SYNTACTIC CHARACTERIZATION OF LEARNABILITY OF STRUCTURES WITH MIND CHANGES

EKATERINA B. FOKINA AND STEFFEN LEMPP

Abstract. We present syntactic characterizations of learnability of some classes of structures under various criteria on the information source and the convergence behavior of the learner.

1. Introduction

The syntactic characterization of computability-theoretic properties of structures has been a long-standing theme in computability theory. Since Post's theorem relating definability in the arithmetical hierarchy with computability from finite jump iterations [25], various notions have been characterized syntactically. In computable structure theory, the characterizations of relative computable categoricity or of relative intrinsically c.e. relations in structures [4, 11] are such examples. In this paper, we characterize several notions of classes of structures that are "learnable with mind changes".

Our general setting is the following. Suppose we are given an at most countable class of countable structures $\mathfrak{K}$. Suppose further that we receive step by step finitely much information about one of the structures $\mathcal{A}$ from $\mathfrak{K}$. Our goal is to correctly identify, after finitely many steps, which structure we are observing. Depending on the way the structure is revealed and on the criteria of correct identification of the structure, we may get different versions of the task. We formalize the setting using the notions of algorithmic learning theory applied to computable structure theory.

Classical algorithmic learning theory goes back to the work of Putnam [26] and Gold [17]. A learner M receives step by step more and more data (a finite amount at each step) on an object X to be learned, and M outputs a sequence of hypotheses that converges to a finitary description of X. The main body of work in algorithmic learning theory has been done for (classes of) formal languages or recursive functions, see the monograph [20].

Within the framework of computable structure theory, the work of Glymour and Martin and Osherson [16, 23, 22] initiated the study of learnable classes of structures. Later on, Stephan and Ventsov [27] investigated the learnability for

2020 *Mathematics Subject Classification.* Primary: 68Q32; Secondary: 03B70.

Key words and phrases. computational learning theory, computable structure theory, learning structures.

The first author's research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/P36781. For open access purposes, the first author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. The second author's research was partially supported by AMS-Simons Foundation Collaboration Grant 626304. Part of this research was carried out while the second author was visiting Technical University of Vienna as a guest professor, and he would like to express his gratitude to this university. The authors would also like to thank Luca San Mauro for a helpful suggestion.

classes of substructures of a given computable structure $\mathcal{S}$. This approach was further developed, e.g., in the papers [18, 15].

Fokina, Kötzing, and San Mauro [13] reworked and generalized the approach from [16, 23]. They studied learnability of various classes $\mathfrak{K}$ of computable equivalence relations. For these $\mathfrak{K}$, they considered *learnability from informant* (or **InfEx**-learnability) and *learnability from text* (**TxtEx**-learnability) for equivalence structures, to be formally defined in Section 2. Further work [9] extended the notion of **InfEx**-learnability to arbitrary countable families of computable structures and resulted in a model-theoretic characterization of **InfEx**-learnability. Details will follow in Section 2. An analogous result for **TxtEx**-learnability recently appeared in [7]. Some of the results from [13, 9, 7] already appeared in [16, 23, 22], however, the authors of [13, 9, 7] rediscovered and reproved the results using modern computability-theoretic terminology and new methods that allowed to deduce new corollaries. More related results can be found, e.g., in [5, 6].

In this paper, we consider several new notions of learning of structures where we allow mind changes. As the main results of the paper, we give syntactic characterizations of the notions of learning arising in various ways.

2. Terminology and set-up

Our structures are countable in a fixed relational at most countable signature $\mathcal{L}$. We denote by $\mathrm{Mod}(\mathcal{L})$ the set of all $\mathcal{L}$-structures $\mathcal{A}$ with $\mathrm{dom}(\mathcal{A}) \subseteq \omega$. We consider at most countable classes $\mathfrak{K} \subseteq \mathrm{Mod}(\mathcal{L})$ of structures and, unless stated otherwise, assume these classes are closed under isomorphism.

Let Atm denote the set of (the Gödel numbers) of all positive and negative atomic sentences in the signature $\mathcal{L} \cup \omega$ (in other words, positive and negative atomic facts about possible $\mathcal{L}$-structures on the domain ω). The restriction of Atm to only positive atomic sentences is denoted by Atm_+. For a structure $\mathcal{A}$ with $\mathrm{dom}(\mathcal{A}) \subseteq \omega$, we denote by $D(\mathcal{A})$ the *atomic* diagram of $\mathcal{A}$, i.e., a subset of Atm of $\mathcal{L} \cup \omega$-sentences true in $\mathcal{A}$, and by $D_+(\mathcal{A})$ the *positive* atomic diagram of $\mathcal{A}$.

We now introduce the components of our learning framework. Let $\mathfrak{K} \subseteq \mathrm{Mod}(\mathcal{L})$ contain precisely κ isomorphism types, where $\kappa \leq \omega$, and denote the types of $\mathcal{L}$-structures as $\mathcal{A}_i$, $i \in \kappa$.

- The *learning domain* (LD) is the collection of all copies $\mathcal{S}$ of the structures from $\mathfrak{K}$ such that $\mathrm{dom}(\mathcal{S}) \subseteq \omega$, i.e.,
$$\mathrm{LD}(\mathfrak{K}) = \bigcup_{i \in \kappa} \{\mathcal{S} \in \mathrm{Mod}(\mathcal{L}) : \mathcal{S} \cong \mathcal{A}_i\}.$$
- The *hypothesis space* (HS) contains the indices i for $\mathcal{A}_i \in \mathfrak{K}$ (an index is viewed as a conjecture about the isomorphism type of an input structure $\mathcal{S}$) and a question mark symbol:
$$\mathrm{HS}(\mathfrak{K}) = \kappa \cup \{?\}.$$
- A *learner* L is a function from the set $(\mathrm{Atm})^{<\omega}$ (i.e., the set of all finite tuples of atomic facts) into $\mathrm{HS}(\mathfrak{K})$. That is, L receives as input some atomic facts about a given structure from $LD(\mathfrak{K})$ and is required to output conjectures from $HS(\mathfrak{K})$ about the observed structure. Notice here that we do not require any effectiveness of the learning function. For this reason we also do not impose any computability-theoretic restrictions on the complexity of the enumeration $\{\mathcal{A}_i\}_{i \in \kappa}$ of structures from $\mathfrak{K}$.

