Robust Learning of Automatic Classes of Languages*

Sanjay Jain**[1] and Eric Martin[2] and Frank Stephan***[3]

[1] School of Computing,
National University of Singapore,
Singapore 117417, Republic of Singapore.
Email: sanjay@comp.nus.edu.sg
[2] School of Computer Science and Engineering,
University of New South Wales,
Sydney 2052, Australia.
Email: emartin@cse.unsw.edu.au
[3] Department of Mathematics and School of Computing,
National University of Singapore,
Singapore 119076, Republic of Singapore.
Email: fstephan@comp.nus.edu.sg

Abstract. One of the most important paradigms in the inductive inference literature is that of robust learning. It has been extensively investigated in the context of learning recursive functions, where it had been originally defined. This paper adapts and investigates the paradigm of robust learning to learning languages from positive data. Broadening the scope of that paradigm is important: robustness captures a form of invariance of learnability under admissible transformations on the object of study; hence, it is a very desirable property. The key to defining robust learning of languages is to impose that the latter be automatic, that is, recognisable by a finite automaton. The invariance property used to capture robustness can then naturally be defined in terms of first-order definable operators, called translators; such operators are natural transformations which map automatic languages to automatic languages while preserving the inclusion relations from the class of original languages to the class of mapped languages. For each learning criterion amongst a selection of some of the most important learning criteria investigated either in the literature on explanatory learning from positive data or in the literature on query learning, we characterise the classes of languages all of whose translations are learnable under that criterion. Such learning criteria include consistent and conservative learning, strong-monotonic learning, strong-monotonic consistent learning, finite learning, learning from subset queries, learning from superset queries and learning from membership queries.

Keywords. Inductive inference, learning in the limit, query learning, robust learning, translations, automatic structures.

---

* A conference version of this paper appeared in the proceedings of "Algorithmic Learning Theory 2011" where the paper was presented.

** Supported in part by NUS grant numbers C252-000-087-001 and R252-000-420-112.

*** Supported in part by NUS grant numbers R146-000-114-112 and R252-000-420-112.

1 Introduction

The present paper considers robust learning in the framework of inductive inference, more precisely, of Gold-style language learning in the limit. Informally, Gold [5] formalised language learning in the limit in a way that the learner is presented with all members, one at a time, of a language selected from a class of languages to be learnt. From this data, the learner has to identify the language in the limit by conjecturing and revising at most finitely often a hypothesis where the last hypothesis describes the language to be learnt correctly. Learning is robust when it is preserved under any admissible transformation of a learnable class: that is, given a learnable class, each of the images of the class under an admissible transformation, is learnable. Of course, the notion of an "admissible transformation" has to be appropriately and naturally defined. A related question is that of which classes of languages could be the object of learning, as the proposed "admissible transformations" will be defined with respect to those classes. This is a familiar theme, as the search for invariants is prominent in many fields of mathematics. For example, Hermann Weyl described Felix Klein's famous Erlangen programme on the algebraic foundation of geometry in these words [22]: "If you are to find deep properties of some object, consider all natural transformations that preserve your object." In the field of inductive inference, Bārzdiņš addressed the question of robust learning in the context of learning classes of recursive functions, and he conjectured the following, see [4, 24]. Let a class of recursive functions be given. Then every image of the class under a general recursive operator is learnable iff the class is a subclass of a recursively enumerable (that is, a uniformly recursive) class of functions. To see where this conjecture comes from, it should be recalled that recursively enumerable classes of functions can be easily identified by a technique called "learning by enumeration" [5]. This technique amounts to simply conjecturing the first function in an effective list of the functions to learn, which is consistent with all data seen so far. The learnability of a class of functions by such an algorithm cannot be destroyed by transforming that class to another class using general recursive operators. So Bārzdiņš' conjecture essentially says that the enumeration technique fully captures robust learnability. Fulk [4] disproved the conjecture and this started a rich and fruitful exploration within the field of function learning [10, 11, 20]. Further refinements, such as uniform robust learnability [11] (where the learner for a transformed class has to be computable in a description of the transformation) and hyperrobust learnability [20] (learnability, by the same learner, of all transformations of a class under primitive recursive operators) have also been investigated.

It is natural to try and generalise robust learning to learning of classes of languages, first because the concept of robustness is an instance of the ubiquitous mathematical quest for invariants, and second because learning of classes of language was the first object of study in inductive inference and has been more broadly investigated than function learning. However, what seems to be the natural extension of the definition from the context of function learning to the context of language learning does not work well, as even the class of singletons would not be robustly learnable according to the resulting definition. This paper proposes a modified approach to robust language learning, focusing on specific classes of languages, to be introduced in the next paragraph. Not only are these classes of languages well suited to the definition of a natural transformation between languages that can adequately capture a notion of robust learning and enjoys appealing characterisations; these classes of languages are also interesting in their own right and are themselves a new important topic of research. Besides the advantages of the restriction to those interesting languages, all concepts defined in this paper are meaningful even with respect to all r.e. languages.

Before we introduce the specific classes of languages which we have identified as the natural object of study for robust learning of languages, recall that sets of finite strings over some finite alphabet are regular if they are recognisable by a finite state automaton. Sets of pairs of finite strings over respective alphabets are regular if they are recognisable by a finite state multi-input automaton that uses two different inputs to read both coordinates of the pair, with a special symbol (say $\star$) being used to pad a shorter coordinate. For instance, to accept the pair $(010, 45)$ an automaton should read 0 from the first input and 4 from the second input and change its state from the start state to some state q_1, then read 1 from the first input and 5 from the second input and change its state from q_1 to some state q_2, finally read 0 from the first input and $\star$ from the second input and change its state from q_2 to an accepting state. It is essential that all inputs involved are read synchronically — one character per input and cycle. One can similarly consider finite state automaton accepting triples, quadruples, and so on. The classes of languages that we focus on in this paper are classes of regular languages of the form $(L_i)_{i \in I}$ such that I and $\{(i, x) : x \in L_i\}$ are regular sets; we refer to such a class as an automatic family of languages. An automatic family of languages is actually a particular kind of automatic structure, an object of study in its own right, which is now a source of many interesting questions and results on definability [6, 14, 15].

What this paper presents is not the first work to create a bridge between inductive inference and automatic structures: learnability of automatic families has recently been studied [7, 8]. It should also be noted that our approach is an instance of a more general theme in inductive inference, that of the learnability of indexed families, a topic which has been extensively investigated in learning theory [1, 17, 18]: automatic families of languages are a special case of indexed families. One major advantage of automatic families over indexed families is that their first-order theory is decidable [6–8, 14] and many of their important properties are first-order definable. In particular, the inclusion structure of an automatic family can be first-order defined. As we will see, this property plays an important role in this paper, and it is a key reason why robust learning can be fruitfully studied with automatic families.

