range limited to 1 and 0 that map to 1 infinitely often and infinite decidable languages. We use these theorems to conclude that there exists an isomorphism between total functions and infinite decidable languages for all delayable and semantic learning restrictions. We further show embeddings for both directions between decidable languages and infinite decidable languages for delayable learning restrictions. This allows for the conclusion that total functions and decidable languages embed each other.

Finally, we show an example of non-isomorphic learning settings. Depending on how we map the hypotheses given out by a learner to decidable languages and the learning restriction the relation between the set of learnable classed within the different hypothesis spaces differ. We use this to conclude that there cannot exist an isomorphism between these hypothesis spaces.

2 Preliminaries

We first introduce the mathematical and computability theory notation used in this thesis. Then we formally introduce our settings for learning in the limit and related notation.

2.1 Mathematical and computability theory notation

With $\mathbb{N}$ we denote the set of natural numbers, namely $\mathbb{N} = \{0, 1, 2, ..\}$. With $\subseteq, \supseteq, \subsetneq$ and $\supsetneq$ we denote the subset, superset, proper subset and proper superset relation. By $\emptyset$ and ϵ we denote the empty set and the empty sequence, respectively. The quantifier $\forall^\infty x$ means "for all but infinitely many x". We use $\langle ., . \rangle$ as a bijective computable pairing function $\mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$. We use π_1 and π_2 as decoding functions, so that for any $x, y \in \mathbb{N}$ we have $\pi_1(\langle x, y \rangle) = x$ and $\pi_2(\langle x, y \rangle) = y$.

Let $\mathcal{P}$ define the set of all partial computable functions and $\mathcal{R}$ the set of all total computable functions. Further $\mathcal{P}_{0,1}$ and $\mathcal{R}_{0,1}$ denote the set of all partial and total computable functions $\mathbb{N} \longrightarrow \{0, 1\}$. If a function f is not defined on x we write $f(x) \uparrow$ or $f(x) = \perp$. We fix a programming system φ for $\mathcal{P}$ and let φ_e denote the function computed by the program with the number e. We also fix a Blum complexity measure Φ for φ, so that for all $e, x \in \mathbb{N}$, $\Phi(e, x)$ describes the number of steps φ_e takes on input x to halt [Blu67].

A formal language $L \subseteq \mathbb{N}$ is *computably enumerable* if its is the domain of a computable function. Let $\mathcal{E}$ define the set of all computably enumerable sets. A set $L \subseteq \mathbb{N}$ is *computable* if there is a $f \in \mathcal{R}$ so that for all $x \in \mathbb{N}$ we have

$$f(x) = \begin{cases} 1, & \text{if } x \in L, \\ 0, & \text{otherwise.} \end{cases}$$

We call this function the characteristic function of L and denote it by χ_L. Let $\mathcal{E}_c$ denote the set of all computable sets.

For a function f and an $n \in \mathbb{N}$, $f[n]$ is the finite sequence of the n first elements of f. We denote the empty sequence as ϵ. The concatenation of to finite sequences σ, τ is denoted by $\sigma^\frown\tau$.

2.2 Inductive inference

A text T is a function $T\colon \mathbb{N} \to \mathbb{N} \cup \{\#\}$, where # resembles a break symbol. The content of a text T is defined as $\mathrm{content}(T) = \mathrm{range}(T) \setminus \{\#\}$. A text is a text for a language L if and only if $\mathrm{content}(T) = L$, we say $T \in \mathbf{Txt}(L)$.

For any set $S \subseteq \mathbb{N} \times \{0,1\}$ we define

$$\mathrm{pos}(S) = \{x \mid \langle x, 1 \rangle \in S\}$$
$$\mathrm{neg}(S) = \{x \mid \langle x, 0 \rangle \in S\}$$

An informant is as function $I\colon \mathbb{N} \to \mathbb{N} \times \{0,1\}$. The content of an informant I is defined as $\mathrm{content}(I) = \mathrm{range}(I)$. If $\mathrm{pos}(I) = L$ and $\mathrm{neg}(I) = \overline{L}$, then I is an informant for a language L [KKS21]. For any language L we denote the set of all informants for the language by $\mathbf{Inf}(L)$. An informant I is *canoncial* if and only if for all $n \in \mathbb{N}$, we have $y \in \{0,1\}$, so that $I : n \mapsto \langle n, y \rangle$. We write the set of canonical informants for a language L as $\mathbf{Inf}_{\mathrm{Can}}(L)$.

We define the graph of a function f as a set $\mathrm{graph}(f) = \{\langle x, f(x) \rangle \mid x \in \mathbb{N}\}$. A function is presented to a learner by a surjection $A\colon \mathbb{N} \to \mathrm{graph}(f)$. We denote the set of arbitrary presentations for a function f by $\mathbf{Arb}(f)$. The canonical representation of a function is defined by a bijection b so that for all $n \in \mathbb{N}$ we have $b(n) = \langle n, f(n) \rangle$. The set of canonical presentations for a function f is denoted by $\mathbf{Can}(f)$. For $e \in \mathbb{N}$ we denote the $\mathbf{W}$-hypothesis of e as $W_e = \mathrm{dom}(\varphi_e)$. Furthermore, in reference to Φ, for all $e, t \in \mathbb{N}$ we let $\mathbf{W}_e^t = \{x \leq t \mid \Phi_e(x) \leq t\}$, so $\mathbf{W}_e^t$ describes the decidable set of all number smaller or equal to t, on which φ_e halts in less than or exactly t steps. If we have $\varphi_e \in \mathcal{R}$ we say that e is a $\mathbf{C}$-index and define the $\mathbf{C}$-hypothesis of e as $\mathbf{C}_e = \{x \in \mathbb{N} \mid \varphi_e(x) = 1\}$[Ber+21]. To allow hypothesis not only for languages but also for computable functions we define the φ-hypothesis of e as φ_e.

We use learning criteria as formalized by Kötzing [Köt09]. A learner is a partial function $h \in \mathcal{P}$. An interaction operator takes a learner $h \in \mathcal{P}$ and a presentation P and outputs a function p. In Gold-style learning [Gol67] a learner has access to the full information given by a presentation, so for $h \in \mathcal{P}$, $P \in \mathbf{Inf} \cup \mathbf{Arb} \cup \mathbf{Txt}$ and $i \in \mathbb{N}$ we have

$$\mathbf{G}(h, P[i]) = h(P[i]).$$

We call $p(i) = \mathbf{G}(h, P[i])$ the hypothesis sequence of h on P.

Learning restrictions are predicates on learners and presentations and determine when a learner is successful on a presentation. The set of classes of items of type $\mathcal{X}$

which can be learned under an interaction operator β and a learning restriction $\delta_{\mathcal{X}}$ is denoted by $[(\beta, \delta_{\mathcal{X}})]_{\mathcal{X}}$.

In Section 3.2 we suggest definitions for learning restrictions which are independent from the hypothesis space in which the restrictions are used.

We define delayable learning restrictions as introduced by Kötzing and Palenta [KP16]. Let $\vec{R}$ be the set of all non-decreasing $r\colon \mathbb{N} \to \mathbb{N}$ with infinite limite inferior, i.e. for all m we have $\forall^{\infty} n\colon r(n) \geq m$. A learning restriction δ is delayable if and only if, for all presentations P and P' with $\mathrm{content}(P) = \mathrm{content}(P')$, all p and all $r \in \vec{R}$, if $(p, T) \in \delta$ and $\forall n\colon \mathrm{content}(P[r(n)]) \subseteq \mathrm{content}(P'[n])$, then $(p \circ r, P') \in \delta$.

3 Generalized learning settings

3.1 Abstract definitions of learning settings

We start with some abstract definitions, culminating in the definition of isomorphic learning tasks. First, we define learning tasks, presentations, hypothesis spaces and learning settings.

▶ **Definition 3.1.** Let a set of objects $\mathcal{X}$ be given. We call a set of objects from $\mathcal{X}$ a *set of learning tasks*; we aim to learn sets $\mathcal{L} \subseteq \mathcal{X}$ by single learners h. A *presentation system* for $\mathcal{X}$ is a function P such that, for all $x \in \mathcal{X}$, $P(x)$ is a set of sequences over $\mathbb{N}$. For each $x \in \mathcal{X}$, we call the elements of $P(x)$ the *presentations* of x. If, for all $x, y \in \mathcal{X}$, we have, if $P(x) \cap P(y) \neq \emptyset$, then $x = y$, then we call P *accurate*. In this case we write, for each $x \in \mathcal{X}$ and each $T \in P(x)$, $\mathrm{obj}(T)$ for x, the *object* described by the presentation T in the presentations system P. A *hypothesis space* for $\mathcal{X}$ is a function $H\colon \mathbb{N} \to \mathcal{X}$. We call e a hypothesis for $H(e)$. A *learning setting* is a triple $\mathcal{X} = (\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ with $\mathcal{X}$ a set of learning tasks, P a presentation system for $\mathcal{X}$ and H a hypothesis space for $\mathcal{X}$. ◀

Intuitively, given an object $x \in \mathcal{X}$, each sequence of $P(x)$ gives more and more information about x. We will only consider accurate presentations systems, where no presentation is a presentation for two different objects.

Now we define learning criteria in dependence of learning settings, especially the learning task of a setting.

▶ **Definition 3.2 (Learning criterion).** Given a learning setting $\mathcal{X} = (\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$, a *learning criterion* I is a pair $(\beta, \delta_{\mathcal{X}})$, where β is an interaction operator and $\delta_{\mathcal{X}}$ is a learning restriction defined in dependence on $\mathcal{X}$. ◀

To allow for mappings between learning criteria and settings we first define embeddings and later isomorphisms on learning settings.

▶ **Definition 3.3 (Embeddings of learning settings).** Let I, J be two learning criteria over two learning settings $(\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ and $(\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$, respectively. Suppose there is a injection $\Theta\colon \mathcal{X} \to \mathcal{Y}$ such that, for all $\mathcal{L} \subseteq \mathcal{X}$,[1]

$$\mathcal{L} \in [I]_{\mathcal{X}} \Leftrightarrow \Theta(\mathcal{L}) \in [J]_{\mathcal{Y}}.$$

We then write $I \hookrightarrow J$ and say that I *embeds* J. ◀

▶ **Theorem 3.4.** A preorder relation over the set of all learning criteria is defined by $\hookrightarrow$. ◀

Proof. Let I be a learning criteria over a learning setting $(\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$. Since id is injective and for any $X \subseteq \mathcal{X}$ we have $X \in [I]_{\mathcal{X}}$ if and only if $id(X) \in [I]_{\mathcal{X}}$, we conclude $I \hookrightarrow I$, so $\hookrightarrow$ is reflexive.

Given three learning criteria I, J and K over the learning settings $(\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$, $(\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$ and $(\mathcal{Z}, P_{\mathcal{Z}}, H_{\mathcal{Z}})$ so that $I \hookrightarrow J$ and $J \hookrightarrow K$. This means we have two injective functions $\Theta\colon \mathcal{X} \to \mathcal{Y}$ and $\Omega\colon \mathcal{Y} \to \mathcal{Z}$, so that for all $X \subseteq \mathcal{X}$ we have $X \in [I]_{\mathcal{X}}$ if and only if $\Theta(X) \in [J]_{\mathcal{Y}}$ and for all $Y \subseteq \mathcal{Y}$ we have $Y \in [J]_{\mathcal{Y}}$, if and only if $\Omega(Y) \in [K]_{\mathcal{Z}}$. Since $\Omega \circ \Theta$ is also injective and $X \in [I]_{\mathcal{Z}}$ if and only if $\Omega \circ \Theta(X) \in [K]_{\mathcal{Z}}$, we have $I \hookrightarrow K$. This means that $\hookrightarrow$ is transitive and thus $\hookrightarrow$ is a preorder relation on the set of all learning criteria. ■

Analogously to embeddings, we define isomorphisms between learning settings.

▶ **Definition 3.5 (Isomorphic learning settings).** Let I, J be two learning criteria over two learning settings $(\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ and $(\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$, respectively. Suppose there is a bijection $\Theta\colon \mathcal{X} \to \mathcal{Y}$ such that, for all $\mathcal{L} \subseteq \mathcal{X}$,

$$\mathcal{L} \in [I]_{\mathcal{X}} \Leftrightarrow \Theta(\mathcal{L}) \in [J]_{\mathcal{Y}}.$$

Then we write $I \cong J$ and call the learning settings *isomorphic*.