Depending on the way the atomic facts are revealed to the learner and on the criteria of what it means to correctly learn the class, we obtain different notions of learning of structures. In classical algorithmic learning theory, two main sources of information are informant and text (e.g., [17, 2, 19]). Adapted to the case of structures, the definitions appear as follows.

- For an $\mathcal{L}$-structure $\mathcal{S}$, an *informant* $\mathbb{I}$ for $\mathcal{S}$ is an arbitrary sequence $\{\psi_i\}_{i\in\omega}$ containing elements from Atm and satisfying
$$\mathcal{D}(\mathcal{S}) = \{\psi_i : i \in \omega\}.$$
- For an $\mathcal{L}$-structure $\mathcal{S}$, a *text* $\mathbb{T}$ for $\mathcal{S}$ is an arbitrary sequence $\{\psi_i\}_{i\in\omega}$ containing elements from Atm_+ and satisfying
$$\mathcal{D}_+(\mathcal{S}) = \{\psi_i : i \in \omega\}.$$
- For $k \in \omega$, by $\mathbb{I} \upharpoonright k$ (or $\mathbb{T} \upharpoonright k$, respectively), we denote the corresponding sequence $\{\psi_i\}_{i<k}$.

The two notions of learning used most frequently are the following.

Definition 2.1 ([10])**.** We say that the family $\mathfrak{K}$ is **InfEx**-*learnable* if there exists a learner L such that for any structure $\mathcal{S} \in \mathrm{LD}(\mathfrak{K})$ and any informant $\mathbb{I}_\mathcal{S}$ for $\mathcal{S}$, the learner eventually stabilizes to a correct conjecture about the isomorphism type of $\mathcal{S}$. More formally, there exists a limit
$$\lim_{n\to\omega} L(\mathbb{I}_\mathcal{S} \upharpoonright n) = i$$
belonging to ω, and $\mathcal{A}_i$ is isomorphic to $\mathcal{S}$.

Definition 2.2 ([7])**.** We say that the family $\mathfrak{K}$ is **TxtEx**-*learnable* if there exists a learner L such that for any structure $\mathcal{S} \in \mathrm{LD}(\mathfrak{K})$ and any text $\mathbb{T}_\mathcal{S}$ for $\mathcal{S}$, the learner eventually stabilizes to a correct conjecture about the isomorphism type of $\mathcal{S}$. More formally, there exists a limit
$$\lim_{n\to\omega} L(\mathbb{T}_\mathcal{S} \upharpoonright n) = i$$
belonging to ω, and $\mathcal{A}_i$ is isomorphic to $\mathcal{S}$.

The prefix "**Ex**" in the above definitions stands for "explanatory" learning, meaning syntactic convergence of the learner in the limit. In Section 4 we will require a different convergence behavior from the learner.

Definition 2.3. A function $F : 2^{<\omega} \to$ Ordinals is an *ordinal mind change counter function* if for all $\sigma \in 2^{<\omega}$, $F(\sigma) \leq F(\sigma^-)$. We say that a learner L, with associated ordinal mind change counter function F, *α-learns* a family $\mathfrak{K}$ from informant if the following three conditions hold:

(1) L **InfEx**-learns $\mathfrak{K}$;
(2) $F(\lambda) = \alpha$, where λ is the empty string;
(3) for every $\mathcal{S} \in \mathrm{LD}(\mathfrak{K})$, for every informant $\mathbb{I}_\mathcal{S}$ for $\mathcal{S}$ and for every $n \in \omega$, if $? \neq L(\mathbb{I}_\mathcal{S} \upharpoonright n) \neq L(\mathbb{I}_\mathcal{S} \upharpoonright n+1)$, we have
$$F(\mathbb{I}_\mathcal{S} \upharpoonright n) > F(\mathbb{I}_\mathcal{S} \upharpoonright n+1).$$

The family $\mathfrak{K}$ is *α-learnable* if there exists a learner L that α-learns it.

The papers [10] and [7] give syntactic characterizations of **InfEx**- and **TxtEx**-learning in terms of infinitary formulas or, respectively, positive infinitary formulas, as defined in [3, 8].

Theorem 2.4 ([10, 7]). *Let $\mathfrak{K} = \{\mathcal{A}_i : i \in \omega\}$ be a family of structures such that $\mathcal{A}_i \not\cong \mathcal{A}_j$ for $i \neq j$. Then $\mathfrak{K}$ is* **InfEx***-learnable (or* **TxtEx***-learnable, respectively) if and only if there exists a family of infinitary (or positive infinitary, respectively) Σ_2-sentences $\{\psi_i : i \in \omega\}$ such that for each i, $\mathcal{A}_i$ is the only member of $\mathfrak{K}$ satisfying ψ_i.*

In this paper we modify the allowed information sources and the convergence behavior of the learner. In Section 3, we allow the information to be given in a Δ^0_2-way. For each atomic sentence from the diagram, we allow a change of mind finitely many times, but we require that from some point on, the correct fact is revealed to the learner correctly. In Section 4, we syntactically characterize n-learning, that is, learning from informant with fixed finite number of answers given by the learner in the attempt to learn the structure.

3. Δ^0_2-learning and d.c.e.-learning

Before we give new definitions, we introduce some notations. Let $p \in 2^\omega$ be a path in Cantor space, let $\{\psi_i\}_{i\in\omega}$ be a fixed sequence of first-order $\mathcal{L} \cup \omega$-formulas, usually intended to be elements of Atm or Atm_+, and let $\mathcal{A}$ be a structure. We say that p *correctly codes the fact whether* $\mathcal{A} \models \psi_i$ if we have that $\mathcal{A} \models \psi_i$ iff $p(i) = 1$. Having fixed a sequence of formulas $\{\psi_i\}_{i\in\omega}$ and a path $p \in 2^\omega$, we write $L(p \upharpoonright n)$ for $L(\psi_0^{p(0)}, \psi_1^{p(1)}, \ldots, \psi_{n-1}^{p(n-1)})$. Here, for a formula ψ, we set $\psi^1 = \psi$ and $\psi^0 = \neg\psi$. Under this convention we consider learners to be functions from $2^{<\omega}$ to $\omega \cup \{?\}$.