With the right classes of languages in hand, we can then suitably define the admissible transformations of one class of languages into another that will capture a natural form of robust learning. We consider any transformation given by an operator Φ which maps sets of strings to sets of strings such that the automatic family $(L_i)_{i \in I}$ to be learnt is mapped to a family $(L'_i)_{i \in I} = (\Phi\langle L_i \rangle)_{i \in I}$, where Φ is definable by a first-order formula, Φ preserves inclusions amongst sets of strings, and Φ preserves noninclusions between members of the family. We call such a Φ a translator. A key result of the theory of automatic structures is that the image $(\Phi\langle L_i \rangle)_{i \in I}$ of an automatic family under such an operator Φ is again an automatic family [14]. An important special case is given by continuous, or text-preserving, translators for which $\Phi\langle L \rangle$

is the union of all $\Phi\langle F\rangle$ where F ranges over the finite subsets of L. Continuity is one of the most important properties in the general theory of functionals, and this work is no exception; it captures the natural requirement of computing more and more of the elements of the mapped language from larger and larger, but always finite, sets of elements of the original language. We study the impact of such translations on learnability.

We proceed as follows. In Sections 2 and 3, we introduce the necessary notation and concepts. In Section 4, we provide an overview of the main results to guide the reader in what comes next. In Section 5, we illustrate the notions with a few examples and provide a general characterisation of robust learnability in the limit of automatic families of languages. In Section 6 to 8, we provide many further characterisations of robust learnability for some of the learning criteria that have been studied in the literature: consistent and conservative learning, strong-monotonic learning, strong-monotonic consistent learning, finite learning. In Section 11, we consider learning from subset queries, learning from superset queries and learning from membership queries.

The characterisations that have been found are all natural as they express a particular constraint on the inclusion structure of the original class. In many cases, they deal not only with transformations of the original class under all possible translations, but also with transformations under text-preserving (continuous) translations.

2 Automatic structures, languages and translations

The languages considered in inductive inference [9] consist of numbers implicitly coding some underlying structure, but the coding is not made explicit. In the context of the present work though, where languages have to be recognised by finite automata, a minimum of structure has to be given to the members of a language: they are assumed to be strings over an *alphabet* denoted by Σ. Let $\Sigma^\star$ denote the set of all finite strings over Σ. It is assumed that Σ is nonempty and finite. For $x \in \Sigma^\star$, the length of x, denoted $|x|$, is the number of symbols occurring in x; for example, $|00121| = 5$. We write xy for the concatenation of two strings x and y. We denote the empty string by ε.

We denote by I a regular subset of $\Sigma^\star$. We assume that Σ is strictly ordered and given $x, y \in \Sigma^\star$, we write $x <_{ll} y$ iff x is length-lexicographically smaller than y, that is, if either $|x| < |y|$ or $|x| = |y|$ and x comes lexicographically before y. We write $x \leq_{ll} y$ iff $x = y$ or $x <_{ll} y$.

In order to capture the constraint that a class of languages is uniformly recognisable by a finite automaton, we make use of a particular kind of automatic structures [15], that for simplicity, is still referred to as automatic structures. The structures under consideration offer enough expressive power to refer to the target language that a learner will be given a presentation of and has to eventually correctly identify, and to refer to the whole class of languages that are the object of learning. A unary predicate symbol and a binary predicate symbol are used to refer to the target language and the class of languages, respectively.

Definition 1. We call *automatic structure* any $\mathcal{V}$-structure $\mathfrak{M}$ whose domain is $\Sigma^\star$, with $\mathcal{V}$ being a relational vocabulary satisfying the following properties.

- $\mathcal{V}$ contains the unary predicate symbol X and the binary predicate symbol Y (and possibly more predicate symbols of any arity).
- The interpretation of X in $\mathfrak{M}$ is included in $\Sigma^\star$ and the interpretation of Y in $\mathfrak{M}$ is included in $I \times \Sigma^\star$, where I is a regular set.
- The interpretation of all predicate symbols in $\mathcal{V}$ in $\mathfrak{M}$ is regular.

By *language* we mean a subset of $\Sigma^\star$. Intuitively, in the above definition, I is a set of indices. X is a predicate for a language $\{x : X(x) = 1\}$, and Y is a predicate for describing a class of languages $\mathbf{I} = (L_i)_{i \in I}$, where $L_i = \{x : Y(i, x) = 1\}$, for $i \in I$.

Definition 2. Let I be a regular set. An *automatic class* is a repetition-free I-family $\mathbf{I} = (L_i)_{i \in I}$ of languages such that $\{(i, x) : i \in I,\ x \in L_i\}$ is recognisable by a finite state multi-input automaton. Members of I are referred to as *indices* for the languages in the class $\mathbf{I}$.

Assuming that Σ has at least 2 elements, say 0 and 1, here are some examples of classes of languages that can be represented as automatic classes, for proper choices of I:

- the class of sets with up to k elements for a constant k;
- the class of all finite and cofinite subsets of $\{0\}^\star$;
- the class of all intervals of an automatic linear order on a regular set.

On the other hand, the class of all finite sets over $\{0, 1\}$ is not automatic.

The constraint that automatic classes be repetition-free is not standard when one considers learning of indexed families. However, it is at no loss of generality in the context of the present work and allows one to substantially simplify the arguments in most proofs.

One advantage of considering automatic structures and families is that first-order definable relations over existing automatic relations are also automatic. Thus, several problems related to automatic families become decidable.

Fact 3 (Khoussainov, Nerode [14]). Any relation that is first-order definable from existing automatic relations is automatic.

We consider transformations of languages that are definable in the language and are used to describe the target language and the class of languages to be learnt. Intuitively, in the next definition, Φ is a translator (using an automatic class $\mathbf{I}$ as a parameter), that maps a language L to $\Phi_{\mathbf{I}}\langle L \rangle$.

Definition 4. Let Φ be any first-order formula over the vocabulary of some automatic structure with the distinguished variable x as unique free variable (this allows one to denote such a formula by Φ rather than by $\Phi(x)$). Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given. For all languages L, denote by $\Phi_{\mathbf{I}}\langle L \rangle$ the language consisting of all strings s such that $\Phi[s/x]$ is true in all automatic structures in which the interpretation of $X(w)$ is $w \in L$ and the interpretation of $Y(i, w)$ is $i \in I \wedge w \in L_i$. We say that Φ is *an automatic* $\mathbf{I}$*-translator* if:

- for all languages L and L', if $L \subseteq L'$ then $\Phi_{\mathbf{I}}\langle L \rangle \subseteq \Phi_{\mathbf{I}}\langle L' \rangle$;

– for all members i and j of I, if $L_i \nsubseteq L_j$ then $\Phi_{\mathbf{I}}\langle L_i\rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_j\rangle$.

As an example consider the translation Φ given by the formula

$$x \in I \wedge \forall y\,[[Y(x,y) = 1] \Rightarrow [X(y) = 1]]$$

Then, for $\mathbf{I} = (L_i)_{i\in I}$, $\Phi_{\mathbf{I}}(L)$ maps L to the language $\{x \in I : L_x \subseteq L\}$.
Let an automatic class $\mathbf{I} = (L_i)_{i\in I}$ be given. Note that though the index set I is not part of the logical vocabulary, "$x \in I$" is first-order expressible in this language as there is at most one index for $\varnothing$: I is either the set of all xs with $\exists y\, Y(x,y)$, or the set of all xs with $\exists y\, Y(x,y) \vee x = i_0$, where i_0 denotes the index of $\varnothing$ in case $\varnothing \in \mathbf{I}$.