If, furthermore, there are computable operators Ψ, Ψ' such that any learner h for $\mathcal{L}$ is mapped to a learner $\Psi(h)$ for $\Theta(\mathcal{L})$, and any learner h' for $\Theta(\mathcal{L})$ is mapped to a learner $\Psi'(h')$ for $\mathcal{L}$, then we write $I \cong_c J$ and call the learning settings *computably isomorphic*. ◀

▶ **Theorem 3.6.** The $\cong$-relation is a equivalence relation over the set of all learning criteria and settings. ◀

1 We use the common notation of element-wise application of a function to a set, that is $\Theta(\mathcal{L}) = \{\Theta(L) \mid L \in \mathcal{L}\}$.

Proof. We get that $\cong$ is transitive and reflexive analogously to the proof of Theorem 3.4. We show that $\cong$ is symmetric. Let the learning criteria I and J over the the learning settings $(\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ and $(\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$ be given so that we have $I \cong J$. This means we have a bijection $\Theta\colon \mathcal{X} \to \mathcal{Y}$ so that for $X \subseteq \mathcal{X}$ we have $X \in [I]_{\mathcal{X}}$ if and only if $\Theta(X) \in [J]_{\mathcal{Y}}$. Since Θ is a bijection, so is its inverse function $\Theta^{-1}\colon \mathcal{Y} \to \mathcal{X}$. For an $Y \in \mathcal{Y}$ we now have $Y \in [J]_{\mathcal{Y}}$ if and only if $\Theta^{-1}(Y) \in [I]_{\mathcal{X}}$, so $J \cong I$. This shows that $\cong$ is symmetric and thus a equivalence relation. ■

▸ **Corollary 3.7.** For any learning criteria I, J, if we have $I \cong J$ we also have $I \hookrightarrow J$. ◂

Proof. This follows directly from Definition 3.3and Definition 3.5 ■

3.2 Definitions of learning restrictions

Learning restrictions are typically given for a concrete hypothesis space and a concrete application. In this section we generalize learning restrictions to the abstract setting. Let $\mathcal{X} = (\mathcal{X}, P, H)$ be a learning setting.

▸ **Definition 3.8 (Consistency).** Given an object $x \in \mathcal{X}$ and information given so far σ, we say that x *is consistent with* σ, denoted as $\mathrm{Cons}(x, \sigma)$ if and only if

$$\exists T \in P(x)\colon T[|\sigma|] = \sigma.$$

Overloading notation, given a hypothesis e and and information given so far σ, we say that e *is consistent with* σ, denoted as $\mathrm{Cons}(e, \sigma)$, if and only if $\mathrm{Cons}(H(e), \sigma)$. ◂

Intuitively, a conjecture is consistent if it is a conjecture for an object which might be the target object, given the information σ. With this definition we can now generalize notions based on consistency as follows.

We define the consistency of hypothesis sequences p on a presentation T so that

$$\mathbf{Cons}(p, T) \Leftrightarrow \forall i \in \mathbb{N}\colon \mathrm{Cons}(p(i), T[i]).$$

If this claim holds for p, we call p *consistent.* Intuitively every hypothesis is consistent with the data seen so far [Ang80].

We define the learning restriction **Conv** so that for a hypothesis sequence p on a presentation T we have

$$\mathbf{Conv}(p, T) \Leftrightarrow \forall i, j \in \mathbb{N}\colon i < j \land \mathrm{Cons}(p(i), T[j]) \to p(i) = p(j).$$

We say p is *conservative* on T, intuitively the learner never changes a consistent hypothesis [Ang80].

For each object $x \in \mathcal{X}$, we let $P^{\text{seq}}(x)$ be the set of possible starting sequences of presentations for x, that is

$$P^{\text{seq}}(x) = \{T[i] \mid T \in P(x), i \in \mathbb{N}\}.$$

We call an presentation system *finitely accurate*, if and only if for all $x, y \in \mathcal{X}$ if $P^{\text{seq}}(x) = P^{\text{seq}}(y)$ then $x = y$.

We define an order on objects of a learning task, depending on their presentations.

▶ **Definition 3.9 (Information Order).** For $x, y \in \mathcal{X}$, we define

$$x \preceq_{\mathcal{X}} y \Leftrightarrow P^{\text{seq}}_{\mathcal{X}}(x) \subseteq P^{\text{seq}}_{\mathcal{X}}(y).$$

◀

Whenever it is clear from the context, we might omit the $\mathcal{X}$ and just write $\preceq$. Intuitively, any finite information given about x might as well be information about y.

▶ **Theorem 3.10.** Given a set of objects $\mathcal{X}$ and a finitely accurate presentation system, then $\preceq_{\mathcal{X}}$ is a partial order relation on $\mathcal{X}$. ◀

Proof. Since for any $x \in \mathcal{X}$ we have $P^{\text{seq}}_{\mathcal{X}}(x) = P^{\text{seq}}_{\mathcal{X}}(x)$, so $x \preceq_{\mathcal{X}} x$, showing that $\preceq_{\mathcal{X}}$ is reflexive. Given $x, y \in \mathcal{X}$ so that $x \preceq y$ and $y \preceq x$ we have $P^{\text{seq}}_{\mathcal{X}}(x) = P^{\text{seq}}_{\mathcal{X}}(y)$. Since P is finitely accurate we have $x = y$, so $\preceq$ is antisymmetric. Given $x, y, z \in \mathcal{X}$ so that $x \preceq_{\mathcal{X}} y$ and $y \preceq_{\mathcal{X}} x$ we have $P^{\text{seq}}_{\mathcal{X}}(x) \subseteq P^{\text{seq}}_{\mathcal{X}}(y)$ and $P^{\text{seq}}_{\mathcal{X}}(y) \subseteq P^{\text{seq}}_{\mathcal{X}}(x)$. The transitivity of the $\subseteq$ relation gives us $P^{\text{seq}}_{\mathcal{X}}(x) \subseteq P^{\text{seq}}_{\mathcal{X}}(z)$ and with this $x \preceq_{\mathcal{X}} z$, so $\preceq_{\mathcal{X}}$ is also transitive and thus a partial order relation. ■

With this definition we can now generalize a lot of notions typically based on set inclusion, which are based on Jain et al. [Jai+99] and Jantke [Jan91], as follows.

We define the learning restriction of strong monotonicity, so that for a hypothesis sequence p on a presentation T, we have

$$\mathbf{SMon}(p, T) \Leftrightarrow \forall i, j \in \mathbb{N}\colon i \leq j \rightarrow H(p(i)) \preceq H(p(j)).$$

We call p *strongly monotone*. Intuitively the hypothesis given out by a learner never get smaller.

We define weak monotonicity so that for a hypothesis sequence p on a presentation T, we have

$$\mathbf{WMon}(p,T) \Leftrightarrow \forall i,j \in \mathbb{N}\colon i \le j \wedge \mathrm{Cons}(p(i),T[j]) \to H(p(i)) \preceq H(p(j)).$$

If a hypothesis sequence p with presentation T fulfills this criterion we call it *weakly monotone*, intuitively the learner can only enlarge consistent hypotheses.

We define cautiousness so that for a hypothesis sequence p on a presentation T we have

$$\mathbf{Caut}(p,T) \Leftrightarrow \forall i,j \in \mathbb{N}\colon i \le j \to \neg(H(p(j)) \prec H(p(i))).$$

We call p *cautious*, intuitively the learner never proceeds to a conjecture with strictly less finite information.

We now get to an important property of learning restrictions, namely being semantic. For a sequence σ we let

$$\mathrm{obj}(\sigma) = \{x \in \mathcal{X} \mid \mathrm{Cons}(x,\sigma)\}$$

be the set of objects consistent with σ.

▶ **Definition 3.11 (Semantic Restriction).** Let learning settings $\mathcal{X} = (\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ and $\mathcal{Y} = (\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$ be given. Let $f\colon \mathcal{X} \to \mathcal{Y}$ be a mapping of objects from $\mathcal{X}$ to objects from $\mathcal{Y}$. A learning restriction δ is called *f-semantic* if and only if the following holds. For all hypothesis sequences p, presentations T for some object $x \in \mathcal{X}$, presentation T' for $f(x)$ and hypothesis sequence p', if

S1: $\forall i \in \mathbb{N}\colon f(H_{\mathcal{X}}(p(i))) = H_{\mathcal{Y}}(p'(i))$,

S2: $\forall i \in \mathbb{N}\colon f(\mathrm{obj}_{\mathcal{X}}(T[i])) = \mathrm{obj}_{\mathcal{Y}}(T'[i])$, and

S3: $\delta_{\mathcal{X}}(p,T)$;

then $\delta_{\mathcal{Y}}(p',T')$.

Furthermore, if this implication holds with the following additional antecedent, then δ is called *f-pseudo-semantic*.

PS4: $\forall i \in \mathbb{N}\colon p(i) = p(i+1) \Rightarrow p'(i) = p'(i+1)$.

If, for any given learning settings $(\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ and $(\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$ and any $f\colon \mathcal{X} \to \mathcal{Y}$ we have that δ is f-(pseudo-)semantic, then we call δ *fully (pseudo-)semantic*.

If, for any given learning settings $(\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ and $(\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$ and any $f\colon \mathcal{X} \to \mathcal{Y}$ such that, for all $x,y \in \mathcal{X}$ we have $x \preceq_{\mathcal{X}} y \Leftrightarrow f(x) \preceq_{\mathcal{Y}} f(y)$, we have that δ is f-semantic, then we call δ *monotone-(pseudo-)semantic*. ◀

Intuitively, semantic learning restrictions only make requirements on the semantic characteristic of a hypothesis sequence, while pseudo-semantic restrictions can also require the hypothesis sequence to stay exactly the same. Monotone semantic restrictions also require the translations used to maintain the information order.

We go on to define further learning restrictions. **Ex** and **Bc**, as introduced by Gold [Gol67], require the convergence of the learner to a semantically correct hypothesis. **Ex** also requires the convergence to a single hypothesis. $\mathbf{Bc}^{\mathrm{c}}$ requires the semantic convergence to a hypothesis, which describes the complement of the object that should be learned. Non-U-shaped, short **NU**, learning requires the learner to never change to a wrong hypothesis after adapting a correct hypothesis and later changing changing back to a semantically equivalent hypothesis [Bal+08]. For strongly non-U-shaped, short **SNU**, learning the learner can never abandon and come back to a correct hypothesis [CM11; Wie90]. For decisive learning, short **Dec**, the learner can never come back to a semantically equivalent hypothesis [OSW86]. In strong decisive, short **SDec**, learning the learner can never come back to a semantically equivalent hypothesis and cannot output two different hypothesis that are semantically equivalent [KP16].

Given a learning setting $\mathcal{X} = (X, P, H)$ and a sequence of hypothesis p and presentations T we define

$$\begin{aligned}
\mathbf{Ex}_{\mathcal{X}}(p, T) &\Leftrightarrow \exists q \forall^{\infty} n\colon H(q) = \mathrm{obj}(T) \wedge p(n) = q \\
\mathbf{Bc}_{\mathcal{X}}(p, T) &\Leftrightarrow \forall^{\infty} n\colon H(p(n)) = \mathrm{obj}(T) \\
\mathbf{Bc}^{\mathrm{c}}_{\mathcal{X}}(p, T) &\Leftrightarrow \forall^{\infty} n\colon H(p(n)) = \overline{\mathrm{obj}(T)} \\
\mathbf{NU}_{\mathcal{X}}(p, T) &\Leftrightarrow \forall r, s, t\colon r \le s \le t \wedge H(p(r)) = H(p(t)) = \mathrm{obj}(T) \Rightarrow H(p(r)) = \\
&\qquad H(p(s)) \\
\mathbf{SNU}_{\mathcal{X}}(p, T) &\Leftrightarrow \forall r, s, t\colon r \le s \le t \wedge H(p(r)) = H(p(t)) = \mathrm{obj}(T) \Rightarrow p(r) = p(s) \\
\mathbf{Dec}_{\mathcal{X}}(p, T) &\Leftrightarrow \forall r, s, t\colon r \le s \le \wedge H(p(r)) = H(p(t)) \Rightarrow H(p(r)) = H(p(s)) \\
\mathbf{SDec}_{\mathcal{X}}(p, T) &\Leftrightarrow \forall r, s, t\colon r \le s \le \wedge H(p(r)) = H(p(t)) \Rightarrow p(r) = p(s).
\end{aligned}$$

3.3 Semantic learning restrictions

▶ **Theorem 3.12.** We have that

1. **Bc**, $\mathbf{Bc}^{\mathrm{c}}$ and **Cons** are fully semantic;
2. **Ex** is fully pseudo-semantic;
3. **SMon** and **Caut** are monotone-semantic;

4. For all injective functions f **Conv** is f-pseudo-semantic;
5. For all injective funtions f **WMon** is f-monotone-semantic.
6. For all injective functions f **Dec** and **NU** are f-semantic;
7. For all injective functions f **SDec** and**SNU** are f-pseudo-semantic.