We are now ready to formally define the notion of Δ^0_2-learnability:

Definition 3.1. Fix an infinite sequence $(\varphi_i)_{i\in\omega}$ of all elements from Atm such that each Atm-formula appears on the list infinitely often. Call a family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) Δ^0_2-*learnable* if there is a learner L with the following property: For any path $p \in 2^\omega$, if there is a model $\mathcal{A}_i$ such that for all j with $\mathcal{A}_i \models \varphi_j$, we have that for cofinitely many instances of φ_j, p correctly codes this fact, then for cofinitely many n, $L(p \upharpoonright n) = i$.

The intuition behind the above definition is that the path p does not reveal the atomic diagram of a model outright, but only in the limit: For each atomic formula, it may give the wrong answer finitely often before settling on the correct answer forever. Still, the learner must be able to guess the intended model after only finitely many wrong guesses.

Perhaps surprisingly, this version of learnability has a very simple syntactic characterization:

Theorem 3.2. *A family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) is Δ^0_2-learnable iff there is a sequence $\{\psi_i\}_{i\in\omega}$ of (finite) existential $\mathcal{L}$-sentences such that $\mathcal{A}_i \models \psi_j$ iff $i = j$.*

Proof. Suppose first there is such a sequence $\{\psi_i\}_{i\in\omega}$ of (finite) existential $\mathcal{L}$-sentences. Then we can define a learner function L which, on input σ, checks for the least $k < |\sigma|$ such that φ_k is one of the formulas ψ_i and such that for all instances $\varphi_{k'}$ of φ_k with $k' \geq k$ occurring along σ, σ correctly evaluates (after replacing the constants from ω by existential quantifiers) the sentences $\varphi_{k'}$ according to $\mathcal{A}_i$.

Conversely, suppose that L is a Δ^0_2-learner function but for the sake of a contradiction also that for some model $\mathcal{A}_i$, there is no such sentence ψ_i. Then we can construct a model $\mathcal{B} \in \mathfrak{K}$, coded by a path $p \in 2^\omega$, as follows: Start by building a copy of $\mathcal{A}_i$ until we reach an initial segment σ of p with $L(\sigma) = i$. Since there is

no ψ_i, there must be some $j \neq i$ such that we can now continue from σ by building a copy of $\mathcal{A}_j$ until we reach an initial segment τ of p with $L(\tau) = j$. There is one fine point: When we extend the fragment σ to the fragment τ, we will first complete σ (mentioning elements $\bar{x}$, say) to the full atomic $(\mathcal{L} \restriction |\bar{x}|)$-diagram σ' of $\bar{x}$, which is finite, and only then start extending to a fragment of $\mathcal{A}_j$ to ensure that we do not change any atomic information about $\bar{x}$ in the language $\mathcal{L} \restriction |\bar{x}|$ in τ; in fact, this will ensure that not only is the information along the path p given in a Δ^0_2-way but actually in a d.c.e. way (in fact, in a weakly d.c.e. way per the definition in Epstein/Haas/Kramer [12]): We change our mind about the truth of an atomic formula at most once (from negative to positive or vice versa), when switching from building a $\mathcal{A}_j$ to building $\mathcal{A}_i$.

Once we see $L(\tau) = j$, we revert to σ, or rather σ', declaring all the atomic sentences in τ but not σ' to be possibly false, and continue again with building a copy of $\mathcal{A}_i$, etc. So in the limit, we will either always build a copy of $\mathcal{A}_i$, or a copy of some $\mathcal{A}_j$ from some point on, while the learner does not output the correct index. Or we switch infinitely often between building a copy of $\mathcal{A}_i$ and a copy of some other $\mathcal{A}_j$, all along truly building a copy of $\mathcal{A}_i$, and the learner does not converge to an index. □

Remark 3.3. As noted in the proof above, for Theorem 3.2, we actually show that the conditions are also equivalent to what one might call "weakly d.c.e.-learning from positive and negative information", namely, the learner is given the information about each atomic sentence infinitely often but along any path p, the guess about the truth or falsity changes at most once, from "false" to "true", or from "true" to "false".

We next explore more restrictive kinds of learning but only from positive information:

Definition 3.4. Adopt the notation of Definition 3.1 but restrict the sentences φ_i considered to elements of Atm_+, i.e., only *positive* atomic sentences.

(1) Call a family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) *c.e.-learnable* if there is a learner L with the following property: For any path $p \in 2^\omega$, if there is a model $\mathcal{A}_i$ such that for all k with $\mathcal{A}_i \models \varphi_k$, we have that for cofinitely many instances of φ_k, p correctly codes this fact, and once p codes that $\mathcal{A}_i \models \varphi_k$ is true, it will never change its mind later along p, then for cofinitely many n, $L(p \restriction n) = i$.

(2) Call a family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) *d.c.e.-learnable* if there is a learner L with the following property: For any path $p \in 2^\omega$, if there is a model $\mathcal{A}_i$ such that for all k with $\mathcal{A}_i \models \varphi_k$, we have that for cofinitely many instances of φ_k, p correctly codes this fact, and once p codes that $\mathcal{A}_i \models \varphi_k$ is true, it will change its mind along p at most once more, then for cofinitely many n, $L(p \restriction n) = i$.