For ease of notation, given two terms t and t', we write $t \in X$ for $X(t)$ and $t' \in Y_t$ for $Y(t,t')$ (equivalently, $Y_t = \{t' : Y(t,t')\}$).

Given an automatic $\mathbf{I}$-translator Φ, we let $\Phi\langle\mathbf{I}\rangle$ denote $(\Phi_{\mathbf{I}}\langle L_i\rangle)_{i\in I}$; we refer to any such family as a *translation of* $\mathbf{I}$. Note that a translation is always defined with respect to $\mathbf{I}$-translators for a particular automatic class $\mathbf{I}$. We drop the reference to $\mathbf{I}$ for ease of notation.

One major advantage of the definability of translators via first-order formulas is that automaticity is preserved. This follows from Fact 3.

Theorem 5. *For all automatic classes* $\mathbf{I}$*, all translations of* $\mathbf{I}$ *are automatic.*

3 Texts and learnability

Let us recall the basic concepts in inductive inference as originally defined in [5] and fix some notation. The only difference with the classical framework of learning from positive data is that we consider languages over strings rather than natural numbers.

Let # be a special symbol not in Σ. We denote by SEQ the set of finite sequences of members of $\Sigma^\star \cup \{\#\}$. Given $\sigma \in$ SEQ, we denote by $\mathrm{rng}(\sigma)$ the set of members of $\Sigma^\star$ that occur in σ (for example, if $\Sigma = \{a,b,c\}$ and $\sigma = (ab, \#, ab, cacba)$ then $\mathrm{rng}(\sigma) = \{ab, cacba\}$). We say that σ is *for* L if $\mathrm{rng}(\sigma) \subseteq L$. Given $\sigma \in$ SEQ and a family $\mathbf{I} = (L_i)_{i\in I}$ of languages, we say that σ is *for* $\mathbf{I}$ iff $\mathrm{rng}(\sigma) \subseteq L_i$ for some $i \in I$. Given a language L, a *text for* L refers to an enumeration of all elements of L of the form $(e_k)_{k\in\mathbb{N}}$, possibly with duplicates and possibly with #, but no other symbol, occurring anywhere in the enumeration. In particular, if $L = \varnothing$ then $e_k = \#$ for all $k \in \mathbb{N}$. The concatenation of $\sigma \in$ SEQ and $\tau \in$ SEQ is denoted by $\sigma \diamond \tau$. For $s \in \Sigma^\star \cup \{\#\}$ and $\sigma \in$ SEQ, we sometimes abuse notation and use $\sigma \diamond s$ to denote the concatenation of σ with (s). A member τ of SEQ is an *initial segment* of another member σ of SEQ iff $\sigma = \tau \diamond \tau'$ for some $\tau' \in$ SEQ; in this case σ is said to *extend* τ.

The notion of translator is quite general and it is worthwhile to examine to which extent it can be constrained to continuous transformations, that is, translators such that any member of the translation can be determined from a finite subset of the original language:

Definition 6. Let an automatic class $\mathbf{I} = (L_i)_{i\in I}$ and an automatic $\mathbf{I}$-translator Φ be given. We say that Φ is *text-preserving* iff for all languages L and for all $s \in \Phi_{\mathbf{I}}\langle L\rangle$, there is a finite subset F of L with $s \in \Phi_{\mathbf{I}}\langle F\rangle$.

We talk about *text-preserving translation of* $\mathbf{I}$ to refer to any family of the form $\Phi\langle\mathbf{I}\rangle$ where Φ is a text-preserving automatic $\mathbf{I}$-translator.

Example 7. Given an automatic class $\mathbf{I} = (L_i)_{i\in I}$, let a formula Φ^{nc} (with x as unique free variable, parameter X for the input language, and parameter Y_i for the i-th element L_i of the indexing) express $x \in I \wedge \exists y(y \in X \setminus Y_x)$, that is, $x \in I \wedge X \nsubseteq Y_x$. Then for all languages L, $\Phi^{nc}_{\mathbf{I}}\langle L\rangle$ is equal to $\{i \in I : L \nsubseteq L_i\}$. Moreover, Φ^{nc} is text-preserving.

Almost all results will involve recursive learners, with one exception (Theorem 30) where we had to drop the recursiveness requirement. This result will be expressed in terms of general learners. Learners of both kinds are defined next.

Definition 8. A *general learner* is a partial function from SEQ into I. A *learner* is any partial recursive function from SEQ into I with a recursive domain.

In the context of automatic structures, the fact that a learner is undefined on some input indicates that the learner cannot make a reasonable guess, rather than the learner being unable to make a guess due to computational infeasibility. This justifies letting learners be partial rather than total functions. We could also let a learner output some special symbol rather than being undefined.

Definition 9 (Gold [5]). Let $\mathbf{I} = (L_i)_{i\in I}$ be an automatic class. A learner M is said to *learn* $\mathbf{I}$ iff for all $i \in I$ and for all texts $(e_k)_{k\in\mathbb{N}}$ for L_i, $M\big((e_0 \dots, e_k)\big)$ is defined and equal to i for cofinitely many $k \in \mathbb{N}$. We say that $\mathbf{I}$ is *learnable* iff some learner learns $\mathbf{I}$.

Note that for simplicity, we use the term "learning" to refer to the notion that in the literature is more precisely called *explanatory learning*. Furthermore, observe that the definition above takes advantage of the one-one indexing of the automatic families considered.

We now recall some of the restrictions on learnability that have been investigated in the literature [1, 3, 12, 13, 23] and that will be considered in this paper, individually or combined.

Definition 10. Let $\mathbf{I} = (L_i)_{i\in I}$ be an automatic class and M be a learner that learns $\mathbf{I}$.

M is *consistent* iff for all $\sigma \in$ SEQ, if σ is for $\mathbf{I}$ then $M(\sigma)$ is defined and $\mathrm{rng}(\sigma) \subseteq L_{M(\sigma)}$.

M is *conservative* iff for all $\sigma, \tau \in$ SEQ, if $\sigma \diamond \tau$ is for $\mathbf{I}$, both $M(\sigma)$ and $M(\sigma\diamond\tau)$ are defined and $L_{M(\sigma\diamond\tau)} \neq L_{M(\sigma)}$, then $\mathrm{rng}(\sigma\diamond\tau) \setminus L_{M(\sigma)} \neq \varnothing$.

M is *confident* iff for all texts e for an arbitrary language, there exists $m \in \mathbb{N}$ such that for all $n \geq m$, $M\big((e(0)\dots e(n))\big)$ is undefined or equal to $M\big((e(0)\dots e(m))\big)$.

M is *strong-monotonic* iff for all $\sigma, \tau \in$ SEQ, if σ is an initial segment of τ, τ is for $\mathbf{I}$ and both $M(\sigma)$ and $M(\tau)$ are defined, then $L_{M(\sigma)} \subseteq L_{M(\tau)}$.

Conservative and strong-monotonic learners do not *overgeneralise*, that is, on any input sequence for $L \in \mathbf{I}$, they do not output a conjecture which is a proper superset of L.

Definition 11. An automatic class $\mathbf{I}$ is said to be *consistently, conservatively, confidently or strong-monotonically learnable* iff some consistent, conservative, confident or strong-monotonic learner learns $\mathbf{I}$, respectively.