◄

Proof. Let learning settings $\mathcal{X} = (\mathcal{X}, P_{\mathcal{X}}, H_{\mathcal{X}})$ and $\mathcal{Y} = (\mathcal{Y}, P_{\mathcal{Y}}, H_{\mathcal{Y}})$ be given. Let $f\colon \mathcal{X} \to \mathcal{Y}$ be a arbitrary translation of objects from $\mathcal{X}$ to objects from $\mathcal{Y}$. Let p, p' be hypothesis sequences, T a presentation for some object $x \in \mathcal{X}$, so $x = \mathrm{obj}_{\mathcal{X}}(T)$, T' a presentation for $f(x)$, so $f(x) = \mathrm{obj}_{\mathcal{Y}}(T')$, so that

SR1 $\forall i \in \mathbb{N}\colon f(H_{\mathcal{X}}(p(i))) = H_{\mathcal{Y}}(p'(i))$,

SR2 $\forall \in \mathbb{N}\colon f(\mathrm{obj}_{\mathcal{X}}(T[i])) = \mathrm{obj}_{\mathcal{Y}}(T'[i])$, and

SR3 $\delta_{\mathcal{X}}(p, T)$.

We now show all claimed characteristics.

1. We start with **Bc**. We want to show $\mathbf{Bc}_{\mathcal{Y}}(p', T')$. Let $n_0 \in \mathbb{N}$ so that for all $n \geq n_0$ we have $H_{\mathcal{X}}(p(n)) = \mathrm{obj}_{\mathcal{X}}(T)$. We now show that we also have $H_{\mathcal{Y}}(p'(n)) = \mathrm{obj}_{\mathcal{Y}}(T')$. As $H_{\mathcal{X}}(p(n)) = \mathrm{obj}_{\mathcal{X}}(T)$ we also have $f(H_{\mathcal{X}}(p(n))) = f(\mathrm{obj}_{\mathcal{X}}(T))$. With **SR1** and **SR2** we now have $H_{\mathcal{Y}}(p'(n)) = \mathrm{obj}_{\mathcal{Y}}(T')$. We have $\mathbf{Bc}_{\mathcal{Y}}(p', T')$, so **Bc** is fully semantic.

   Next, we show the claim for $\mathbf{Bc}^{\mathrm{c}}$. We want to show $\mathbf{Bc}^{\mathrm{c}}_{\mathcal{Y}}(p', T')$. Let $n_0 \in \mathbb{N}$ so that for all $n \geq n_0$ we have $H_{\mathcal{X}}(p(n)) = \overline{\mathrm{obj}_{\mathcal{X}}(T)}$. We now show that we also have $H_{\mathcal{Y}}(p'(n)) = \overline{\mathrm{obj}_{\mathcal{Y}}(T')}$. As $H_{\mathcal{X}}(p(n)) = \overline{\mathrm{obj}_{\mathcal{X}}(T)}$ we also have $f(H_{\mathcal{X}}(p(n))) = f(\overline{\mathrm{obj}_{\mathcal{X}}(T)})$. With **SR1** and **SR2** we now have $H_{\mathcal{Y}}(p'(n)) = \overline{\mathrm{obj}_{\mathcal{Y}}(T')}$. We have $\mathbf{Bc}^{\mathrm{c}}_{\mathcal{Y}}(p', T')$, so $\mathbf{Bc}^{\mathrm{c}}$ is fully semantic.

   We now show that **Cons** is semantic. We want to show $\mathbf{Cons}_{\mathcal{Y}}(p', T')$. We have $\mathbf{Cons}_{\mathcal{X}}(p, T)$, so for all $i \in \mathbb{N}$ we have $\mathbf{Cons}(p(i), T[i])$. This means that there exist an $I \in P(H_{\mathcal{X}}(p(i)))$ for which we have $I[i] = T[i]$, so $H_{\mathcal{X}}(p(i)) \in \mathrm{obj}_{\mathcal{X}}(T[i])$. By condition **SR1**, for any $i \in \mathbb{N}$, we have $H_{\mathcal{Y}}(p'(i)) = f(H_{\mathcal{X}}(p(i)))$ and, by condition **SR2** we get that $\mathrm{obj}_{\mathcal{Y}}(T'[i]) = f(\mathrm{obj}_{\mathcal{X}}(T([i])))$. This leads us to the conclusion that we have $H_{\mathcal{Y}}(p'(i)) \in \mathrm{obj}_{\mathcal{Y}}(T'[i])$ for any $i \in \mathbb{N}$, so there exists an $I' \in P(H_{\mathcal{Y}}(p'(i)))$ so that $I'[i] = T'[i]$. We have $\mathbf{Cons}_{\mathcal{Y}}(p', T')$, so **Cons** is semantic.

2. In addition to the conditions **SR1**, **SR2** and **SR3**, let the following also hold

   PS $\forall i \in \mathbb{N} \colon p(i) = p(i+1) \Rightarrow p'(i) = p'(i+1)$.

   We start with **Ex**. We have $\mathbf{Ex}_{\mathcal{X}}(p, T)$, let $n_0 \in \mathbb{N}$ so that for all $n \geq n_0$ we have $p(n_0) = p(n)$ and $H_{\mathcal{X}}(p(n_0)) = \mathrm{obj}_{\mathcal{X}}(T)$. We now show that we also have $H_{\mathcal{Y}}(p'(n_0)) = \mathrm{obj}_{\mathcal{Y}}(T')$. We have $H_{\mathcal{X}}(p(n_0)) = \mathrm{obj}_{\mathcal{X}}(T)$ and thus $f(H_{\mathcal{X}}(p(n_0))) = f(\mathrm{obj}_{\mathcal{X}}(T))$, by conditions **SR1** we also have $H_{\mathcal{Y}}(p'(n_0)) = obj(T')$. Since for any $n \geq n_0$ we have $p(n) = p(n+1)$ we also conclude $p'(n) = p'(n+1)$. So we have $p'(n_0) = p'(n)$. This lets us conclude $\mathbf{Ex}_{\mathcal{Y}}(p', T')$, **Ex** is fully pseudo-semantic.

3. Let $f \colon \mathcal{X} \to \mathcal{Y}$ be so that, for all $x, y \in \mathcal{X}$ we have $x \leq_{\mathcal{X}} y \Leftrightarrow f(x) \leq_{\mathcal{Y}} f(y)$.

   We show that **SMon** is monotone-semantic. We have $\mathbf{SMon}_{\mathcal{X}}(p, T)$, so for all $i, j \in \mathbb{N}$ with $i \leq j$ we have $H_{\mathcal{X}}(p(i)) \leq_{\mathcal{X}} H_{\mathcal{X}}(p(j))$. By the characteristics of f we now have $f(H_{\mathcal{X}}(p(i))) \leq_{\mathcal{Y}} f(H_{\mathcal{X}}(p(j)))$. By condition **SR1** we have $H_{\mathcal{Y}}(p'(i)) \leq_{\mathcal{Y}} H_{\mathcal{Y}}(p'(j))$ so we have $\mathbf{SMon}_{\mathcal{Y}}(p', T')$. **SMon** is monotone-semantic.

   We continue with **Caut**. We have $\mathbf{Caut}_{\mathcal{X}}(p, T)$, so for any $i, j \in \mathbb{N}$ with $i \leq j$ we have $\neg(H(p(j)) \leq H(p(i)))$. By the characteristics of f we now have $\neg(f(H_{\mathcal{X}}(p(j))) \leq f(H_{\mathcal{X}}(p(i))))$ and by condition **SR1** we have $\neg(H_{\mathcal{Y}}(p'(j)) \leq_{\mathcal{Y}} H_{\mathcal{Y}}(p'(i)))$, so $\mathbf{Caut}_{\mathcal{Y}}(p', T')$ holds. **Caut** is monotone-semantic.

4. In addition to the conditions **SR1**, **SR2** and **SR3**, let the following also hold

   PS $\forall i \in \mathbb{N} \colon p(i) = p(i+1) \Rightarrow p'(i) = p'(i+1)$

   and let f be injective.

   We continue by showing the claimed characteristics for **Conv** We have $\mathbf{Conv}_{\mathcal{X}}(p, T)$. Let $i, j \in \mathbb{N}$ be given, so that $i < j$ and $\mathbf{Cons}_{\mathcal{Y}}(p'(i), T'[j])$. This means that there exist an $I' \in P(H_{\mathcal{Y}}(p'(i)))$ for which we have $I'[i] = T[i']$, so $H_{\mathcal{Y}}(p'(i)) \in \mathrm{obj}_{\mathcal{Y}}(T'[i])$. By **SR1** and **SR2** we get $f(H_{\mathcal{X}}(p(i))) \in f(\mathrm{obj}_{\mathcal{Y}}(T[i]))$, since f is injective we get $H_{\mathcal{X}}(p(i)) \in \mathrm{obj}_{\mathcal{Y}}(T[i])$, we have $\mathbf{Cons}_{\mathcal{X}}(p(i), T[j])$. For all $x \in \mathbb{N}$ with $i \leq x < j$ we have $\mathbf{Cons}_{\mathcal{X}}(p(i), T[x])$ and, by $\mathbf{Conv}_{\mathcal{X}}(p, T)$ and induction, $p(x) = p(x+1)$. By **PS** we now get $p'(x) = p'(x+1)$, so $p'(i) = p'(j)$. We have that $\mathbf{Conv}_{\mathcal{Y}}(p', T')$, **Conv** is f-pseudo-semantic.

5. Let $f \colon \mathcal{X} \to \mathcal{Y}$ be so that, for all $x, y \in \mathcal{X}$ we have $x \leq_{\mathcal{X}} y \Leftrightarrow f(x) \leq_{\mathcal{Y}} f(y)$ and let f be injective.

We continue by showing the claim is also true for **WMon**. Given $i, j \in \mathbb{N}$ with $i \leq j$ and $\mathbf{Cons}_{\mathcal{Y}}(p'(i), T'[j])$. Analogously to the proof that **Conv** is f-pseudo-semantic, we have $\mathbf{Cons}_{\mathcal{X}}(p(i), T[j])$. Thus, we have $H_{\mathcal{X}}(p(i)) \leq H_{\mathcal{X}}(p(j))$, so we also have $f(H_{\mathcal{X}}(p(i))) \leq f(H_{\mathcal{X}}(p(j)))$. By condition **SR1** we can conclude $H_{\mathcal{Y}}(p(i)) \leq H_{\mathcal{Y}}(p(j))$, so $\mathbf{WMon}_{\mathcal{Y}}(p', T')$. **WMon** is f-monotone-semantic.

6. Let f be injective.

   Now we show the claim for **Dec**. We have $\mathbf{Dec}_{\mathcal{X}}(p, T)$, so for any $r, s, t \in \mathbb{N}$ with $r \leq s \leq t$ we get that if $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(t))$ then $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(s))$. Given r, s, t with $r \leq s \leq t$ so that $H_{\mathcal{Y}}(p'(r)) = H_{\mathcal{Y}}(p'(t))$, by condition **SR1** we get $f(H_{\mathcal{X}}(p(r))) = f(H_{\mathcal{X}}(p(t)))$ and by the definition of **DEC** we have $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(s))$. By condition **SR1** this gives us $H_{\mathcal{Y}}(p(r)) = H_{\mathcal{Y}}(p(s))$. From this we conclude $\mathbf{Dec}_{\mathcal{Y}}(p', T')$, so for all injective functions f we have that **Dec** is f-semantic.

   We continue by showing the claim for **NU**. We have $\mathbf{NU}_{\mathcal{X}}(p, T)$, so for any $r, s, t \in \mathbb{N}$ with $r \leq s \leq t$ we get that if $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(t)) = \text{obj}_{\mathcal{X}}(T)$ then we also have $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(s))$. Let $r, s, t \in \mathbb{N}$ with $r \leq s \leq t$ so that $H_{\mathcal{Y}}(p'(r)) = H_{\mathcal{Y}}(p'(t)) = \text{obj}_{\mathcal{Y}}(T')$. By conditions **SR1** we have $f(H_{\mathcal{X}}(p(r))) = f(H_{\mathcal{X}}(p(t))) = \text{obj}_{\mathcal{X}}(T)$ and since f is injective $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(t)) = \text{obj}_{\mathcal{X}}(T)$. Because of $\mathbf{NU}_{\mathcal{Y}}(p, T)$ this gives us $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(s))$, and by condition **SR1** $H_{\mathcal{Y}}(p'(r)) = H_{\mathcal{Y}}(p'(s))$. So we have $\mathbf{NU}_{\mathcal{Y}}(p', T')$ and, thus, for all injective functions f **NU** is f-semantic.