Interestingly, these two notions of learnability behave quite differently:

Theorem 3.5. *A family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) is c.e.-learnable iff there is a sequence $\{\psi_i\}_{i\in\omega}$ of positive infinitary Σ_2-sentences in $\mathcal{L}$ such that $\mathcal{A}_i \models \psi_j$ iff $i = j$, i.e., iff it is* **TxtEx***-learnable.*

Proof. From [9], we know that $\mathfrak{K}$ is **TxtEx**-learnable iff there is a sequence $\{\psi_i\}_{i\in\omega}$ of positive infinitary Σ_2-sentences in $\mathcal{L}$ such that $\mathcal{A}_i \models \psi_j$ iff $i = j$. Now clearly,

if $\mathfrak{K}$ is c.e.-learnable, then it is **TxtEx**-learnable. Conversely, suppose $\mathfrak{K}$ is **TxtEx**-learnable, and suppose we are given a labeled tree coding positive information about the structures $\mathcal{A}_i$ in the sense of c.e.-learning. Now convert a regular learner function L from text into a learner function $\hat{L}$ from positive information given in a c.e. way in the sense of Definition 3.4 as follows: Given $\sigma \in 2^{<\omega}$ coding positive information in a c.e. way, let $\mathbb{T}_\sigma$ be the collection of all positive facts about the structure that σ has given. Then $\mathbb{T} = \bigcup_{|\sigma|\to\infty} \mathbb{T}_\sigma$ will be a text which, as $\sigma \subset p$ increases along a path coding a model in $\mathfrak{K}$, represents true positive information about the structure. So $\hat{L}$ uses $L(\mathbb{T}_\sigma)$ for longer and longer sequences σ to compute the correct index i of the structure $\mathcal{A}_i$ coded by p in the limit. □

Unfortunately, the situation for d.c.e.-learnability is quite involved, and we do not have a complete syntactic characterization of it. Just to demonstrate the difficulties, we will start by giving a complete characterization in a very special case only, the language of one unary predicate U:

Proposition 3.6. *A family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) in the language $\mathcal{L} = \{U\}$ is d.c.e.-learnable iff $k \leq 2$; and if $k = 2$, then $U^{\mathcal{A}_0}$ is finite in $\mathcal{A}_0$ and $U^{\mathcal{A}_1}$ is cofinite in $\mathcal{A}_1$ or vice versa.*

Proof. The result is trivial if $k \leq 1$, so assume $k \geq 2$.

We will consider three cases, each assuming the existence of models $\mathcal{A}_0$ and $\mathcal{A}_1$ in the class $\mathfrak{K}$, which will exhaust all possibilities up to symmetry:

Case 1: $U^{\mathcal{A}_0}$ is finite in $\mathcal{A}_0$, $U^{\mathcal{A}_1}$ is coinfinite in $\mathcal{A}_1$, and $|U^{\mathcal{A}_0}| < |U^{\mathcal{A}_1}|$: We will need to show that $\mathfrak{K}$ is not d.c.e.-learnable. Toward a contradiction, we assume the existence of a d.c.e. learner L and show that it can fail: Build a model $\mathcal{A}$ (which will be either $\mathcal{A}_0$ or $\mathcal{A}_1$) as follows, letting $m = |U^{\mathcal{A}_0}|$: Declare the first m elements of ω to satisfy U once and for all but no others for now, waiting for the learner to guess that $\mathcal{A} = \mathcal{A}_0$ at a stage s_0, say. From now on, declare every other new element to be in U until we reach $|U^{\mathcal{A}_1}|$ many (if this number is finite) and wait for the learner to guess that $\mathcal{A} = \mathcal{A}_1$ at a stage s_1, say. Then declare these new elements of $U^{\mathcal{A}}$ (excluding the first m many) to not satisfy U after all, and wait again for the learner to guess that $\mathcal{A} = \mathcal{A}_0$, etc. So either, the learner eventually fails to make a guess and we build one of $\mathcal{A}_0$ or $\mathcal{A}_1$; or the learner does not converge in its guesses and we actually build $\mathcal{A}_0$.

Case 2: $U^{\mathcal{A}_0}$ is cofinite in $\mathcal{A}_0$, $U^{\mathcal{A}_1}$ is infinite in $\mathcal{A}_1$, and $|A_0 - U^{\mathcal{A}_0}| < |A_1 - U^{\mathcal{A}_1}|$: The argument is similar to Case 1 but not symmetric due to the asymmetry about positive information: Again, toward a contradiction, we assume the existence of a d.c.e. learner L and show that it can fail: Build a model $\mathcal{A}$ (which will be either $\mathcal{A}_0$ or $\mathcal{A}_1$) as follows, letting $m = |A_0 - U^{\mathcal{A}_0}|$: Commit to never declare the first m elements of ω to satisfy U and then declare each new element to satisfy U until the learner guesses that $\mathcal{A} = \mathcal{A}_0$. Now, for new elements, declare only every other element to satisfy U until we reach $|A_1 - U^{\mathcal{A}_1}|$ many elements in $A - U^{\mathcal{A}}$ (if this number is finite) and wait for the learner to guess that $\mathcal{A} = \mathcal{A}_1$. Then declare these new elements of $A - U^{\mathcal{A}}$ (excluding the first m many) to satisfy U after all, and wait again for the learner to guess that $\mathcal{A} = \mathcal{A}_0$, etc. So either, the learner eventually fails to make a guess and we build one of $\mathcal{A}_0$ or $\mathcal{A}_1$; or the learner does not converge in its guesses and we actually build $\mathcal{A}_0$.