For robust learning, one requires that each translation $\Phi\langle\mathbf{I}\rangle$ of the family $\mathbf{I}$ is learnable (according to the given criterion), where Φ ranges over all automatic $\mathbf{I}$-translators. Note that requiring the learnability of each translation demands that $\mathbf{I}$ itself be learnable, as the identity is a particular translator. In some cases, we consider learnability of $\Phi\langle\mathbf{I}\rangle$ only for all *text-preserving* Φs.

The characterisation of learnability of indexed families of languages in terms of tell-tales given by Angluin [1] can easily be adapted to the current setting, with indexed families replaced by automatic classes. The characterisation is simpler here because the tell-tales do not have to be assumed to be computable from the languages in the class, as they are necessarily so.

Definition 12 (Angluin [1]). Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class. Given $i \in I$, a *tell-tale for* L_i (with respect to $\mathbf{I}$) is a finite $F \subseteq L_i$ such that for all $i' \in I$, if $F \subseteq L_{i'} \subseteq L_i$ then $L_i = L_{i'}$.

If $\mathbf{I}$ is clear from context, then, for ease of notation, we often drop "(with respect to $\mathbf{I}$)," when considering tell-tale sets. If every language in an automatic class $\mathbf{I}$ has a tell-tale set, then we say that the class satisfies Angluin's tell-tale condition.

Theorem 13 (Jain, Luo and Stephan [7]; based on Angluin [1]). *Let* $\mathbf{I} = (L_i)_{i \in I}$ *be an automatic class. Then* $\mathbf{I}$ *is learnable iff for all* $i \in I$*, there exists a tell-tale for* L_i*. Moreover, if* $\mathbf{I}$ *is learnable then* $\mathbf{I}$ *is consistently and conservatively learnable by a set-driven learner* (*whose conjecture on an input* σ *only depends on* $\mathrm{rng}(\sigma)$).

Alternatively, one could also describe the tell-tale by an upper bound in order to get a first-order formula which expresses learnability. An automatic class $\mathbf{I} = (L_i)_{i \in I}$ is learnable iff for all members i of I, there is a bound $b_i \in \Sigma^\star$ such that $\{y \in L_i : y \leq_{ll} b_i\} \subseteq L_j \subset L_i$ for no $j \in I$. This is equivalent to

$$(\forall i \in I)\,(\exists b_i \in \Sigma^\star)\,(\forall j \in I)\big[\exists y \in L_i \setminus L_j(y \leq_{ll} b_i)\ \vee\ \exists y \in L_j \setminus L_i\ \vee\ \forall y \in L_i(y \in L_j)\big].$$

In order not to clutter notation, we will from now on abstain from breaking subset-relations down into first-order formulas as exemplified with the previous formula; we leave it to the reader to formalise subset-relations via quantified predicates using membership.

Example 14. Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given. There are two learners, M_{smon} and M_{ex} (that use an automatic description of $\mathbf{I}$ as a parameter), which learn $\mathbf{I}$ whenever $\mathbf{I}$ is strong-monotonically and explanatorily learnable, respectively. These two learners are defined as follows.

- In response to $\sigma \in \mathrm{SEQ}$, M_{smon} outputs the unique $i \in I$ that satisfies (1) $\mathrm{rng}(\sigma) \subseteq L_i$ and (2) for all $j \in I$, if $\mathrm{rng}(\sigma) \subseteq L_j$ then $L_i \subseteq L_j$; if such an i does not exist then M_{smon} is undefined.
- In response to $\sigma \in \mathrm{SEQ}$, M_{ex} outputs the unique $i \in I$ such that $\mathrm{rng}(\sigma) \subseteq L_i$ and there is no $j \in I$ with (1) $\mathrm{rng}(\sigma) \subseteq L_j$ and (2) either $L_j \subset L_i$ or $j <_{ll} i$ and $L_i \not\subset L_j$; if such an index i does not exist then M_{ex} is undefined.

Proof. Clearly, the learner M_{smon} is partial recursive and has a recursive domain. Note that by definition, any hypothesis output by M_{smon} is the smallest one (with respect to $\subseteq$) which contains the input data seen so far; hence, any further hypothesis output by M_{smon} is a superset of the current one and therefore M_{smon} is strong-monotonic. Suppose that $\mathbf{I}$ is learnable by a (possibly non partial recursive) strong-monotonic learner N. Let us show that for all σ for $\mathbf{I}$, if $N(\sigma) = i$ then either $\mathrm{rng}(\sigma) \nsubseteq L_i$ or $M_{smon} = i$; this implies that M_{smon} also learns $\mathbf{I}$ as whenever N converges to i on a text for L_i, so does M_{smon}. So assume that $N(\sigma)$ outputs i and L_i contains $\mathrm{rng}(\sigma)$. Let j be any index with $\mathrm{rng}(\sigma) \subseteq L_j$. There is a text for L_j which starts with σ; hence, N outputs j on some τ extending σ. It follows from the strong-monotonicity of N that $L_i \subseteq L_j$. In other words, i is the index of the $\subseteq$-minimal language containing $\mathrm{rng}(\sigma)$; hence, $M_{smon}(\sigma) = i$. Therefore M_{smon} is a strong-monotonic learner for the class $\mathbf{I}$ whenever $\mathbf{I}$ has a strong-monotonic learner at all.

Clearly, the learner M_{ex} is partial recursive and has a recursive domain. Now assume that the class $\mathbf{I}$ is explanatorily learnable. Therefore it satisfies Angluin's tell-tale condition. Fix $i \in L_i$ and a text for L_i; to complete the verification of the claim of the example, it is necessary and sufficient to show that M_{ex} converges on this text to i. For every sufficiently long initial segment σ of the given text for L_i, it holds that (a) $\mathrm{rng}(\sigma)$ contains the tell-tale of L_i and (b) $\mathrm{rng}(\sigma)$ contains some element of $L_i \setminus L_j$ for every $j <_{ll} i$ with $L_i \nsubseteq L_j$. Condition (a) implies that there is no j with $\mathrm{rng}(\sigma) \subseteq L_j \subset L_i$, and condition (b) implies that there is no $j <_{ll} i$ with $\mathrm{rng}(\sigma) \subseteq L_j$ and $L_i \not\subset L_j$. Hence, for every $i \in I$ and every text for L_i, $M_{ex}(\sigma)$ conjectures i on almost all initial segments of the text. $\square$

Note that not all explanatorily learnable classes are strong-monotonically learnable. Hence, the learner M_{smon} is not as powerful as M_{ex}. An example of a class which is explanatorily learnable but not strong-monotonically learnable is $\big\{\{0,1\}^\star \setminus \{x\} : x \in \{0,1\}^\star\big\}$. Furthermore, the above learners can of course also operate on classes of the form $\Phi\langle\mathbf{I}\rangle$ (using $\Phi\langle\mathbf{I}\rangle$ as a parameter instead of $\mathbf{I}$) whenever such a class is learnable under the corresponding condition.