7. In addition to the conditions **SR1**, **SR2** and **SR3**, let the following also hold

   PS $\forall i \in \mathbb{N}\colon p(i) = p(i+1) \Rightarrow p'(i) = p'(i+1)$

   and let f be injective. We now show that **SNU** is f-pseudo-semantic. We have $\mathbf{SNU}_{\mathcal{X}}(p, T)$, so for any $r, s, t \in \mathbb{N}$ with $r \leq s \leq t$ we get that if $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(t)) = \text{obj}_{\mathcal{X}}(T)$ then we also have $p(r) = p(s)$. Let $r, s, t \in \mathbb{N}$ with $r \leq s \leq t$ so that $H_{\mathcal{Y}}(p'(r)) = H_{\mathcal{Y}}(p'(t)) = \text{obj}_{\mathcal{Y}}(T')$. By conditions **SR1** we have $f(H_{\mathcal{X}}(p(r))) = f(H_{\mathcal{X}}(p(t))) = \text{obj}_{\mathcal{X}}(T)$. We have $p(r) = p(r)$ and since f is injective for any $n \in \mathbb{N}$ with $r \leq n \leq t$ we also have $p(n) = p(r)$. From this we can conclude that $p'(n) = p'(r)$, so $p'(s) = p'(r)$. We have $\mathbf{SNU}_{\mathcal{Y}}(p', T')$, so **SNU** is f-pseudo-semantic.

   Finally we show that the claim also holds for **SDec**. We have $\mathbf{SDec}_{\mathcal{X}}(p, T)$, so for any $r, s, t \in \mathbb{N}$ with $r \leq s \leq t$ we get that if $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(t))$ then $p(r) = p(s)$. Given r, s, t with $r \leq s \leq t$ so that $H_{\mathcal{Y}}(p'(r)) = H_{\mathcal{Y}}(p'(t))$, by condition **SR1** we get $f(H_{\mathcal{X}}(p(r))) = f(H_{\mathcal{X}}(p(t)))$. Since f is injective we now have $H_{\mathcal{X}}(p(r)) = H_{\mathcal{X}}(p(t))$ and thus for any $n \in \mathbb{N}$ with $r \leq n \leq t$ we also have $p(n) = p(r)$. From this we can conclude that $p'(n) = p'(r)$, so $p'(s) = p'(r)$. We have $\mathbf{SDec}_{\mathcal{Y}}(p', T')$, so **SDec** is f-pseudo semantic.

■

4 Mappings between Learning Tasks

In this section we are first going to identify embeddings and isomorphisms between languages and sets of functions under different learning settings and restrictions. Later we show that for certain learning settings and restrictions there does not exist and isomorphism between decidable languages learned under **W**- and **C**-indices.

4.1 Isomorphic learning tasks

We begin by showing isomorphisms between sets of functions and sets of functions and decidable languages. Because of the transitivity of isomorphisms we can show that total computable functions and infinite decidable languages are isomorph under semantic and delayable learning criteria and **C**-indices.

4.1.1 Isomorphism between sets of total functions

Let $\mathcal{R}_{0,1}^{\infty}$ denote the set of total computable functions $f\colon \mathbb{N} \to \{0,1\}$, where, for infinitely many $n \in \mathbb{N}$, we have $f(n) = 1$.

Intuitively, in this section we first show that for delayble function learning the order of the presentation does not matter, i.e. the same classes of functions can be learned from arbitrary informants as from canonical informants. We define a function Θ translating from $\mathcal{R}$ to $\mathcal{R}_{0,1}^{\infty}$. We show that this function is a bijection. We continue by also defining a function transforming a presentation for a $f \in \mathcal{R}$ to a presentation for $\Theta(f) \in \mathcal{R}_{0,1}^{\infty}$. We show that, for delayable learning restrictions, if a learner can learn a function f it can also learn it from the translated presentation for $\Theta(f)$. From this we go on to show the main theorem of this section, that is, that we have $(\mathbf{Arb}_{\mathcal{R}_{0,1}^{\infty}}, \mathbf{G}, \delta) \cong (\mathbf{Arb}_{\mathcal{R}}, \mathbf{G}, \delta)$.

▶ **Theorem 4.1.** For all delayable learning restrictions δ we have $[\mathbf{ArbG}\delta] = [\mathbf{CanG}\delta]$. ◀

Proof. Clearly $[\mathbf{ArbG}\delta] \subseteq [\mathbf{CanG}\delta]$. To prove the other inclusion, let h be a $\mathbf{CanG}\delta$-learner. For all sequences σ we define $c(\sigma)$ to return the longest initial sequence of a canonical representation of coded tuples contained in σ. We define a learner g so that, for all finite sequences σ,

$$g(\sigma) = h(c(\sigma)).$$

Let $f \in \mathbf{CanG}\delta(h)$. Let T_c denote the canonical representation of f and p_g the learning sequence of g on T_c. Since for all $n \in \mathbb{N}$ we have $T_c[n] = c(T_c[n])$ and h learns f from T_c, we conclude that g also learns f from T_c, so $(p_g, T_c) \in \delta$. For any $T \in \mathbf{Arb}(f)$ we now define $r_T : \mathbb{N} \to \mathbb{N}$ as

$$r_T(n) = \max\{n' \in \mathbb{N} \mid \mathrm{content}(T_c[n']) \subseteq \mathrm{content}(T[n])\}.$$

Intuitively r_T always returns the length of the longest initial sequence of a canonical representation contained in content$(T[n])$. We show that $r_T \in \vec{R}$. As r_T is the maximum of a set which grows monotonically in regards to the $\subseteq$-relation, r_T is non-decreasing. To show that r_T has an infinite limit inferior, we show that for all $i \in \mathbb{N}$ there exists an $m \in \mathbb{N}$ so that $r_T(m) \geq i$ by induction. Since r_T is non-decreasing this suffices to show that for all $i \in \mathbb{N}$ there exists an $m \in \mathbb{N}$ so that for all $n \geq m$ we have $r_T(n) \geq i$. For $i = 0$ this claim holds, as any sequence contains the empty sequence. Let the claim be true for an arbitrary i. We show that it holds for $i + 1$ as well. Let m' so that $r_T(m') \geq i$. If $r_T(m') > i$ we have $r_T(m') \geq i+1$. If $r_T(m') = i$ there exists an $m'' \geq m'$ with $T(m'') = \langle i+1, f(i+1) \rangle$, so $r_T(m'') \geq i+1$. So the claim holds in both cases, r_T has an infinite limit inferior.

Following the definition of r_T we have that, for all $n \in \mathbb{N}$, content$(T_c[r_T(n)]) \subseteq$ content$(T[n])$. Since δ is a delayable learning restriction this leads us to the conclusion that $(p_g \circ r_T, T) \in \delta$, so h $\mathbf{ArbG}\delta$-learns f. ■

We define a function $u : \mathbb{N} \to \mathbb{N}$ which translates a Gödel number of a function $f \in \mathcal{R}$ into the Gödel number of a function $f' \in \mathcal{R}^{\infty}_{0,1}$. With the *s-m-n* Theorem we can define $u : \mathbb{N} \to \mathbb{N}$ so that, for all $e, x \in \mathbb{N}$, we have

$$\varphi_{u(e)}(x) = \begin{cases} \bot, & \text{if } \exists t \leq x : \varphi_e(t) = \bot, \\ 1, & \text{if } \exists t \leq x : \sum_{i=0}^{t-1}(\varphi_e(i) + 1) + \varphi_e(t) = x, \\ 0, & \text{otherwise.} \end{cases}$$

Intuitively, u maps every Gödel number e of a total computable function to a total function which first returns $\varphi_e(0)$ many 0s, then a 1 as a separator, $\varphi_e(1)$ many 0s, a 1 as a separator, and so on.

We define $u' : \mathbb{N} \to \mathbb{N}$ so that for any e' with $\varphi_{e'} \in \mathcal{R}^{\infty}_{0,1}$ and $n \in \mathbb{N}$ we have that $\varphi_{u'(e')}(n)$ is undefined if $\varphi_{e'}$ is undefined for any value before the $n+1$th occurrence of 1. Otherwise it maps to the number of 0s between the nth and $n+1$th occurrence of 1 in $\varphi_{e'}$. So u' is the computable inverse function of u with regards to e'.

For this section we define another function $\Theta\colon \mathcal{R} \to \mathcal{R}_{0,1}^{\infty}$ on the basis of u, but working on functions instead of Gödel numbers. So for any $\varphi_e \in \mathcal{R}$ we have

$$\Theta(\varphi_e) = \varphi_{u(e)}.$$

Since $u(e)$ only depends on the semantic meaning of the program encoded in e, not the syntactic meaning of the number e itself, Θ is a well-defined function. For any $f \in \mathcal{R}$, there are never infinitely many 0s in a row in $\Theta(f)$, meaning that there are infinitely many $n \in \mathbb{N}$ so that $\Theta(f)(n) = 1$. This results in $\Theta(f) \in \mathcal{R}_{0,1}^{\infty}$.

To show that Θ is injective we take two functions $f, g \in \mathcal{R}$ with $f \neq g$. We take $x \in \mathbb{N}$ to be the smallest number so that $f(x) \neq g(x)$ and assume without loss of generality that $f(x) \leq g(x)$. Let $n_0 \coloneqq \sum_{i=0}^{x-1}(f(i)+1) + f(x)$. Since, up until x, f and g map to the same result and $f(x) < g(x)$ for any $n' \leq n_0$ we have $\Theta(f)(n') = \Theta(g)(n')$. However $\Theta(f)(n+1) = 1$, but $\Theta(g)(n+1) = 0$ so $\Theta(f) \neq \Theta(g)$. Thus, Θ is injective. To show that Θ is also surjective we take any $f' \in \mathcal{R}_{0,1}^{\infty}$. Since f' maps to 1 at infinitely many points, we can create a new function $f \in \mathcal{R}$ for which $f(0)$ maps the number of 0s before the first 1 appears as a function value for $f'(0)$, $f(1)$ is the number of 0s between the first and second occurrence of 1 as a function value, and so on. By the definition of Θ it becomes clear that $\Theta(f) = f'$.

So Θ is injective and surjective and thus bijective. This implies the existence of the inverse function of Θ, denoted by Θ^{-1}.

Intuitively, Θ^{-1} translates each function $\varphi_{e'} \in \mathcal{R}_{0,1}^{\infty}$ into a function φ_e, so that $\varphi_e(0)$ is the number of 0s before the first occurrence of 1 in $\varphi_{e'}$ and for any $n \in \mathbb{N}^+$ $\varphi_e(n)$ is the number of 0s between the nth and $n+1$th occurrence of 1. So Θ^{-1} uses the same principle of translation for functions as u' does on Gödel numbers. For any $\varphi_{e'} \in \mathcal{R}_{0,1}^{\infty}$ we have

$$\Theta^{-1}(\varphi_{e'}) = \varphi_{u'(e')}.$$

For this section we define a translator function $\vartheta\colon \mathbf{Can}(\mathcal{R}) \to \mathbf{Can}(\mathcal{R}_{0,1}^{\infty})$, which translates canonical presentations for functions $f \in \mathcal{R}$ into canonical presentations for $\Theta(f) \in \mathcal{R}_{0,1}^{\infty}$. Given $C \in \mathbf{Can}(\mathcal{R})$ and $n \in \mathbb{N}$, for all $n' \leq n$ we have

$$\vartheta(C[n])[0] = \epsilon$$

$$x_{n'} = \begin{cases} 1, & \text{if } \exists t \leq n' : \sum_{i=0}^{t-1}(\pi_1(C[i]) + 1) + \pi_1(C[t]) = n', \\ 0, & \text{else.} \end{cases}$$

$$\vartheta(C[n])(n'+1) = \langle n', x'_n \rangle.$$

Since every value in $C[n]$ will be encoded with at least on value in $\vartheta(C[n])$ it is ensured that ϑ always builds a sequence of length n and $\vartheta \in \mathcal{R}$. Since ϑ uses the same principle of translation as u and Θ, it maps presentations for a $f \in \mathcal{R}$ to presentations for $\Theta(f) \in \mathcal{R}^\infty_{0,1}$.