Case 3: $U^{\mathcal{A}_0}$ is finite in $\mathcal{A}_0$, $U^{\mathcal{A}_1}$ is cofinite in $\mathcal{A}_1$, and $k = 2$: In this case, we need to show that $\mathfrak{K} = \{\mathcal{A}_0, \mathcal{A}_1\}$ is d.c.e.-learnable, so let $m = |U^{\mathcal{A}_0}|$ and $n = |A_1 - U^{\mathcal{A}_1}|$. Given a path p describing a model $\mathcal{A}$ in $\mathfrak{K}$, the informant has to approximate for each element of ω whether it is in $U^{\mathcal{A}}$ or not. So the learner keeps track of the numerically smallest element $x \in \omega$ such that either there are $m+1$ many elements $\leq x$ currently in $U^{\mathcal{A}}$; or such that there are currently $n+1$ many elements $\leq x$ not in $U^{\mathcal{A}}$. In the first case, the learner guesses that $\mathcal{A} = \mathcal{A}_1$, in the latter that $\mathcal{A} = \mathcal{A}_0$. □

We conjecture that our technique from Proposition 3.6 can be used to characterize classes of models in the language of finitely many unary relation symbols, and also classes of models on which there is one equivalence relation. For the latter, we have the following partial result, completely characterizing the d.c.e.-learnable classes of equivalence structures of size 2, which suggests that even for equivalence structures, a full characterization is nontrivial:

Proposition 3.7. *Let $\mathfrak{K} = \{A_0, A_1\}$ be a class of two non-isomorphic equivalence structures on ω with equivalence relations E_0 and E_1, respectively. Then $\mathfrak{K}$ is d.c.e.-learnable iff neither of the following two conditions holds.*

1. *For $i \leq 1$, there is an injection $h : \omega/E_i \to \omega/E_{1-i}$ such that for all E_i-equivalence classes $[x]_{E_i}$, we have $|[x]_{E_i}| \leq |h([x]_{E_{1-i}})|$.*
2. *At least one of E_0 or E_1 has equivalence classes of unbounded finite size.*

Proof. First assume that (1) holds (by symmetry with $i = 0$), and, for the sake of a contradiction, assume there is d.c.e. learner L. We build an equivalence structure $\mathcal{A}$ on ω as follows: We start copying $\mathcal{A}_0$ into $\mathcal{A}$ until the learner guesses $\mathcal{A}_0$ at a stage s_0, say. Then, for each E_0-equivalence class C built or partially built by us for now, we copy C into $h(C)$ (which is possible by our condition on h) and start copying $\mathcal{A}_1$ into $\mathcal{A}$ consistent with what we have built so far. Now wait for the learner to guess $\mathcal{A}_1$. Then we let each element created after stage s_0 be in a new distinct equivalence class (using the fact that $\mathcal{A}$ is given only in a d.c.e. way) if E_0 has infinitely many equivalence classes, or into one infinite E_0-equivalence class (otherwise, since then E_0 has an infinite equivalence class) and again copy more of $\mathcal{A}_0$ into $\mathcal{A}$. We now repeat this process, changing strategy each time the learner seems to have a new correct guess.

There are now two possibilities: If the learner eventually fails to guess the structure we have started building from some point on, then we clearly win. On the other hand, if the learner changes his mind infinitely often, he loses by divergence but we end up building a copy of $\mathcal{A}_0$.

Now assume that (2) holds (by symmetry for E_0). The proof is essentially the same, except that when we switch from copying $\mathcal{A}_0$ into $\mathcal{A}$ to copying $\mathcal{A}_1$ into $\mathcal{A}$, we dynamically define a partial function h from the finitely many finite (parts of) E_0-equivalence classes already built to E_1-equivalence classes of larger size, which is possible by assumption.

Now suppose that both (1) and (2) fail; so both E_0 and E_1 can have at most finitely many infinite equivalence classes, and there is a fixed bound on the finite equivalence classes in each. So for the equivalence relation E_0, we now define a "reverse size sequence" $\{\kappa_j\}_{j<k}$ (for $k \leq \omega$) such that, listing the equivalence classes of E_0 is non-increasing size, the sequence lists the first k many. (Here, $k < \omega$ only if E_0 has only finitely many equivalence classes; if $k = \omega$, then $\kappa = \lim_j \kappa_j < \infty$

exists, and the κ_j need not exhaust all finite sizes of E_0-equivalence classes, namely, those of size $< \kappa$.) We define a similar sequence $\{\lambda_j\}_{j<l}$ for E_1, and we call these sequences "compatible" if for all $j < k$, $\kappa_j \leq \lambda_j$, or for all $j < l$, $\lambda_j \leq \kappa_j$. One can now easily check that if the sequences are compatible, then one can build a function h as in (1). So there are now two cases:

Case 1: $k = l$: Then there must $j < l$ with $\kappa_j < \lambda_j$, *and also* $j' < l$ with $\lambda_{j'} < \kappa_{j'}$. By symmetry, assume that $j < j'$. Fix $m \in \omega$ greater than the size of any finite E_0- or E_1-equivalence class and replace all values of ∞ in each sequence by m. Now we can define positive existential formulas saying "There are equivalence classes $C_0, C_1, \ldots, C_j$ of size at least κ_i for each $i \leq j$", and "There are equivalence classes $D_0, D_1, \ldots, D_{j'}$ of size at least λ_i for each $i \leq j'$", separating $\mathcal{A}_0$ and $\mathcal{A}_1$ as required for c.e.-learnability.

Case 2: $k \neq l$, and so by symmetry $k < l$: Then there must $j < k$ with $\kappa_j < \lambda_j$. Again, fix $m \in \omega$ greater than the size of any finite E_0- or E_1-equivalence class and replace all values of ∞ in each sequence by m. Then the learner, given an equivalence structure B, will simply look for the least parameters $\bar{x} \in \omega$ satisfying (1) $|\bar{x}| = (j+1) \cdot \kappa_j$ and $\bar{x}$ consists of $j+1$ many subsets of equivalence classes of size κ_j such that the subsets are pairwise inequivalent, or (2) $|\bar{x}| = k+1$ and all coordinates of $\bar{x}$ are pairwise inequivalent. At least one of these must eventually happen, and if both, the learner chooses the least such $\bar{x}$ and guesses $\mathcal{A}_0$ in case (1), and $\mathcal{A}_1$ in case (2).

□

4. n-LEARNING

We now take a closer look at mind changes made by the learner. Recall Definition 2.3: we consider n-learning, which is explanatory learning from informant, where we fix a bound n on how many times the learner changes its mind and outputs a new hypothesis. For learning of sets this convergence behavior of the learner was studied, e.g., in [1, 14, 21]. For classes of structures, some earlier results appeared in [24]. A descriptive set-theoretic interpretation of n-learning of structures recently appeared in [5], following the ideas from [6].