4 Overview of the main results

Theorem 13, proved in [7], characterises the learnability of an automatic class in terms of tell-tales, defined in Definition 12. We will establish other characterisations of the learnability of an automatic class or of one of its translations:

- Theorem 21 shows that every automatic class has some strong-monotonically learnable translation.
- Theorem 31 characterises the finite learnability of some translation of an automatic class in terms of the class being an antichain.
- Theorem 34 shows that every automatic class is learnable from equivalence queries.
- Theorem 39 shows that every automatic class has a translation learnable using membership queries.
- Theorem 40 and Corollary 41 characterise the learnability of an automatic class from subset and superset queries, respectively, in terms of tell-tale-like conditions.

Another family of results characterise the learnability of all translations of an automatic class in terms of tell-tale-like conditions.

- Theorem 17, dealing with arbitrary automatic classes, is the most general result, and it also holds for text-preserving translations.
- Theorem 20 provides such a characterisation for strong-monotonic learnability, which also holds for text-preserving translations.
- Theorem 26 provides such a characterisation for strong-monotonic and confident learnability.
- Theorem 46 provides such a characterisation for learnability from membership queries.

Besides, characterising the learnability of all translations of an automatic class in terms of tell-tale-like conditions, we also do some other characterisations as follows:

- In terms of well orderings of the set of indices, or of the class of languages under inclusion:
  - Theorem 19 provides such a characterisation for consistent and conservative learnability.
  - Theorem 24 provides such a characterisation for strong-monotonic consistent learnability, which also holds for text-preserving translations.
- In terms of the finiteness of the class, with Theorem 29 providing such a characterisation for confident, conservative and consistent learnability.
- In terms of the existence of least upper bounds whenever every finite collection of languages in the class are bounded (with respect to inclusion) by a language in the class:
  - Theorem 25 provides such a characterisation when some translation of the class is consistently and strong-monotonically learnable.
  - Corollary 27 provides such a characterisation when some translation of the class is consistently, confidently and strong-monotonically learnable.

5 General characterisation

We start with two examples of conditions that guarantee robust learnability.

Theorem 15. *Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class.*

- *If $\big(\{L_i : i \in I\}, \supset\big)$ is well ordered (for the superset, not the subset relation) then all translations of $\mathbf{I}$ are learnable.*
- *If for all $i, j \in I$, $L_i \subseteq L_j \Leftrightarrow L_i = L_j$, then all translations of $\mathbf{I}$ are learnable.*

Proof. Let an automatic $\mathbf{I}$-translator Φ be given.

Suppose that $\big(\{L_i : i \in I\}, \supset\big)$ is well ordered. Let ordinal κ and $(L'_\lambda)_{\lambda<\kappa}$ be a well ordering of $\big(\{L_i : i \in I\}, \supset\big)$. Then $(\Phi_{\mathbf{I}}\langle L'_\lambda\rangle)_{\lambda<\kappa}$ is a well ordering of $\big(\{\Phi_{\mathbf{I}}\langle L_i\rangle : i \in I\}, \supset\big)$. Given $i \in I$, let $\lambda < \kappa$ be such that $L_i = L'_\lambda$, and let $s_i \in \Phi_{\mathbf{I}}\langle L'_\lambda\rangle$ be such that if $\lambda + 1 < \kappa$ then $s_i \notin \Phi_{\mathbf{I}}\langle L'_{\lambda+1}\rangle$. Clearly, for all $i \in \mathbb{N}$, $\{s_i\}$ is a tell-tale for $\Phi_{\mathbf{I}}\langle L_i\rangle$. We conclude using Theorems 5 and 13 that $\Phi\langle\mathbf{I}\rangle$ is learnable.

Suppose that for all members i and j of I, $L_i \subseteq L_j$ and $L_i = L_j$ are equivalent. This property is inherited by the translations: $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$ iff $\Phi_{\mathbf{I}}\langle L_i\rangle = \Phi_{\mathbf{I}}\langle L_j\rangle$. Then for all $i \in \mathbb{N}$, $\varnothing$ is a tell-tale for $\Phi_{\mathbf{I}}\langle L_i\rangle$, and we conclude using Theorems 5 and 13 again that $\Phi\langle\mathbf{I}\rangle$ is learnable. $\square$

As can be expected, learning does not imply robust learning, even if restricted to text-preserving translations:

Theorem 16. *There exists a strong-monotonically, consistently and confidently learnable automatic class one of whose text-preserving translations is not learnable.*

Proof. Take I equal to $\{0,1\}^\star$. Clearly, there exists an automatic class $\mathbf{I} = (L_i)_{i \in I}$ such that $L_\varepsilon = \{0,1\}^\star$ (recall that ε denotes the empty string), $L_0 = \emptyset$, and for all $i \in I \setminus \{\varepsilon, 0\}$, $L_i = \{i\}$. Define Φ as $\exists v(v \in X \wedge v \neq x)$ (a formula with x as unique free variable). It is immediately verified that $\mathbf{I}$ is strong-monotonically, consistently and confidently learnable. Furthermore, Φ is an automatic $\mathbf{I}$-translator, and $\Phi\langle\mathbf{I}\rangle$ consists of I, $\emptyset$, and all cosingletons of I except for $I \setminus \{\varepsilon\}$ and $I \setminus \{0\}$. This implies that $\Phi\langle\mathbf{I}\rangle$ is not learnable (as there is no tell-tale of I with respect to $\Phi\langle\mathbf{I}\rangle$) [1]. $\square$

Theorem 17 offers a general characterisation of robust learning in this framework along the lines of Theorem 13. It is worth comparing the third condition in Theorem 17 with the condition on tell tales spelled out after Theorem 13. Given $i \in I$, the existence of a language L_j such that $\{y \in L_i : y \leq_{ll} b_i\} \subseteq L_j \subset L_i$ is not ruled out but still under control: L_i rather than L_j will be output as a hypothesis on the basis of $\{y \in L_i : y \leq_{ll} b_i\}$ only if L_j has been "killed" thanks to the appearance of a member of L_i that does not belong to a superset L_k of L_j.

Theorem 17. *Given an automatic class $\mathbf{I} = (L_i)_{i \in I}$, the three conditions below are equivalent.*

1. *Every translation of $\mathbf{I}$ is learnable.*
2. *Every text-preserving translation of $\mathbf{I}$ is learnable.*
3. *For all $i \in I$, there exists $b_i \in I$ such that for all $j \in I$, either $L_j \not\subset L_i$ or there exists $k \in I$ with $k \leq_{ll} b_i$, $L_i \nsubseteq L_k$ and $L_j \subseteq L_k$.*

Proof. It suffices to prove that 3. implies 1. and 2. implies 3.

We first show that 3. implies 1. Assume that 3. holds. Without loss of generality, for all $i \in I$, let b_i be $\leq_{ll}$-least member of I that satisfies the third condition of the theorem. Note that for all $i \in I$, b_i is first-order definable from i; therefore the mapping $i \mapsto b_i$ is recursive (Fact 3). Let Φ be an automatic $\mathbf{I}$-translator. Let a learner M be such that in response to $\sigma \in \mathrm{SEQ}$, M outputs the $\leq_{ll}$-least $i \in I$, if any, such that $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle$ and for all $j \in I$, if $L_j \subset L_i$ then $\mathrm{rng}(\sigma)$ contains a member of $\Phi_{\mathbf{I}}\langle L_i\rangle \setminus \Phi_{\mathbf{I}}\langle L_k\rangle$ for some $k \leq_{ll} b_i$ with $L_j \subseteq L_k$. It is easily verified that M learns $\Phi\langle\mathbf{I}\rangle$.