The counterpart function $\vartheta' \colon \mathbf{Can}(\mathcal{R}^\infty_{0,1}) \to \mathbf{Can}(\mathcal{R})$ translates the canonical presentations for functions $f' \in \mathcal{R}^\infty_{0,1}$ to the canonical presentation for $\Theta^{-1}(f') \in \mathcal{R}$. So, for any given $C' \in \mathbf{Can}(\mathcal{R}^\infty_{0,1})$ and $n \in \mathbb{N}$, let $i := |\{x \in \mathbb{N} \mid \langle x, 1 \rangle \in \mathrm{content}(C'[n])\}|$. For $0 \leq i' < i$, let x'_i denote the amount of 0s between the i'th and $i'+1$th occurrence of 1, or before the first 1 for x_0. For all $i' \in \mathbb{N}$ with $0 \leq i' < i$ we have

$$\vartheta'(C'[n])[0] = \epsilon,$$
$$\vartheta'(C'[n])[i'+1] = \vartheta'(C[n])[i']^\frown\langle i', x_{i'} \rangle.$$

Analogously to ϑ, ϑ' computably maps presentations of $f' \in \mathcal{R}^\infty_{0,1}$ to presentations for $\Theta^{-1}(f') \in \mathcal{R}$ as it uses the same principle of translation as u' and Θ^{-1}, just on sequences instead of functions or Gödel numbers.

In the following lemma we show that for delayable learning criteria a learner which can learn a function $f \in \mathcal{R}$ from its canonical presentation can also learn f from a canonical presentation for $\Theta(f)$ that is translated using ϑ'.

▶ **Lemma 4.2.** Let δ be a delayable learning restriction. Given $f \in \mathcal{R}$, $C \in \mathbf{Can}(f)$, $C' \in \mathbf{Can}(\Theta(f))$ and a learner h. Let p_C and $p_{\vartheta'}$ be the hypothesis sequences of h on C and $\vartheta'(C')$, respectively. Then we have

$$(p_C, C) \in \delta \Rightarrow (p_{\vartheta'}, \vartheta'(C')) \in \delta.$$

◀

Proof. First, we define a delaying function r_C and show $r_C \in \vec{R}$. Then, we use the delayability of δ to show $(p_C \circ r_C, C \circ r_C) \in \delta$ and finally conclude $(p_{\vartheta'}, \vartheta'(C')) \in \delta$. We define $r_C \colon \mathbb{N} \to \mathbb{N}$ so that for all $n \in \mathbb{N}$ we have

$$r_C(n) = |\{x \in \vartheta(C[n]) \mid \pi_2(x) = 1\}|.$$

Intuitively, r_C returns the length of the longest sequence of a canonical representation for f that can be reconstructed from $\vartheta(C)$. We show that $r_C \in \vec{R}$. As r_C returns the size of a set growing monotonically in regards to the $\subseteq$-relation, it is non-decreasing. Since $\vartheta(C)$ is a presentation for a function $\Theta(f) \in \mathcal{R}^\infty_{0,1}$, it contains infinitely many pairs with $\langle x, 1 \rangle \in \mathbb{N} \times \{1\}$ which encode arguments x for which $\Theta(f)(x) = 1$ Since r_C counts the 1s appearing in $\vartheta(C)$ and every function value of $\Theta(f)$ appears in C, r_C also has an infinite limit inferior. We have $r_C \in \vec{R}$. Because $\vartheta(C) \in \mathbf{Can}(\mathcal{R}^{\infty}_{0,1})$, we have that there exists no $n \in \mathbb{N}$ so that $r_C(x) + 1 < r_C(x + 1)$, so we have $\text{content}(C) = \text{content}(C \circ r_C)$. Following the definition of r_C for all $n \in \mathbb{N}$ we have $\text{content}(C[r_C(n)]) = \text{content}(C \circ r_C[n])$. Since δ is delayable we have $(p_C \circ r_C, C \circ r_C) \in \delta$.

Both p_C and $p_{\vartheta'}$ are hypothesis sequences of the same learner h on C and $\vartheta'(C')$. Since $C' \in \mathbf{Can}(\Theta(f))$ and since, by definition, ϑ' translates canonical presentations for $\Theta(f)$ to canonical presentations for f, we have that $C = \vartheta'(C')$. The delay caused by r_C is also the length of the subsequence of C that can be reconstructed from C' with the use of $\vartheta(C')$. This leads us to the conclusion, that for all $n \in \mathbb{N}$ we have $C \circ r_C[n] = \vartheta'(C'[n])$ and $p_C \circ r_C = p_{\vartheta'}$. So we have $(p_{\vartheta'}, \vartheta'(C')) \in \delta$. ■

▶ **Theorem 4.3.** Let δ be Θ-Θ^{-1}-semantic and delayable. Then we have

$$(\mathbf{Arb}_{\mathcal{R}^{\infty}_{0,1}}, \mathbf{G}, \delta) \cong (\mathbf{Arb}_{\mathcal{R}}, \mathbf{G}, \delta).$$

◀

Proof. We use Θ as the bijection between $\mathcal{R}$ and $\mathcal{R}^{\infty}_{0,1}$. Since δ is delayable, by Theorem 4.1 it suffices to show that for all $\mathcal{F} \subseteq \mathcal{R}$ we have $\mathcal{F} \in [\mathbf{CanG}\delta_{\mathcal{R}}]$ if and only if $\Theta(\mathcal{F}) \in [\mathbf{CanG}\delta_{\mathcal{R}^{\infty}_{0,1}}]$.

To show that if $\mathcal{F} \in [\mathbf{CanG}\delta_{\mathcal{R}}]$ then $\Theta(\mathcal{F}) \in [\mathbf{CanG}\delta_{\mathcal{R}^{\infty}_{0,1}}]$, let h be a learner and $\mathcal{F} = \mathbf{CanG}\delta_{\mathcal{R}}(h)$. We define the learner g so that for any $C \in \mathbf{Can}(\Theta(\mathcal{F}))$ and all $n \in \mathbb{N}$ we have

$$g(C[n]) = u(h(\vartheta'(C[n]))).$$

Let p_h be the learning sequence of h on $\vartheta'(C)$. Because of Lemma 4.2 we have $(p_h, \vartheta'(C)) \in \delta$. Let p_g be the hypothesis sequence of g on C. For any $i \in \mathbb{N}$ we now have $p_g(i) = u(h(\vartheta'(C[n]))) = u(p_h(i))$. Since u and Θ apply the same transformation to Gödel numbers and function respectively, this means that we have $\varphi_{p_g(i)} = \varphi_{u(p_h(i))} = \Theta(\varphi_{p_h(i)})$. We also have $\Theta(\text{obj}(p_h(i))) = \text{obj}(p_g(i))$. Since δ is Θ-semantic we conclude $(p_g, C) \in \delta$ and, thus, g $\mathbf{CanG}\delta_{\mathcal{R}}$-learns $\Theta(\mathcal{F})$.

To show that if $\Theta(\mathcal{F}) \in [\mathbf{CanG}\delta_{\mathcal{R}^{\infty}_{0,1}}]$ we have $\mathcal{F} \in [\mathbf{CanG}\delta_{\mathcal{R}}]$ let g be a learner and $\mathcal{F} = \mathbf{CanG}\delta_{\mathcal{R}^{\infty}_{0,1}}(g)$. We now define a learner h so that for any $C \in \mathbf{Can}(\Theta^{-1}(\mathcal{F}))$ and all $n \in \mathbb{N}$ we have

$$h(C[n]) = u'(g(\vartheta(C[n]))).$$

Let p_g be the learning sequence of g on $\vartheta(C)$. By the definition of ϑ and since g learns $\mathcal{F}$ we have $(p_g, \vartheta(C[n])) \in \delta$. Let p_h be the learning sequence of h on

C. For any $i \in \mathbb{N}$, we now have $p_h(i) = u'(h(\vartheta(C[n]))) = u'(p_g(i))$. This gives us $\varphi_{p_h(i)} = \varphi_{u'(p_g(i))} = \Theta^{-1}(\varphi_{p(i)})$. We also have $\Theta^{-1}(\mathrm{obj}(p_g(i))) = \mathrm{obj}(p_h(i))$. We conclude that $(p_h, C) \in \delta$ and, thus, h **CanG**δ-learns $\Theta^{-1}(\mathcal{F})$. ■

4.1.2 Isomorphism between total functions and decidable languages

We first define a function translating between $\mathcal{R}_{0,1}$ and $\mathcal{E}_c$ an then go on to show an isomorphism between these sets for all semantic learning restrictions. Finally we use this isomorphism to also build an isomorphism between the set of infinite decidable languages, denoted by $\mathcal{E}_c^\infty$, and $\mathcal{R}_{0,1}^\infty$, allowing us to generalize the theorems from this section into an isomorphism between $\mathcal{R}$ and $\mathcal{E}_c^\infty$.

In this section we define the function $\Theta\colon \mathcal{R}_{0,1} \to \mathcal{E}_c$, so that for all $f \in \mathcal{R}_{0,1}$ we have

$$\Theta(f) = \{x \in \mathbb{N} \mid f(x) = 1\}. \tag{4.1}$$

Intuitively Θ interprets f as the characteristic function for a language and maps it to this language. Since for any $f, g \in \mathcal{R}_{0,1}$ with $f \neq g$ there exists an $x \in \mathbb{N}$ so that $f(x) \neq g(x)$, we have, without loss of generality, $x \in \Theta(f)$ and $x \notin \Theta(g)$ and thus $\Theta(f) \neq \Theta(g)$, so Θ is injective. For any $L \in \mathcal{E}_c$ there exists a characteristic function $\chi_L \in \mathcal{R}_{0,1}$ for L and by definition $\Theta(\chi_L) = L$. Thus, Θ is surjective and bijective. So there exists an inverse function for Θ translating decidable languages to their characteristic function, which we denote by Θ^{-1}.

▶ **Theorem 4.4.** Let δ be Θ-Θ^{-1}-semantic. Then we have

$$(\mathbf{Arb}_{\mathcal{R}_{0,1}}, \mathbf{G}, \delta) \cong (\mathbf{Inf}_{\mathcal{E}_c}, \mathbf{G}, \delta)_{\mathrm{C}}.$$

◀

Proof. We observe that by the definition of informants and presentations for functions for any $f \in \mathcal{R}_{0,1}$ we have $\mathbf{Arb}_{\mathcal{R}_{0,1}}(f) = \mathbf{Inf}_{\mathcal{E}_c}(\Theta(f))$. So an arbitrary presentations for a $f \in \mathcal{R}_{0,1}$ is always also an informant for $\Theta(f) \in \mathcal{E}_c$ and vice versa.

We first show that for any $\mathcal{F} \subseteq \mathcal{R}_{0,1}$, if $\mathcal{F} \in [\mathbf{Arb}_{\mathcal{R}_{0,1}}\mathbf{G}\delta]$ then we have that $\Theta(\mathcal{F}) \in [\mathbf{Inf}_{\mathcal{E}_c}\mathbf{G}\delta]_{\mathrm{C}}$. Let h be a learner and $\mathcal{F} = \mathbf{Arb}_{\mathcal{R}_{0,1}}\mathbf{G}\delta(h)$. We now define the learner g so that for any $I \in \mathbf{Inf}_{\mathcal{E}_c}(\Theta(\mathcal{F}))$ and $n \in \mathbb{N}$ we have

$$g(I[n]) = h(I[n]).$$

So the learner g forwards the φ-hypothesis given by h, which is now interpreted as a $\mathbf{C}$-hypothesis.

Let p_h and p_g be the learning sequences of h and g on I respectively. Since $I \in \mathbf{Arb}_{\mathcal{R}_{0,1}}(\mathcal{F})$ and h $\mathbf{Arb}_{\mathcal{R}_{0,1}}\mathbf{G}\delta$-learns $\mathcal{F}$, we have $(p_h, I) \in \delta$. For all $i \in \mathbb{N}$, by the definition of Θ, we now have $\Theta(\varphi_{p_h(i)}) = \mathbf{C}_{p_g(i)}$ and $\Theta(\mathrm{obj}_{\mathcal{R}_{0,1}}(I[i])) = \mathrm{obj}_{\mathcal{E}_c}(I[i])$. Since δ is Θ-semantic we conclude $(p_g, I) \in \delta$. So g $\mathbf{Inf}_{\mathcal{E}_c}\mathbf{G}\delta_{\mathbf{C}}$-learns $\Theta(\mathcal{F})$.