As in previous sections, we identify sequences $p \in 2^\omega$ with atomic diagrams of structures via a fixed sequence of sentences from Atm. Under this convention, we extend the definition of n-learning to arbitrary families of infinite binary strings to allow our proof of the syntactic characterization of n-learning to proceed smoothly by induction:

Definition 4.1. Fix a family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$), and let $P \subseteq 2^\omega$ be a family of strings coding atomic diagrams of copies of models in $\mathfrak{K}$. (Note that for technical reasons, we only require here that P codes models in $\mathfrak{K}$ but not that P codes all models of $\mathfrak{K}$, nor that P is closed under isomorphism.)

We then define P to be *n-learnable* if there is a learner $L : 2^{<\omega} \to \omega \cup \{?\}$ with the property that for any $p \in P$:

- $L(\lambda) = ?$ for the empty string λ;
- for any l, $L(p \upharpoonright l) = ?$ implies $L(p \upharpoonright l') = ?$ for any $l' < l$;
- there are at most n many $\sigma \subset p$ with $L(\sigma) \neq L(\sigma^-)$ and $L(\sigma^-) \neq ?$; and
- $L(\sigma) = i$ for the model $\mathcal{A}_i$ coded by p and all sufficiently long $\sigma \subset p$.

We note that if P is the set of all codes of atomic diagrams of a family $\mathfrak{K}$, then the above definition coincides with Definition 2.3 for finite n.

For our syntactic characterization of n-learning, we define special families of $\mathcal{L}$-sentences:

Definition 4.2. We define two classes of $\mathcal{L}$-sentences (allowing constants from ω for elements of $\mathcal{L}$-models):

- An S_0-*sentence* is an infinitary Σ_1-sentence in $\mathcal{L}$.
- An S_1-*sentence* is an (infinitary) $\mathcal{L}$-sentence of the form
$$\bigvee_l (\varphi_l \wedge \bigwedge_m \neg\chi_{l,m}),$$
where each φ_l, and each $\chi_{l,m}$, is an S_0-sentence.

We first make some comments that will simplify the proof of Theorem 4.3: First of all, let $\mathcal{L} = \bigcup_n \mathcal{L}_n$ be an increasing union of finite languages $\mathcal{L}_n$. (This will be irrelevant for finite $\mathcal{L}$ but makes the argument more uniform for infinite $\mathcal{L}$.) Next, we call a (finitary) existential sentence *complete* if it mentions precisely the variables $x_0, \ldots, x_n$ for some n, is an $\mathcal{L}_n$-sentence, and its matrix completely specifies the atomic $\mathcal{L}_n$-diagram of $x_0, \ldots, x_n$.

Now, in the definition of an S_0-sentence $\bigvee_l \varphi_l$, after possibly further expanding the disjunction, we may assume that each φ_l is a (finitary) complete existential sentence, and also (by dropping sentences with "longer" matrices) that the matrices of φ_l and φ_m (for $l \neq m$) are pairwise incompatible.

Similarly, in the definition of an S_1-sentence $\bigvee_l(\varphi_l \wedge \bigwedge_m \neg\chi_{l,m})$, we may assume that all φ_l are as above, and that, fixing l, the S_0-sentences $\chi_{l,m}$ also satisfy the above (fixing l and letting m vary), and furthermore, that the matrix of each $\chi_{l,m}$ properly extends the matrix of φ_l (since otherwise, if the matrices of φ_l and $\chi_{l,m}$ are contradictory, then the latter can be dropped, and if the matrix of $\chi_{l,m}$ is a subformula of the matrix of φ_l, then $\varphi_l \wedge \neg\chi_{l,m}$ would be contradictory).

We are now ready to state our first result.

Theorem 4.3.
- *A family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) is 0-learnable iff there is a sequence $\{\psi_i\}_{i\in\omega}$ of S_0-sentences such that $\mathcal{A}_i \models \psi_j$ iff $i = j$.*
- *If a family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) is n-learnable (for some $n > 0$) then there is a sequence $\{\psi_i\}_{i\in\omega}$ of S_1-sentences such that $\mathcal{A}_i \models \psi_j$ iff $i = j$.*

Proof. As alluded to above, we will actually, and tacitly, prove the result for all sets $P \subseteq 2^\omega$ coding atomic diagrams of models in $\mathfrak{K}$ in order for the inductive step to work correctly (i.e., not necessarily only for sets P coding all atomic diagrams of $\mathfrak{K}$).

For 0-learnability, suppose first that a family $\mathfrak{K}$ is 0-learnable. So fix a 0-learner $L : 2^{<\omega} \to \omega \cup \{?\}$. Then $L(\sigma) \neq ?$ and $\sigma \subset \tau$ implies $L(\sigma) = L(\tau)$. For each i, let M_i be the set of minimal σ with $L(\sigma) = i$, and let ψ_i be the (possibly infinite) disjunction of the atomic sentences coded by all $\sigma \in M_i$, written as existential sentences ψ_σ to replace the parameter constants for elements of ω. (Recall that we restrict ourselves here to complete S_0-sentences!) Then $M_i \cap M_j = \emptyset$ for $i \neq j$, and $\bigcup_i M_i$ forms an antichain. For the sake of a contradiction, suppose now that $\mathcal{A}_i \models \psi_j$ for $i \neq j$. So, in particular, $\mathcal{A}_i \models \psi_\sigma$ for some σ with $L(\sigma) = j$, which would contradict that L 0-learns $\mathcal{A}_i$.

Conversely, suppose there is such a sequence $\{\psi_i\}_{i\in\omega}$ of S_0-sentences. Then we can define a 0-learner for $\mathfrak{K}$ by setting $L(\sigma) = i$ for each disjunct ψ_σ (i.e., σ codes the corresponding finite fragment of an atomic diagram) of ψ_i. We then extend L to a total function by setting $L(\tau) = L(\sigma)$ if $L(\sigma)$ is already defined for some $\sigma \subset \tau$, and $L(\tau) = ?$ otherwise. The conditions on our sequence $\{\psi_i\}_{i\in\omega}$ now ensure that L is a 0-learner for $\mathfrak{K}$, since if σ and τ are comparable and ψ_σ and ψ_τ are defined, then they characterize the same model in $\mathfrak{K}$.