We now show that 2. implies 3. For a contradiction, assume that there exists $i \in I$ for which there exists no $b_i \in I$ such that for all $j \in I$, either $L_j \not\subset L_i$ or there exists $k \in I$ with $k \leq_{ll} b_i$, $L_j \subseteq L_k$ and $L_i \nsubseteq L_k$. Thus, I is infinite. We have to exhibit a text-preserving automatic $\mathbf{I}$-translator Φ such that $\Phi\langle\mathbf{I}\rangle$ is not learnable. Define Φ as follows: given a language L, $\Phi_{\mathbf{I}}\langle L\rangle$ consists of all $(p,n) \in I^2$ such that at least one of the following conditions holds:

(a) For all s with $s \leq_{ll} n$, if $s \in L_p$ then $s \in L$.
(b) For all s with $s \leq_{ll} n$, if $s \in L_p$ then $s \in L_i$. Furthermore, for all $k \in I$ with $k \leq_{ll} \max(p,n)$, either $L_i \subseteq L_k$ or $L \nsubseteq L_k$.

This is a first-order definition. Let $H_j = \Phi_{\mathbf{I}}\langle L_j \rangle$. It follows from the definition of Φ that if $L \subseteq L'$ then $\Phi_{\mathbf{I}}\langle L \rangle \subseteq \Phi_{\mathbf{I}}\langle L' \rangle$. Let $j, j', n \in I$ be such that there exists $s \leq_{ll} n$ with $s \in L_j \setminus L_{j'}$. Then at least one of the following conditions holds:

- For all s' with $s' \leq_{ll} n$, if $s' \in L_i$ then $s' \in L_{j'}$. Hence, $s \in L_j \setminus (L_i \cup L_{j'})$ and $s \leq n$; therefore (j, n) belongs to $H_j \setminus H_{j'}$.
- There exists s' such that $s' \leq_{ll} n$, $s' \in L_i$ and $s' \notin L_{j'}$. Let $j'' = \max(j', n)$. Clearly, $(j, j'') \in H_j$. Furthermore, setting $p = j$ and $n = j''$ in conditions (a) and (b) above and instantiating k to j' in the last part of condition (b), it follows from the existence of s' and the inclusion $L_{j'} \subseteq L_{j'}$ that (j, j'') cannot be in $H_{j'} = \Phi_{\mathbf{I}}\langle L_{j'} \rangle$. Hence, (j, j'') belongs to $H_j \setminus H_{j'}$.

Hence, by case distinction, $H_j \not\subseteq H_{j'}$.

We conclude that Φ is an automatic $\mathbf{I}$-translator. Moreover, it is immediately verified that Φ is text-preserving. Now for every $b_i \in I$, there is a $j \in I$ with $L_j \subset L_i$ such that for all $k \in I$ with $k \leq_{ll} b_i$, either $L_i \subseteq L_k$ or $L_j \not\subseteq L_k$. It follows that for all elements (p, n) of H_i, if both p and n are $\leq_{ll}$-smaller than b_i then (p, n) is also in H_j. However, H_j is still a proper subset of H_i. It follows that H_i does not have a finite tell-tale. Hence, by Theorem 13, $\Phi\langle \mathbf{I} \rangle$ is not learnable. □

6 Characterisations of learnability variously constrained

Consistency is a rather weak constraint on learners, and is often combined with other desirable properties. We first combine consistency with conservativeness. In Section 8, we will combine it with strong-monotonicity. Note that here, "a class is consistently and conservatively learnable" means that the class is learnable by a learner which is both consistent and conservative (rather than having two different learners, one satisfying consistency and the other satisfying conservativeness). A similar convention applies to combining other constraints on learners. Let us first illustrate the notion with an example.

Example 18. Take I equal to $\{1^n, 2^n : n \in \{1, 2, 3, \ldots\}\}$. Let $\mathbf{I} = (L_i)_{i \in I}$ be defined by $L_{1^n} = \{0^m : m > n\}$ and $L_{2^n} = \{0^m : m < n\}$ for all $n > 0$. Note that $\mathbf{I}$ is an automatic class that is neither $\subset$- nor $\supset$-well founded. Let Φ be a text-preserving automatic $\mathbf{I}$-translator. Some consistent and conservative learner M learns $\Phi\langle \mathbf{I} \rangle$, proceeding as follows in response to $\sigma \in \mathrm{SEQ}$:

If σ extends τ with $M(\tau)$ being defined and $\mathrm{rng}(\sigma) \subseteq L_{M(\tau)}$, then M outputs $M(\tau)$,
else if there is $n \in \mathbb{N}$ with $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_{1^n} \rangle$ and $\mathrm{rng}(\sigma) \not\subseteq \Phi_{\mathbf{I}}\langle L_{1^{n+1}} \rangle$, then M outputs 1^n,
else M conjectures 2^n for the least $n > 0$ with $\mathrm{rng}(\sigma)$ included in $\Phi_{\mathbf{I}}\langle L_{2^n} \rangle$.

Since Φ is text-preserving, for all $n > 0$, every finite subset of $\Phi_{\mathbf{I}}\langle L_{1^n} \rangle$ is contained in $\Phi_{\mathbf{I}}\langle L_{2^m} \rangle$ for some $m \in \mathbb{N}$; hence, M is consistent. By the first clause in the definition of M, M is conservative.

We now show that M learns $\Phi\langle \mathbf{I} \rangle$. Let $n > 0$ be given. Presented with a text for $\Phi_{\mathbf{I}}\langle L_{1^n} \rangle$, M eventually observes a datum outside $\Phi_{\mathbf{I}}\langle L_{1^{n+1}} \rangle$, at which point M either conjectures 1^n or outputs the previous hypothesis — of the form 2^m for some $m > 0$ — until $\Phi_{\mathbf{I}}\langle L_{2^m} \rangle$ becomes inconsistent with the data observed, at which point M makes a mind change to 1^n. Presented with a text for $\Phi_{\mathbf{I}}\langle L_{2^n} \rangle$, M eventually conjectures 2^n as soon as the data observed become inconsistent with

$\Phi_{\mathbf{I}}\langle L_1\rangle$ and $\Phi_{\mathbf{I}}\langle L_{2^m}\rangle$ for all nonzero $m < n$, which is guaranteed to happen as there are only finitely many languages of the latter type.

Combined with Theorem 17, the next result characterises robust learnability by consistent, conservative learners.

Theorem 19. *Let* $\mathbf{I}$ *be a learnable automatic class all of whose translations are learnable. Then every translation of* $\mathbf{I}$ *is consistently and conservatively learnable iff the set of members of* $\mathbf{I}$ *is well founded under inclusion.*

Proof. Set $\mathbf{I} = (L_i)_{i\in I}$. First assume that $\mathbf{I}$ is well founded under inclusion. Note that the inclusion structure is preserved under all translations and it is therefore sufficient to let a learner exploit no other information on $\mathbf{I}$ but the inclusion structure of $\mathbf{I}$ and its automaticity. Let Φ be an automatic $\mathbf{I}$-translator.