We now show the other implication, that is if $\Theta(\mathcal{L}) \in [\mathbf{Inf}_{\mathcal{E}_c}\mathbf{G}\delta_{\mathbf{C}}]$ then $\mathcal{F} \in [\mathbf{Arb}_{\mathcal{R}_{0,1}}\mathbf{G}\delta]$. Given a learner g and $\mathcal{L} = \mathbf{Inf}_{\mathcal{E}_c}\mathbf{G}\delta(g)$, for any $A \in \mathbf{Arb}_{\mathcal{R}_{0,1}}(\Theta^{-1}(\mathcal{L}))$ and $n \in \mathbb{N}$ we define an new learner h so that

$$h(A[n]) = g(A[n]).$$

So h takes any $\mathbf{C}$-hypothesis made by g and returns it as an φ-hypothesis.

Let p_g and p_h be the learning sequences of g and h on A respectively. Since $A \in Inf_{\mathcal{E}_c}(\mathcal{L})$ we have $(p_g, A) \in \delta$. For all $i \in \mathbb{N}$ we have $\Theta^{-1}(C_{p_g(i)}) = \varphi_{p_h(i)}$ and $\Theta^{-1}(\mathrm{obj}_{\mathcal{E}_c}(A[i])) = \mathrm{obj}_{\mathcal{R}_{0,1}}(A[i])$. Since δ is Θ^{-1}-semantic, we conclude $(p_h, A) \in \delta$. So h $\mathbf{Arb}_{\mathcal{R}_{0,1}}\mathbf{G}\delta$-learns $\Theta^{-1}(\mathcal{L})$. ■

Recall Θ from the proof of Equation (4.1). We consider $\Theta|_{\mathcal{R}_{0,1}^{\infty}} : \mathcal{R}_{0,1}^{\infty} \to \mathcal{E}_c^{\infty}$. This restricted function remains injective. Every $L \in \mathcal{E}_c^{\infty}$ has a characteristic function χ_L which maps to 1 infinitely often, as there are infinitely many elements in L. So we have $\chi_L \in \mathcal{R}_{0,1}^{\infty}$ and by the definition of Θ we get $\Theta|_{\mathcal{R}_{0,1}^{\infty}}(\chi_L) = L$. So $\Theta|_{\mathcal{R}_{0,1}^{\infty}}$ is also surjective and thus a bijection. This implies the existence of an inverse function which we denote by $(\Theta|_{\mathcal{R}_{0,1}^{\infty}})^{-1}$.

▶ **Corollary 4.5.** Let δ be $\Theta|_{\mathcal{R}_{0,1}^{\infty}}$-$(\Theta|_{\mathcal{R}_{0,1}^{\infty}})^{-1}$-semantic. Then we have

$$(\mathbf{Arb}_{\mathcal{R}_{0,1}^{\infty}}, \mathbf{G}, \delta) \cong (\mathbf{Inf}_{\mathcal{E}_c^{\infty}}, \mathbf{G}, \delta)_{\mathbf{C}}.$$

◀

Proof. From Theorem 4.4 we directly conclude that, for all $\mathcal{L} \subseteq \mathcal{R}_{0,1}^{\infty}$, we have that if and only if $\mathcal{L} \in [\mathbf{Arb}_{\mathcal{R}_{0,1}}\mathbf{G}\delta]$ then $\Theta(\mathcal{L}) \in [\mathbf{InfG}\delta]_{\mathbf{C}}$. So $(\mathbf{Arb}_{\mathcal{R}_{0,1}^{\infty}}, \mathbf{G}, \delta) \cong (\mathbf{Inf}_{\mathcal{E}_c}^{\infty}, \mathbf{G}, \delta)_{\mathbf{C}}$. ■

▶ **Corollary 4.6.** Let δ be semantic and delayable. We have

$$(\mathbf{Arb}_{\mathcal{R}}, \mathbf{G}, \delta) \cong (\mathbf{Inf}_{\mathcal{E}_c}^{\infty}, \mathbf{G}, \delta)_{\mathbf{C}}$$

. ◀

Proof. This follows directly from Theorem 4.3, Corollary 4.5 and the transitivity of $\cong$ shown in Theorem 3.6. ■

▶ **Corollary 4.7.** We have

$$(\mathbf{Arb}_{\mathcal{R}_{0,1}}, \mathrm{G}, \delta) \cong_c (\mathbf{Inf}_{\mathcal{E}_c}, \mathrm{G}, \delta)_{\mathrm{C}}, \text{ and}$$
$$(\mathbf{Arb}_{\mathcal{R}_{0,1}^\infty}, \mathrm{G}, \delta) \cong_c (\mathbf{Inf}_{\mathcal{E}_c}^\infty, \mathrm{G}, \delta)_{\mathrm{C}}.$$

◀

Proof. This follows directly from the proofs for Theorem 4.4 and Corollary 4.5. ■

4.2 Embeddings

In this section we show that $\mathcal{E}_c$ embeds $\mathcal{E}_c^\infty$ and vice versa.

▶ **Theorem 4.8.** Let δ be a learning restriction and P be either a text or informant presentation system. Then, we have

$$(P_{\mathcal{E}_c^\infty}, \mathrm{G}, \delta)_{\mathrm{C}} \hookrightarrow (P_{\mathcal{E}_c}, \mathrm{G}, \delta)_{\mathrm{C}}.$$

◀

Proof. First, please note that $\mathcal{E}_c^\infty \subseteq \mathcal{E}_c$. We now consider the function $id\colon \mathcal{E}_c^\infty \to \mathcal{E}_c$, the injective identity function. Given a learner h and $\mathcal{L} = P\mathrm{G}\delta_{\mathcal{E}_c^\infty}(h)$, h learns every $L = id(L) \in \mathcal{L}$. Thus, $id(L) \in P\mathrm{G}\delta_{\mathcal{E}_c}(h)$ and $P\mathrm{G}\delta_{\mathcal{E}_c^\infty}(h) \subseteq P\mathrm{G}\delta_{\mathcal{E}_c}(h)$.

Given a learner g and $\mathcal{L}' = P\mathrm{G}\delta_{\mathcal{E}_c}(g) \cap \mathcal{E}_c^\infty$, g learns every $L' \in \mathcal{L}'$. Thus, $id(L') \in P\mathrm{G}\delta_{\mathcal{E}_c^\infty}(h)$ and $P\mathrm{G}\delta_{\mathcal{E}_c^\infty}(h) = P\mathrm{G}\delta_{\mathcal{E}_c}(h) \cap \mathcal{E}_c^\infty$. ■

▶ **Corollary 4.9.** For delayable and semantic learning restrictions δ we have

$$(\mathbf{Arb}_{\mathcal{R}}, \mathrm{G}, \delta) \hookrightarrow (\mathbf{Arb}_{\mathcal{R}_{0,1}^\infty}, \mathrm{G}, \delta) \hookrightarrow (\mathbf{Inf}_{\mathcal{E}_c}^\infty, \mathrm{G}, \delta)_{\mathrm{C}}$$
$$\hookrightarrow (\mathbf{Inf}_{\mathcal{E}_c}, \mathrm{G}, \delta)_{\mathrm{C}} \hookrightarrow (\mathbf{Arb}_{\mathcal{R}_{0,1}}, \mathrm{G}, \delta)$$

◀

Proof. This follows directly from Theorem 4.8, the isomorphism shown in Theorem 4.3, Corollary 4.5 and Theorem 4.4 and Corollary 3.7. ■

To show the that $\mathcal{E}_c^\infty$ embeds $\mathcal{E}_c$ we first define a function translating between these two sets of functions and a reverse function. Just as in the proof of Theorem 4.3 we also define methods to translate between the canonical presentations of functions from $\mathcal{E}_c$ and $\mathcal{E}_c^\infty$ and show that delayable learners which can learn an $L \in \mathcal{E}_c$ or $\mathcal{E}_c^\infty$ from its canonical informant can also learn it from the translated presentation. With this we then show that $\mathcal{E}_c \hookrightarrow \mathcal{E}_c^\infty$.

In this section we define $\Theta\colon \mathcal{E}_c \to \mathcal{E}_c^\infty$ so that for all $L \in \mathcal{E}_c$ we have

$$\Theta(L) = \{2x + 1 \mid x \in L\} \cup 2\mathbb{N}.$$

Since L is decidable, so is $\Theta(L)$. As $2\mathbb{N}$ is infinite, we have $\Theta(L) \in \mathcal{E}_c^\infty$. Given $L, M \in \mathcal{E}_c$ with $L \neq M$ without loss of generality there exists an $x \in L$ with $x \notin M$. This means that $2x + 1 \in \Theta(L)$ but $2x + 1 \notin \Theta(M)$, so we have $\Theta(L) \neq \Theta(M)$. Thus, Θ is also injective.

We define another function $\Omega\colon \Theta(\mathcal{E}_c) \to \mathcal{E}_c$ so that for all $L' \in \mathrm{range}(\Theta)$ we have

$$\Omega(L') = \left\{ \frac{x-1}{2} \,\middle|\, x \in L \setminus 2\mathbb{N} \right\}.$$

Intuitively Ω is the inverse function of Θ . Given $L', M' \in \Theta(\mathcal{E}_c)$ with $L' \neq M'$ without loss of generality there exists an $x \in L'$ with $x \notin M'$. Since, by the definition of Θ, we have that $2\mathbb{N} \subseteq L'$ and $2\mathbb{N} \subseteq M'$, so x is uneven. So we have $\frac{x-1}{2} \in \Omega(L')$ and $\frac{x-1}{2} \notin \Omega(M')$. We conclude that Ω is injective.

We define a function $\vartheta\colon \mathbf{Inf}(\mathcal{E}_c) \to \mathbf{Inf}(\mathcal{E}_c^\infty)$ so that for this section for any $I \in \mathbf{Inf}(\mathcal{E}_c)$ and $n \in \mathbb{N}$ and all $n' \leq 2n$ we have

$$\begin{aligned} \vartheta(I[n])[0] &= \epsilon, \\ \vartheta(I[n])[2(n'+1)] &= \vartheta(I[n])[2(n')]^\frown\langle\, 2\pi_1(I[n]) - 1, \pi_2(I[n]) \,\rangle^\frown\langle\, 2\pi_1(I[n]), 1 \,\rangle. \end{aligned}$$

Intuitively, for every $L' \in \mathcal{E}_c^\infty$ and $I \in \mathbf{Inf}(\Omega(L'))$, ϑ constructs a new informant from I. It uses every information $\langle\, x, y \,\rangle \in \mathrm{content}(I)$ to include tuples $\langle\, 2x, 1 \,\rangle$ and $\langle\, 2x + 1, y \,\rangle$ in $\vartheta(I)$. It follows the same translation method as Θ, so we have $\vartheta(I) \in \mathbf{Inf}(L')$. Further for any $L \in \mathcal{E}_c$ and $I \in \mathbf{Inf}_{\mathbf{Can}}((L))$ we have $\vartheta(I) \in \mathbf{Inf}_{\mathbf{Can}}(\Theta(L))$.

We define another function $\omega\colon \mathbf{Inf}(\mathcal{E}_c^\infty) \to \mathbf{Inf}(\mathcal{E}_c)$ so that for any $I' \in \mathbf{Inf}(\mathcal{E}_c^\infty)$, $n \in \mathbb{N}$ and $n' \leq |\{\langle\, x, y \,\rangle \in I'[n] \mid x \notin 2\mathbb{N}\}|$ we have

$$\begin{aligned} \rho(\langle\, x, y \,\rangle) &= \langle\, \frac{x-1}{2}, y \,\rangle, \\ next(I'[n], n') &= \{i \leq n \mid \pi_1(I'[i]) \notin 2\mathbb{N} \wedge \rho(I'[i]) \notin \mathrm{content}(\omega(I'[n])[n'])\}, \\ \omega(I'[n])[0] &= \epsilon, \end{aligned}$$

$$\omega(I'[n])[n'+1] = \omega(I'[n])[n']^\frown \begin{cases} \rho(\min(next(I'[n], n'))), & \text{if } next(I'[n], n') \neq \emptyset, \\ \epsilon, & \text{else.} \end{cases}$$

Intuitively, ρ transforms tuples containing information about uneven number, mimicking the translations performed by Ω on tuples instead of sets, while $next(I'[n], n')$ gives us a set of the values not yet contained in $\omega(I'[n-1])$. This, we use so that $\omega(I'[n])$ gives us a sequence containing information deduced from all uneven numbers contained in $I'[n]$. By following the same principles of translation as Ω, it maps the informant I' for any $L' \in \mathcal{E}_c^\infty$ to an informant for $\Theta(L')$. Further for any $I' \in \mathbf{Inf}_{\mathrm{Can}}(L')$ we have $\omega(I') \in \mathbf{Inf}_{\mathrm{Can}}(\Omega(L'))$. Especially for any $L' \in \mathrm{range}(\Theta)$ and $I' \in \mathbf{Inf}_{\mathrm{Can}}(L')$ we have $\vartheta(\omega(I')) = I'$ and for $L \in \mathcal{E}_c$ and $I \in \mathbf{Inf}_{\mathrm{Can}}(L)$ we have $\omega(\vartheta(I)) = I$.