For n-learning for $n > 0$, suppose that $L : 2^{<\omega} \to \omega \cup \{?\}$ is an n-learner function for $\mathfrak{K}$ and the set P of paths coding the atomic diagram of a model in $\mathfrak{K}$. Let $M_i = \{\sigma \mid L(\sigma) = i \text{ and } L(\sigma^-) \neq i\}$. Now define ψ_i as the disjunction of all S_1-formulas of the form $\psi_\sigma \wedge \bigwedge_\tau \neg\psi_\tau$ where $\sigma \in M_i$ and the τ range over all minimal $\tau \supset \sigma$ with $L(\tau) \neq i$. First of all, clearly $\mathcal{A}_i \models \psi_i$ since for any path p coding $\mathcal{A}_i$, there is a longest $\sigma \subset p$ such that $L(\tau) = i$ for all τ with $\sigma \subseteq \tau \subset p$. Conversely, if $\mathcal{A}_i \models \psi_j$ and p is any path coding $\mathcal{A}_i$, then $L(\tau) = j$ for all sufficiently long $\tau \subset p$, so by our assumption on L, we must have $i = j$. □

In order to prove a full syntactic characterization of n-learning, we need a still finer classification of definability by S_1-sentences; it will no longer be enough to look at the complexity of the individual sentences ψ_i but to consider them all at once:

Definition 4.4. Given a sequence $\{\psi_i\}_{i\in\omega}$ of S_1-sentences of the form $\varphi_i \wedge \bigwedge_l \neg\chi_{i,l}$, corresponding to nodes σ_i (for φ_i) and $\tau_{i,l} \supset \sigma_i$ (for $\chi_{i,l}$, respectively), we define the *depth* of the sequence $\{\psi_i\}_{i\in\omega}$ as the supremum of the length n of a sequence $\rho_0 \subset \rho_1 \subset \cdots \subset \rho_n$, where the ρ_m are any of the nodes σ_i and $\tau_{i,l}$. (This supremum can be infinite; in the cases of interest to us, it will always be finite. Note that our conventions force, for $n = 0$, the sequence to only consist of S_0-sentences since all nodes ρ involved will then form an antichain.)

We are now in a position to state the complete syntactic characterization of n-learning:

Theorem 4.5. *For $n > 0$, a family $\mathfrak{K} = \{\mathcal{A}_i \mid i < k\}$ (for $k \leq \omega$) is n-learnable iff there is a sequence $\{\psi_i\}_{i\in\omega}$ of S_1-sentences of depth at most n such that $\mathcal{A}_i \models \psi_j$ iff $i = j$.*

Proof. The proof requires a more careful analysis of the above proof of Theorem 4.3.

Suppose first that $\mathfrak{K}$ is n-learnable. Define the formulas ψ_i as in the last paragraph of the proof of Theorem 4.3; the fact that L is an n-learner function then clearly implies that the depth of the sequence $\{\psi_i\}_{i\in\omega}$ is at most n.

Conversely, suppose that we have a sequence $\{\psi_i\}_{i\in\omega}$ of S_1-sentences of depth at most n; say, each ψ_i is of the form $\bigwedge_l(\varphi_{i,l} \wedge \bigwedge_m \neg\chi_{i,l,m})$, corresponding to nodes $\sigma_{i,l}$ (for $\varphi_{i,l}$) and $\tau_{i,l,m} \supset \sigma_{i,l}$ (for $\chi_{i,l,m}$, respectively). Without affecting the conditions on the sentences ψ_i, we can modify them as follows, by induction on the length of the nodes $\rho = \sigma_{i,l}$ or $\rho = \tau_{i,l,m}$: If we encounter a node $\sigma_{j,l'} \supset \sigma_{i,l}$ without a node $\tau_{i,l,m}$ with $\sigma_{i,l} \subset \tau_{i,l,m} \subset \sigma_{j,l'}$, then for any path $p \supset \sigma_{j,l'}$ corresponding to a model in $\mathfrak{K}$, we know that there must be some $\rho = \tau_{i,l,m}$ or some $\rho = \tau_{j,l',m'}$ with $\sigma_{j,l'} \subset \rho \subset p$. We now change ψ_j to, in place of $\sigma_{j,l'}$, refer to all such (minimal) $\tau_{i,l,m}$; this is fine since ψ_j can only start applying once such $\tau_{i,l,m}$ makes ψ_i no longer apply, and cannot apply at all as soon as we encounter some $\tau_{j,l',m'}$. On the other hand, if we encounter $\tau_{i,l,m}$ which is not of the form $\sigma_{j,l'}$ at the same time, then for any path

$p \supset \tau_{i,l,m}$ corresponding to a model in $\mathfrak{K}$, we know that there must be some $\sigma_{j,l'}$ with $\tau_{i,l,m} \subset \sigma_{j,l'} \subset p$. We now change ψ_i to, in place of $\tau_{i,l,m}$, refer to all such (minimal) $\sigma_{j,l'}$; this is fine since some ψ_j must apply once such $\tau_{i,l,m}$ makes ψ_i no longer apply.

We now define a learner function $L : 2^{<\omega} \to \omega$ as follows: Given $\sigma \in 2^{<\omega}$, find the longest $\rho \subseteq \sigma$ of the form $\sigma_{i,l}$. If there is none, then set $L(\sigma) = ?$. If ρ is of the form $\sigma_{i,l}$, then set $L(\sigma) = i$.