Let a learner M process an input σ for $\Phi\langle\mathbf{I}\rangle$ as follows. If there are τ, s with $\sigma = \tau \diamond s$ and $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_{M(\tau)}\rangle$, then let $M(\sigma) = M(\tau)$, else let $M(\sigma)$ be the length-lexicographically least $i \in I$, if any, such that

(a) $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle$ and
(b) no $j \in I$ satisfies $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle \subset \Phi_{\mathbf{I}}\langle L_i\rangle$.

Now it is shown that M is consistent and conservative. Consistency needs that M be defined on all relevant input. To see that this is the case, consider any input σ for the class. Due to the well foundedness of $\mathbf{I}$ under inclusion and the definition of M, there is some i satisfying (a) and (b). Hence, M is defined on σ. Furthermore, M is consistent, as a mind change is forced whenever the old hypothesis becomes inconsistent. On the other hand, a consistent old hypothesis will not be withdrawn; hence, the learner is conservative.

So it remains to show that M actually learns the class. Consider a set to be learnt of the form $\Phi_{\mathbf{I}}\langle L_i\rangle$, as well as a text for this set. Then there is an initial segment σ of this text such that the tell-tale of $\Phi_{\mathbf{I}}\langle L_i\rangle$ is contained in $\mathrm{rng}(\sigma)$ and $\mathrm{rng}(\sigma) \nsubseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ for every $k <_{ll} i$ with $\Phi_{\mathbf{I}}\langle L_i\rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_k\rangle$. Note that i is then the length-lexicographically least index satisfying (a) and (b) in the definition of M above. Set $j = M(\sigma)$. If $j = i$ then the learner has converged to the correct index. If $j \neq i$ then $\Phi_{\mathbf{I}}\langle L_j\rangle$ cannot be a superset of $\Phi_{\mathbf{I}}\langle L_i\rangle$ due to the definition of M. Thus the learner will eventually observe a datum inconsistent with the current hypothesis. Let τ be the least initial segment of the given text with $\mathrm{rng}(\tau) \nsubseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. Then τ extends σ and M updates its hypothesis on τ to i, as i is the length-lexicographic least index satisfying (a) and (b) above with $\mathrm{rng}(\tau)$ in place of $\mathrm{rng}(\sigma)$. Hence, M converges to i and M is indeed a learner for $\Phi\langle\mathbf{I}\rangle$ as required.

Conversely, assume for a contradiction that there exists a sequence $(i_n)_{n\in\mathbb{N}}$ of members of I such that $(L_{i_n})_{n\in\mathbb{N}}$ is a $\subseteq$-descending chain. Let $(j_n)_{n\in\mathbb{N}}$ be a sequence of members of I such that for all $n \in \mathbb{N}$, the following conditions hold:

- $|j_n| < |j_{n+1}|$ and $L_{j_{n+1}} \subset L_{j_n}$;
- Infinitely many members of $\{i_0, i_1, \ldots\}$ extend the initial segment of j_{n+1} of length $|j_n|$;

– j_{n+2} extends the initial segment of j_{n+1} of length $|j_n|$.

The existence of $(j_n)_{n\in\mathbb{N}}$ can be shown by induction. Suppose we have defined $j_0, j_1, \ldots, j_{n+1}$ (where we take j_0 to be i_0, and for purposes of definition, set j_{-1} to ε). To define j_{n+2} for $n \geq -1$, note that by the inductive hypothesis, there exists an extension h of the initial segment of j_{n+1} of length $|j_n|$ which is both longer than j_{n+1} and an initial segment of infinitely many i_rs. Choose j_{n+2} to be one of these i_rs such that $L_{j_{n+2}} \subset L_{j_{n+1}}$. We now describe a coding of $(j_n)_{n\in\mathbb{N}}$ in three ω-words α, β and γ. The ω-word α is such that for all $n \in \mathbb{N}$, its initial segment of length $|j_n|$ is an initial segment of j_{n+1}. The ω-word β consists of j_0 followed by the $|j_1| - |j_0|$ last symbols of j_1 followed by the last $|j_2| - |j_1|$ symbols of j_2 and so on. The ω-word γ is an ω-word over $\{0, 1\}$ such that for all $m \in \mathbb{N}$, $\gamma(m) = 1$ iff there is $n \in \mathbb{N}$ with $|j_n| = m$. Clearly, $(j_n)_{n\in\mathbb{N}}$ can be retrieved from these three ω-words, and thus there exists a Rabin automaton which recognises all triples of ω-words which code an infinite descending chain of indices in I, see [21]. It follows that there exists an infinite regular language $R \subseteq I$ which consists of indices of an infinite descending chain of sets. Consider a first-order formula Φ (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which expresses the following condition:

if there exists $j \in R$ with $Y_j \subseteq X$,
then x is either the empty string or of the form $0y$ for some $y \in X$,
else x is of the form $0y$ for some $y \in X$.

It is easily verified that Φ is an automatic $\mathbf{I}$-translator.

Suppose a consistent learner M learns $\Phi\langle\mathbf{I}\rangle$. In response to the empty string, M must output some $i \in I$ for which there is $j \in R$ with $L_j \subseteq L_i$, and, hence, $\Phi_{\mathbf{I}}\langle L_j\rangle \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle$. But by the choice of R, the empty word can be extended to a text for a language L_h with $L_h \subset L_j$, and, hence, $\Phi_{\mathbf{I}}\langle L_h\rangle \subset \Phi_{\mathbf{I}}\langle L_j\rangle$. This implies that M cannot be conservative, completing the proof of the theorem. $\Box$

7 Strong-monotonic learning

For the learning criterion of strong-monotonicity, we first consider the concept by itself and then, in Section 8, combined with consistency. Again, we can characterise robust learnability under these restrictions, and provide further insights.

Theorem 20. *Given an automatic class $\mathbf{I} = (L_i)_{i\in I}$, clauses 1–3 are equivalent.*

1. *Every translation of $\mathbf{I}$ is strong-monotonically learnable.*
2. *Every text-preserving translation of $\mathbf{I}$ is strong-monotonically learnable.*
3. *For all $i \in I$, there exists $b_i \in I$ such that for all $j \in I$ with $L_i \not\subseteq L_j$, there exists $k \in I$ with $k \leq_{ll} b_i$, $L_i \not\subseteq L_k$ and $L_j \subseteq L_k$.*

Proof. First it is shown that 3. implies 1. and 2., by verifying that the learner M_{smon} from Example 14 learns the class. Note that M_{smon} only exploits the inclusion-structure and automaticity of the class $\Phi\langle\mathbf{I}\rangle$, making the same algorithm (using parameter Φ) work on all translations of $\mathbf{I}$. Without loss of generality, for all $i \in I$, let b_i be $\leq_{ll}$-least member of I that satisfies the third condition of the theorem. Note that for all $i \in I$, b_i is first-order definable from i; therefore the mapping $i \mapsto b_i$ is recursive (Fact 3).

Let Φ be an automatic $\mathbf{I}$-translator. Recall that M_{smon} from Example 14 works as follows: $M_{smon}(\sigma) = i$ iff i is the unique index such that $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$ for all j with $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. If such an index i does not exist then $M_{smon}(\sigma)$ is undefined. It is clear that M_{smon} is partial recursive. Furthermore, if i and j are subsequent hypotheses of M_{smon}, then $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$; hence, M_{smon} is strong-monotonic.