In the following two lemmata we show that, for delayable learning criteria, if a learner can learn a language $L \in \mathcal{E}_c$ or $L' \in \Theta(\mathcal{E}_c)$ from a canonical informant for L or L', it can also learn the language from the canonical informant for $\Theta(L')$ or $\Omega(L')$ which was translated using ω or ϑ, respectively.

▶ **Lemma 4.10.** Let δ be a delayable learning restriction. Given any $L \in \mathcal{E}_c$ and $I \in \mathbf{Inf}_{\mathrm{Can}}(L)$ as well as $I' \in \mathbf{Inf}_{\mathrm{Can}}(\Theta(L))$ and a learner h, let $p_I, p_{\omega(I)}$ be the learning sequences of h on I and $\omega(I')$ respectively. We have

$$(p_I, I) \in \delta \Rightarrow (p_{\omega(I)}, \omega(I')) \in \delta.$$

◀

Proof. First we define a delaying function $r^\omega_{I,I'}$ and show $r^\omega_{I,I'} \in \vec{R}$. We use the delayability of δ to show $(p_I \circ r^\omega_{I,I'}, \omega(I')) \in \delta$ and finally conclude $(p_{\omega(I)}, \omega(I')) \in \delta$. We define $r^\omega_{I,I'} \colon \mathbb{N} \to \mathbb{N}$ so that for all $n \in \mathbb{N}$ we have

$$r^\omega_{I,I'}(n) = \max\{n' \in \mathbb{N} \mid \mathrm{content}(I[n']) \subseteq \mathrm{content}(\omega(I[n]))\}.$$

Intuitively $r^\omega_{I,I'}$ always returns the length of the longest initial sequence of $I[n]$ which is also contained in $\mathrm{content}(\omega(I'[n]))$. We show that $r^\omega_{I,I'} \in \vec{R}$. As $r^\omega_{I,I'}$ ist the maximum of a set which grows monotonically in regards to the $\subseteq$-relation, $r^\omega_{I,I'}$ is non-decreasing. To show that $r^\omega_{I,I'}$ has an infinite limit inferior, we show that for all $i \in \mathbb{N}$ there exists an $m \in \mathbb{N}$ so that $r^\omega_{I,I'}(m) \geq i$. Since $r^\omega_{I,I'}$ is non-decreasing this suffices to show that for all $i \in \mathbb{N}$ there exists an $m \in \mathbb{N}$ so that for all $n \geq m$ we have $r^\omega_{I,I'}(n) \geq i$. For $i = 0$ this claim holds, as any sequence contains the content of the empty sequence. Let the claim be true for an arbitrary i. We show that it holds for $i+1$ as well. Let m' so that $r^{\omega}_{I,I'}(m') \geq i$. If $r^{\omega}_{I,I'}(m') > i$ we have $r^{\omega}_{I,I'}(m') \geq i+1$. Else, let $x := \pi_1(I[i+1])$. Then I' contains information about $2x+1$, so there exist an $m'' \geq m'$ so that without loss of generality $I'(m'') = \langle 2x+1, 0 \rangle$. By the construction of ω we now have $\langle x, 0 \rangle \in \text{content}(\omega(I'[m'']))$. We have $r^{\omega}_{I,I'}(m'') \geq i+1$. So the claim holds in both cases, thus, $r^{\omega}_{I,I'}$ has an infinite limit inferior.

Following the definition of $r^{\omega}_{I,I'}$ we have that, for all $n \in \mathbb{N}$, $\text{content}(I[r^{\omega}_{I,I'}(n)]) \subseteq \text{content}(I'(n))$. Because $\omega(I) \in \mathbf{Inf}_{\mathrm{Can}}(\mathcal{E}^{\infty}_c)$ we have that there exists no $n \in \mathbb{N}$ so that $r^{\omega}_{I,I'}(x) + 1 < r^{\omega}_{I,I'}(x+1)$, so we have $\text{content}(I) = \text{content}(I \circ r^{\omega}_{I,I'})$. Since δ is a delayable learning restriction, this leads us to the conclusion that $(p_I \circ r^{\omega}_{I,I'}, I') \in \delta$. Since, for all $n \in \mathbb{N}$, $r^{\omega}_{I,I'}$ sets $p_I \circ r^{\omega}_{I,I'}$ back to only use the information from $I[n]$ that is also available in $\omega(I'[n])$ we have $p_{\omega(I)} = p_I \circ r^{\omega}_{I,I'}$. This leads us to the conclusion that $(p_{\omega(I)}, \omega(I')) \in \delta$. ■

▶ **Lemma 4.11.** Given any $L \in \mathcal{E}_c$ and $I' \in \mathbf{Inf}_{\mathrm{Can}}(L)$ as well as $I \in \mathbf{Inf}_{\mathrm{Can}}(\Theta(L))$ and a learner h. Let $p_I, p_{\vartheta(I')}$ be the learning sequences of h on I and $\vartheta(I')$ respectively. We have

$$(p_I, I) \in \delta \Rightarrow (p_{\vartheta(I')}, \vartheta(I')) \in \delta.$$

◀

Proof. We define $r^{\vartheta}_{I,I'} : \mathbb{N} \to \mathbb{N}$ so that for all $n \in \mathbb{N}$ we have

$$r^{\vartheta}_{I,I'}(n) = \max\{n' \in \mathbb{N} \mid \text{content}(I[n']) \subseteq \text{content}(\vartheta(I'[n]))\}.$$

Analogous to the proof for Lemma 4.10 we see that $r^{\vartheta}_{I,I'}$ is non-decreasing and has an infinite limit inferior. Following the definition of $r^{\vartheta}_{I,I'}$ for all $n \in \mathbb{N}$ we have $\text{content}(I[r^{\vartheta}_{I,I'}(n)]) \subseteq \text{content}(I'(n))$ and, following the arguements used in the proof for Lemma 4.10, we have $\text{content}(I \circ r^{\vartheta}_{I,I'})(n) = \text{content}(I'(n))$. Since δ is a delayable learning restriction, this leads us to the conclusion, that $(p_I \circ r^{\vartheta}_{I,I'}, I') \in \delta$. Since, for all $n \in \mathbb{N}$, $r^{\vartheta}_{I,I'}$ sets $p_I \circ r^{\vartheta}_{I,I'}$ back to only use the information from $I[n]$ that is also available in $\omega(I'[n])$ we have $p_{\vartheta(I')} = p \circ r^{\vartheta}_{I,I'}$. This leads us to the conclusion that $(p_{\vartheta(I')}, \omega(I')) \in \delta$.

■

▶ **Theorem 4.12.** Let δ be a Θ-Ω-semantic and delayable learning restriction. Then, we have

$$(\mathbf{Inf}_{\mathcal{E}_c}, \mathbf{G}, \delta)_{\mathrm{C}} \hookrightarrow (\mathbf{Inf}_{\mathcal{E}^{\infty}_c}, \mathbf{G}, \delta)_{\mathrm{C}}.$$

◀

Proof. As Aschenbach et al. [AKS18] have shown, it suffices to show that, for an $\mathcal{L} \subseteq \mathcal{E}_c$ we have $\mathcal{L} \in [\mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c}]_{\mathrm{C}}$ if and only if $\Theta(\mathcal{L}) \in [\mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c^\infty}]_{\mathrm{C}}$. We first show that if $\mathcal{L} \in [\mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c}]_{\mathrm{C}}$, then $\Theta(\mathcal{L}) \in [\mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c^\infty}]_{\mathrm{C}}$. Let h be a learner and $\mathcal{L} = \mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c}\mathrm{C}(h)$. We define a new learner g so that for all $I' \in \mathbf{Inf}(\Theta(\mathcal{L}))$ and $x, n \in \mathbb{N}$ we have

$$\varphi_{g(I'[n])}(x) = \begin{cases} 1, & \text{if } x \in 2\mathbb{N}, \\ \varphi_{h(\omega(I'[n]))}(\frac{x-1}{2}), & \text{else.} \end{cases}$$

So g uses the predictions made by h on $\omega(I')$ to predict which elements belong to $\Theta(L)$. Let p_h and p_g be the learning sequence of h on $\omega(I')$ and g on I', respectively. Since h learns $\mathcal{L}$ by Lemma 4.10 we have $(p_h, \omega(I')) \in \delta$. By the definition of g, for all $i \in \mathbb{N}$, we have $\mathrm{C}_{p_g(i)} = \left\{x \in \mathbb{N} \,\middle|\, \varphi_{h(\omega(I'[i]))}(\frac{x-1}{2}) = 1\right\} \cup 2\mathbb{N} = \Omega(\mathrm{C}_{p(i)})$. We also have $\Omega(\mathrm{obj}_{\mathcal{E}_c^\infty}(\omega(I'[i]))) = \mathrm{obj}_{\mathcal{E}_c}(I'[i])$. Since δ is Ω-semantic, we have $(p_g, I') \in \delta$, g $\mathbf{Inf}\mathbf{G}\delta_{\mathrm{C}}$-learns $\Theta(\mathcal{L})$.

We now show that for any $\mathcal{L} \in \mathcal{E}_c$ if $\Theta(\mathcal{L}) \in [\mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c^\infty}]_{\mathrm{C}}$ then $\mathcal{L} \in [\mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c}]_{\mathrm{C}}$. Let g be a learner and $\mathcal{L}$ so that $\Theta(\mathcal{L}) = \mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathcal{E}_c^\infty}\mathrm{C}(g) \cap \mathrm{range}(\Theta)$. We define a new learner h so that for all $I' \in \mathbf{Inf}_{\mathrm{Can}}(\mathcal{L})$ and $x, n \in \mathbb{N}$ we have

$$\varphi_{h(I'[n])}(x) = \varphi_{g(\vartheta(I'[n]))}(2x+1).$$

So h modifies and passes on the information contained in I' and works on the prediction made by g. Let p_g and p_h be the learning sequence of g and h on $\omega(I')$ and I', respectively. Since g learns $\Theta(\mathcal{L})$, by Lemma 4.11 we have $(p_g, \omega(I')) \in \delta$. For all $i \in \mathbb{N}$ we now have $\mathrm{C}_{p_h(i)} = \left\{x \in \mathbb{N} \,\middle|\, \varphi_{g(\omega(I'[i]))}(2x+1) = 1\right\} = \Theta(C_{p_g(i)})$. We also have $\Theta(\mathrm{obj}_{\mathcal{E}_c}(\omega(I'[i]))) = \mathrm{obj}_{\mathcal{E}_c^\infty}(I'[i])$. As δ is Θ-sematic, we have $(p_h, I') \in \delta$, so h $\mathbf{Inf}_{\mathrm{Can}}\mathbf{G}\delta_{\mathrm{C}}$-learns $\mathcal{L}$. ■

▶ **Corollary 4.13.** For delayable and semantic learning restriction we have

$$(\mathbf{Arb}_{\mathcal{R}_{0,1}}, \mathbf{G}, \delta) \hookrightarrow (\mathbf{Inf}_{\mathcal{E}_c}, \mathbf{G}, \delta)_{\mathrm{C}} \hookrightarrow (\mathbf{Inf}_{\mathcal{E}_c^\infty}, \mathbf{G}, \delta)_{\mathrm{C}}$$
$$\hookrightarrow (\mathbf{Arb}_{\mathcal{R}_{0,1}^\infty}, \mathbf{G}, \delta) \hookrightarrow (\mathbf{Arb}_{\mathcal{R}}, \mathbf{G}, \delta)$$

◀

Proof. This follows directly from Theorem 4.12, the isomorphism shown in Theorem 4.4, Corollary 4.5 and Theorem 4.3 and Corollary 3.7. ■

4.3 Non isomorphic learning tasks

In this section we show that there are sets of learning settings between which no isomorphism exists. We proof this for semantic learning restrictions δ and $\mathbf{InfG}\delta_{\mathbf{W}}$ and $\mathbf{InfG}\delta_C$ by finding two semantic restrictions with different relationships inf **W**- and C-index learning.

▶ **Lemma 4.14.** We have $[\mathbf{InfGBc}]_{\mathrm{C}} = [\mathbf{InfGBc^c}]_{\mathrm{C}}$. ◀

Proof. With *s-m-n* we define $f : \mathbb{N} \to \mathbb{N}$ so that for all $e, n' \in \mathbb{N}$ we have

$$\varphi_{f(e)}(n') = \begin{cases} 1, & \text{if } \varphi_e(n') = 0, \\ 0, & \text{if } \varphi_e(n') = 1, \\ \bot, & \text{else.} \end{cases}$$

For an any $L \in \mathcal{E}_c$ and $e \in \mathbb{N}$ for which $\varphi_e = \chi_L$, we now have $\varphi_{f(e)} = \chi_{\overline{L}}$.