We first note that the assumption of the depth being at most n implies that L is an n-learner function. Now fix a path $p \in 2^\omega$ for the atomic diagram of a model $\mathcal{A} \in \mathfrak{K}$. Let $\rho \subset p$ be the longest string of the form $\sigma_{i,l}$. (Such ρ must exist by our assumption on the ψ_i and on the depth being finite.) Then $\mathcal{A} \models \varphi_{i,l}$ for this l but $\mathcal{A} \not\models \chi_{i,l,m}$ for any m, so $\mathcal{A} \models \psi_i$ and thus $\mathcal{A} \cong \mathcal{A}_i$. □

References

[1] Andris Ambainis, Sanjay Jain, and Arun Sharma. Ordinal mind change complexity of language identification. In Shai Ben-David, editor, *Computational Learning Theory*, pages 301–315, Berlin, Heidelberg, 1997. Springer Berlin Heidelberg.

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

[3] C. J. Ash and J. Knight. *Computable structures and the hyperarithmetical hierarchy*, volume 144 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, 2000.

[4] Chris Ash, Julia Knight, Mark Manasse, and Theodore Slaman. Generic copies of countable structures. *Ann. Pure Appl. Logic*, 42(3):195–205, 1989.

[5] Nikolay Bazhenov, Vittorio Cipriani, and Luca San Mauro. Calculating the mind change complexity of learning algebraic structures. In Ulrich Berger, Johanna N. Y. Franklin, Florin Manea, and Arno Pauly, editors, *Revolutions and Revelations in Computability*, pages 1–12, Cham, 2022. Springer International Publishing.

[6] Nikolay Bazhenov, Vittorio Cipriani, and Luca San Mauro. Learning algebraic structures with the help of Borel equivalence relations. *Theoretical Computer Science*, 951, 2023.

[7] Nikolay Bazhenov, Ekaterina Fokina, Dino Rossegger, Alexandra Soskova, and Stefan Vatev. Learning families of algebraic structures from text. In Ludovic Levy Patey, Elaine Pimentel, Lorenzo Galeotti, and Florin Manea, editors, *Twenty Years of Theoretical and Practical Synergies*, pages 166–178, Cham, 2024. Springer Nature Switzerland.

[8] Nikolay Bazhenov, Ekaterina Fokina, Dino Rossegger, Alexandra A. Soskova, and Stefan V. Vatev. A Lopez-Escobar theorem for continuous domains. *The Journal of Symbolic Logic*, page 1–18, 2024.

[9] Nikolay Bazhenov, Ekaterina Fokina, and Luca San Mauro. Learning families of algebraic structures from informant. *Information and Computation*, 275:104590, 2020.

[10] Nikolay Bazhenov and Luca San Mauro. On the Turing complexity of learning finite families of algebraic structures. *Journal of Logic and Computation*, 31(7):1891–1900, 2021.

[11] John Chisholm. Effective model theory vs. recursive model theory. *J. Symbolic Logic*, 55(3):1168–1191, 1990.

[12] Richard L. Epstein, Richard Haas, and Richard L. Kramer. Hierarchies of sets and degrees below $\mathbf{0}'$. In *Logic Year 1979–80 (Proc. Seminars and Conf. Math. Logic, Univ. Connecticut, Storrs, Conn., 1979/80)*, volume 859 of *Lecture Notes in Math.*, pages 32–48. Springer, Berlin, 1981.

[13] Ekaterina Fokina, Timo Kötzing, and Luca San Mauro. Limit learning equivalence structures. In Aurélien Garivier and Satyen Kale, editors, *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, volume 98 of *Proceedings of Machine Learning Research*, pages 383–403. PMLR, 22–24 Mar 2019.

[14] R. Freivalds and C.H. Smith. On the role of procrastination in machine learning. *Information and Computation*, 107(2):237–271, 1993.

[15] Ziyuan Gao, Frank Stephan, Guohua Wu, and Akihiro Yamamoto. Learning families of closed sets in matroids. In Michael J. Dinneen, Bakhadyr Khoussainov, and André Nies, editors,

Computation, Physics and Beyond - International Workshop on Theoretical Computer Science, WTCS 2012, volume 7160 of *LNCS*, pages 120–139, Berlin, 2012. Springer.
[16] Clark Glymour. *Inductive Inference in the Limit*, pages 23–31. Springer Netherlands, Dordrecht, 1985.
[17] E. Mark Gold. Language identification in the limit. *Inf. Control*, 10(5):447–474, 1967.
[18] Valentina S. Harizanov and Frank Stephan. On the learnability of vector spaces. *J. Comput. Syst. Sci.*, 73(1):109–122, 2007.
[19] Sanjay Jain and Efim Kinber. Learning languages from positive data and negative counterexamples. *Journal of Computer and System Sciences*, 74(4):431–456, 2008. Carl Smith Memorial Issue.
[20] Sanjay Jain, Daniel Osherson, James S. Royer, and Arun Sharma. *Systems that learn*. MIT Press, Cambridge, MA, 1999.
[21] Wei Luo and Oliver Schulte. Mind change efficient learning. *Information and Computation*, 204(6):989–1011, 2006.
[22] Eric Martin and Daniel Osherson. Scientific Discovery on Positive Data via Belief Revision. *Journal of Philosophical Logic*, 29(5):483–506, 2000.
[23] Eric Martin and Daniel N. Osherson. *Elements of scientific inquiry*. MIT Press, 1998.
[24] Eric Martin and Arun Sharma. On a syntactic characterization of classification with a mind change bound. In Peter Auer and Ron Meir, editors, *Learning Theory*, pages 413–428, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.
[25] Emil L. Post. Degrees of recursive unsolvability: preliminary report. *Bull. Amer. Math. Soc.*, 54:641–642, 1948.
[26] Hilary Putnam. Trial and error predicates and the solution to a problem of Mostowski. *J. Symb. Log.*, 30(1):49–57, 1965.
[27] Frank Stephan and Yuri Ventsov. Learning algebraic structures from text. *Theor. Comput. Sci.*, 268(2):221–273, 2001.

(Fokina) Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstrasse 8-10, 1040 Wien, AUSTRIA
Email address: ekaterina.fokina@tuwien.ac.at
URL: https://www.dmg.tuwien.ac.at/fokina/

(Lempp) Department of Mathematics, University of Wisconsin, Madison, Wisconsin 53706-1325, USA
Email address: lempp@math.wisc.edu
URL: http://www.math.wisc.edu/~lempp