To show $[\mathbf{InfGBc}]_{\mathrm{C}} \subseteq [\mathbf{InfGBc^c}]_{\mathrm{C}}$, let a learner h be given. Let $\mathcal{L} = \mathbf{InfGBc}(h)$. This means that for all $L \in \mathcal{L}, I \in \mathbf{Inf}(L)$, with p_h denoting the hypothesis sequence of h on I, there exists an $n_0 \in \mathbb{N}$ so that for all $n \geq n_0$ we have $C_{p_h(n)} = L$. We define a new learner g so that for all Informants I and $n \in \mathbb{N}$ we have

$$g(I'[n]) = f(h(I'[n])).$$

Given an $I \in \mathbf{Inf}(\mathcal{L})$, let p_g be the hypothesis sequence of g on I. For all $n \geq n_0$ we now have $\varphi_{p_g(n)} = \chi_{\overline{L}}$, so $\mathrm{C}_{p(n)} = \overline{L}$. This shows that $(p_g, I) \in \mathbf{Bc^c}$, so $\mathbf{InfGBc^c}(g) \supseteq \mathbf{InfGBc}(h)$. The other inclusion, that is $[\mathbf{InfGBc}]_{\mathrm{C}} \supseteq [\mathbf{InfGBc^c}]_{\mathrm{C}}$, follows analogously. ■

▶ **Lemma 4.15.** We have $[\mathbf{InfGBc}]_{\mathbf{W}} \setminus [\mathbf{InfGBc^c}]_{\mathbf{W}} \neq \emptyset$. ◀

Proof. We define

$$\mathcal{L}_0 = \left\{ M \in \mathcal{E}_c \;\middle|\; M \neq \emptyset \wedge \mathbf{W}_{\min(M)} = M \right\}. \tag{4.2}$$

It is well known that $\mathcal{L}_0$ is $\mathbf{TxtGEx}_{\mathbf{W}}$-learnable [CL82; Jai+99]. We include the proof, that it is $\mathbf{InfGBc}_{\mathbf{W}}$-learnable for completeness. We want to show that there is a learner h, so that $\mathcal{L}_0 \subseteq \mathbf{InfGBc}(h)$. We define h so that for all Informants I and $n \in \mathbb{N}$ we have

$$h(I[n]) = \min(\mathrm{pos}(I[n])).$$

Let $L \in \mathcal{L}_0$, $I \in \mathbf{Inf}(L)$ and p be the hypothesis sequence of h on I. There is an $n_0 \in \mathbb{N}$ so that $I(n_0) = \langle \min(L), 1 \rangle$. So for all $n \geq n_0$ we have $p(n) = \min(L)$. By the definition of $\mathcal{L}_0$ we have $\mathbf{W}_{p(n)} = \mathbf{W}_{\min(L)} = L$. Thus, we have $(p, I) \in \mathbf{Bc}$, so h **InfGBc**-learns $\mathcal{L}_0$.

We show that there is no learner g so that $\mathcal{L}_0 \subseteq \mathbf{InfGBc}^{\mathrm{c}}(g)$ by contradiction. Suppose there is such a learner g. First we define a function so that for a finite sequence σ_n and $s \in \mathbb{N}$ we have that for an $I \in \mathbf{Inf}_{\mathbf{Can}}(\mathrm{range}(\sigma))$

$$ci(\sigma, s) = I[s].$$

So $ci(\sigma_n, s)$ gives us a canonical informant of length s with $\mathrm{pos}(ci(\sigma_n, s)) = \mathrm{range}(\sigma_n)$ if $s \geq \max(\mathrm{range}(n))$. Further, for any finite sequence σ_n let

$$\mathrm{notin}(\sigma_n) = \{0, 1, \ldots, \max(\sigma_n)\} \setminus \mathrm{range}(\sigma_n).$$

Now the Operator Recursion Theorem [Cas74] yields a computable function σ and a Gödel number e_0 so that the following holds for all $a, n \in \mathbb{N}$

$$\varphi_{e(0)}(a) = \begin{cases} \sigma(a), & \text{if } \exists b \in \mathbb{N} \colon a = \sigma(b), \\ \bot, & \text{else.} \end{cases}$$

$$\sigma(0) = e_0,$$

$$\sigma(n+1) = \begin{cases} x, & \text{if } \exists s, t, x \in \mathbb{N} \colon x \geq t \geq s \geq \sigma(n) \\ & \wedge\, x \in \mathbf{W}^t_{g(ci(\sigma[n], s))} \setminus \mathrm{notin}(\sigma[n]), \\ \bot, & \textit{else}. \end{cases}$$

Not that σ is either total or only defined for all $n \in \mathbb{N}$ until some $n_0 \in \mathbb{N}$. Futhermore it is strongly monotone for all n that it is defined on. Let $A = \mathrm{range}(\sigma)$. By definition of σ we have $\sigma(0) = e_0 = \min(\mathrm{range}(\sigma)) = \min(A)$ and by the definition A we have $\mathbf{W}_{e_0} = \mathrm{range}(\sigma) = A$. Also A is either finite or the range of a total strictly monotone function, so $A \in \mathcal{E}_c$. Thus, we have $A \in \mathcal{L}_0$. By our assumption g $\mathbf{InfGBc}^{\mathrm{c}}$-learns A. Let I be the canonical informant for A, and let p_g be the hypotheses sequence of q on I. We have $(p_g, I) \in \mathbf{Bc}^{\mathrm{c}}$. There exists an $n_0 \in \mathbb{N}$ so that for all $n \geq n_0$ we have $W_{p_g(n)} = \overline{A}$.

Suppose σ is infinite. For n_0 there exists $x, s, t \in \mathbb{N}$ with $x \geq s \geq t$, so that $x \in \mathbf{W}^t_{g(ci(\sigma(n_0),s))}$, but by definition of σ we have $x \in A$. Thus, we have $\mathbf{W}_{p(n_0)} \neq \overline{A}$. In this case $A \notin \mathbf{InfGBc}^{\mathrm{c}}(g)$.

In the other case σ is finite. So there exists an $n' \in \mathbb{N}$ for which we have $\sigma(n') \downarrow$, but $\sigma(n'+1) \uparrow$. For all $s \geq n'$ we now have that $W_{g(ci(\sigma(n'),s))}$ is finite. This especially means that $\mathbf{W}_{p(\max(n',n_0))}$ is finite, so we have $\mathbf{W}_{p(\max(n',n_0))} \neq \overline{A}$, as A is finite. So in this case g also can not $\mathbf{InfGBc^c}$-learn A. ■

▶ **Theorem 4.16.** There does not exist an isomorphism $\Theta : \mathcal{E} \to \mathcal{E}$ so that for all semantic learning restrictions δ and $\mathcal{L} \subseteq \mathcal{E}$ we have $L \in [\mathbf{InfG}\delta]_\mathrm{C}$ if and only if $\Theta(L) \in [\mathbf{InfG}\delta]_\mathrm{W}$. ◀

Proof. We proof this claim by contradiction. Suppose we have such an isomorphism Θ. Then for all $L \subseteq \mathcal{E}$ we would have $L \in [\mathbf{InfGBc}]_\mathrm{C} \Leftrightarrow \Theta(L) \in [\mathbf{InfGBc}]_\mathrm{W}$ and $L \in [\mathbf{InfGBc^c}]_\mathrm{C} \Leftrightarrow \Theta(L) \in [\mathbf{InfGBc^c}]_\mathrm{W}$. With Lemma 4.14 we have $[\mathbf{InfGBc}]_\mathrm{C} = [\mathbf{InfGBc^c}]_\mathrm{C}$, which gives us $[\mathbf{InfGBc}]_\mathrm{W} = [\mathbf{InfGBc^c}]_\mathrm{W}$, a contradiction to Lemma 4.15. ■

▶ **Corollary 4.17.** There does not exist an isomorphism $\Theta : \mathcal{E} \to \mathcal{R}_{0,1}$ so that for all semantic learning restrictions δ and $\mathcal{L} \subseteq \mathcal{E}$ we have $L \in [\mathbf{InfG}\delta]_\mathrm{W}$ if and only if $\Theta(L) \in [\mathbf{ArbG}\delta]$. ◀

Proof. This follows directly from the isomorphism between $(\mathbf{Inf}_{\mathcal{E}_c}, \mathbf{G}, \delta)_\mathrm{W}$ and $(\mathbf{Arb}_{\mathcal{R}_{0,1}}, \mathbf{G}, \delta)$ shown in Theorem 4.4, the trasitivity of $\cong$ shown in Theorem 3.6 and Theorem 4.16. ■

Bibliography

[AKS18] Martin Aschenbach, Timo Kötzing, and Karen Seidel. **Learning from informants: relations between learning success criteria**. *CoRR* (2018). ArXiv preprint (see pages 1, 30).

[Ang80] Dana Angluin. **Inductive inference of formal languages from positive data**. *Information and control* 45:2 (1980), 117–135 (see pages 11, 12).

[Bal+08] Ganesh Baliga, John Case, Wolfgang Merkle, Frank Stephan, and Rolf Wiehagen. **When unlearning helps**. *Information and Computation* 206:5 (2008), 694–709 (see page 14).

[Ber+21] Julian Berger, Maximilian Böther, Vanja Doskoč, Jonathan Gadea Harder, Nicolas Klodt, Timo Kötzing, Winfried Lötzsch, Jannik Peters, Leon Schiller, Lars Seifert, et al. **Learning languages with decidable hypotheses**. In: *Conference on Computability in Europe*. Springer. 2021, 25–37 (see page 6).

[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).

[Cas74] John Case. **Periodicity in generations of automata**. *Mathematical Systems Theory* 8:1 (1974), 15–32 (see page 32).

[CL82] John Case and Christopher Lynes. **Machine inductive inference and language identification**. In: *International Colloquium on Automata, Languages, and Programming*. Springer. 1982, 107–115 (see page 31).

[CM11] John Case and Samuel E Moelius III. **Optimal language learning from positive data**. *Information and Computation* 209:10 (2011), 1293–1311 (see page 14).

[Gol67] E Mark Gold. **Language identification in the limit**. *Information and control* 10:5 (1967), 447–474 (see pages 1, 6, 14).

[Jai+99] Sanjay Jain, Daniel Osherson, James S Royer, Arun Sharma, et al. **Systems that learn: an introduction to learning theory**. MIT press, 1999 (see pages 12, 31).

[Jan91] Klaus P Jantke. **Monotonic and non-monotonic inductive inference**. *New Generation Computing* 8:4 (1991), 349–360 (see page 12).

[KKS21] Ardalan Khazraei, Timo Kötzing, and Karen Seidel. **Towards a map for incremental learning in the limit from positive and negative information**. In: *Conference on Computability in Europe*. Springer. 2021, 273–284 (see pages 1, 6).

[Köt09] Timo Kötzing. **Abstraction and complexity in computational learning in the limit**. University of Delaware, 2009 (see pages 1, 6).

[Köt17] Timo Kötzing. **A solution to Wiehagen's thesis**. *Theory of Computing Systems* 60:3 (2017), 498–520 (see page 2).

[KP16] Timo Kötzing and Raphaela Palenta. **A map of update constraints in inductive inference**. *Theoretical Computer Science* 650 (2016), 4–24 (see pages 7, 14).

[KS16] Timo Kötzing and Martin Schirneck. **Towards an atlas of computational learning theory**. In: *Symposium on Theoretical Aspects of Computer Science*. Vol. 47. Leibniz International Proceedings in Informatics. Dagstuhl, Germany, 2016, 47:1–47:13 (see page 1).

[KSS17] Timo Kötzing, Martin Schirneck, and Karen Seidel. **Normal forms in semantic language identification**. In: *International Conference on Algorithmic Learning Theory*. PMLR. 2017, 493–516 (see page 2).

[Moh22] Niklas Mohrin. **Behaviorally correct learning from informants**. Bachelor's thesis. Universität Potsdam, 2022 (see page 1).

[OSW86] Daniel N Osherson, Michael Stob, and Scott Weinstein. **Systems that learn: An introduction to learning theory for cognitive and computer scientists.** The MIT Press, 1986 (see page 14).

[Wie90] Rolf Wiehagen. **A thesis in inductive inference**. In: *International Workshop on Nonmonotonic and Inductive Logic*. Springer. 1990, 184–207 (see page 14).

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

Paula Marten