So the main task is to show that M_{smon} indeed learns the class $\Phi\langle\mathbf{I}\rangle$. Let $i \in I$ and a text for $\Phi_{\mathbf{I}}\langle L_i\rangle$ be given. Let σ be any initial segment of the text which is so long that $\mathrm{rng}(\sigma) \not\subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ for every $k \leq_{ll} b_i$ with $\Phi_{\mathbf{I}}\langle L_i\rangle \not\subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$. Then for all j with $\Phi_{\mathbf{I}}\langle L_i\rangle \not\subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$, there exists $k \leq_{ll} b_i$ with $\Phi_{\mathbf{I}}\langle L_i\rangle \not\subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ and $\Phi_{\mathbf{I}}\langle L_j\rangle \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$. By assumption, $\mathrm{rng}(\sigma) \not\subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ and, hence, $\mathrm{rng}(\sigma) \not\subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. Thus there exists a $\subseteq$-minimal language in the class which contains $\mathrm{rng}(\sigma)$, and i is the index of that set; so $M_{smon}(\sigma) = i$. It follows that M_{smon} strong-monotonically learns $\Phi\langle\mathbf{I}\rangle$. Hence, 3. implies 1. and 2.

Now it is shown that 2. implies 3., from which it follows that 1. implies 3. Let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. Let M be a strong-monotonic learner that learns $\Phi^{nc}\langle\mathbf{I}\rangle$. Let $i \in I$ be given, and let b_i be the $\leq_{ll}$-least member of I for which there exists $\sigma \in \mathrm{SEQ}$ such that $\mathrm{rng}(\sigma) \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_i\rangle$, $M(\sigma) = i$ and for all $s \in \mathrm{rng}(\sigma)$, $s \leq_{ll} b_i$. Then for any $j \in I$ with $\Phi^{nc}_{\mathbf{I}}\langle L_i\rangle \not\subseteq \Phi^{nc}_{\mathbf{I}}\langle L_j\rangle$, there exists $k \in \mathrm{rng}(\sigma)$ with $k \leq_{ll} b_i$ and $k \in \Phi^{nc}_{\mathbf{I}}\langle L_i\rangle \setminus \Phi^{nc}_{\mathbf{I}}\langle L_j\rangle$, implying that $L_j \subseteq L_k$ and $L_i \not\subseteq L_k$. Hence, condition 3. holds, completing the proof of the theorem. $\square$

The following theorem shows that for every automatic class, which may or may not be learnable, one can find some automatic translation which can be strong-monotonically learnt.

Theorem 21. *Every automatic class has some strong-monotonically learnable translation.*

Proof. Let $\mathbf{I} = (L_i)_{i\in I}$ be an automatic class and let Φ be the formula (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) that is defined as $\forall z(z \in Y_x \Rightarrow z \in X)$. So for all languages L, $\Phi_{\mathbf{I}}\langle L\rangle$ is the set of all $j \in I$ with $L_j \subseteq L$. Clearly, Φ is an automatic $\mathbf{I}$-translator. Let M be a learner such that for all $k \in \mathbb{N}$ and members $i, i_0, \ldots, i_k$ of I, $M\big((i_0, \ldots, i_k)\big)$ is defined and equal to i iff $L_{i_0}, \ldots, L_{i_k}$ are all subsets of L_i and i is the index of the $\subseteq$-minimal member of $\mathbf{I}$ that contains $L_{i_0}, \ldots, L_{i_k}$. It is easily verified that M learns $\Phi\langle\mathbf{I}\rangle$ and that M is strong-monotonic. $\square$

The following theorem shows that a text-preserving translation of an automatic class is strong-monotonically learnable only if the class itself is strong-monotonically learnable.

Theorem 22. *If some text-preserving translation of an automatic class* $\mathbf{I}$ *is strong-monotonically learnable, then* $\mathbf{I}$ *itself is strong-monotonically learnable.*

Proof. Set $\mathbf{I} = (L_i)_{i \in I}$. Let Φ be a text-preserving $\mathbf{I}$-translator such that $\Phi\langle\mathbf{I}\rangle$ is strong-monotonically learnable. Then for all $i \in I$, there exists a finite subset F_i of $\Phi_{\mathbf{I}}\langle L_i\rangle$ such that for all $j \in I$, if $F_i \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$ then $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. Since Φ is text-preserving, for all $i \in I$, there exists a finite subset E_i of L_i with $F_i \subseteq \Phi_{\mathbf{I}}\langle E_i\rangle$. For all members i and j of I, if $E_i \subseteq L_j$ then $\Phi_{\mathbf{I}}\langle E_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$; thus $F_i \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$, and, hence, $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$ and $L_i \subseteq L_j$. Thus, for every $i \in I$, there is a finite subset E_i of L_i such that for all $j \in I$, if $E_i \subseteq L_j$ then $L_i \subseteq L_j$. As $\mathbf{I}$ is automatic, one can determine E_i effectively from i. Thus there is a learner M which, on input $\sigma \in \mathrm{SEQ}$, outputs the $\leq_{ll}$-least $i \in I$, if it exists, such that $E_i \subseteq \mathrm{rng}(\sigma) \subseteq L_i$ — if such an i does not exist, then the learner M is undefined on σ. It is easily verified that M is a strong-monotonic learner for $\mathbf{I}$. $\Box$

8 Strong-monotonic consistent learning

In this section, we consider learners which are both strong-monotonic and consistent. The following example shows that consistency adds a genuine constraint to strong-monotonicity.

Example 23. Take $I = \{0, 1\} \cup \{2\}^\star$. Let $\mathbf{I} = (L_i)_{i \in I}$ be defined by $L_0 = \{0\}$, $L_1 = \{1\}$ and $L_{2^n} = \{0, 1\} \cup \{2^m : m \geq n\}$ for all $n \in \mathbb{N}$. Then $\mathbf{I}$ is an automatic class. Let Φ be an automatic $\mathbf{I}$-translator. Then M_{smon} from Example 14 (using $\Phi\langle\mathbf{I}\rangle$ as a parameter) learns $\Phi\langle\mathbf{I}\rangle$, due to the following behaviour on $\sigma \in \mathrm{SEQ}$: if $\mathrm{rng}(\sigma)$ is contained in exactly one of $\Phi_{\mathbf{I}}\langle L_0\rangle$ and $\Phi_{\mathbf{I}}\langle L_1\rangle$, then M_{smon} outputs 0 or 1, respectively. If there is $n \in \mathbb{N}$ with $\mathrm{rng}(\sigma)$ contained in $\Phi_{\mathbf{I}}\langle L_{2^n}\rangle$ but not in $\Phi_{\mathbf{I}}\langle L_{2^{n+1}}\rangle$, then $M_{smon}(\sigma)$ outputs 2^n. In any other case, $M_{smon}(\sigma)$ is undefined. Clearly, M_{smon} is a strong-monotonic learner that learns $\Phi\langle\mathbf{I}\rangle$.

But no consistent, strong-monotonic learner M learns $\Phi\langle\mathbf{I}\rangle$. Indeed, suppose otherwise, and let $\sigma \in \mathrm{SEQ}$ be such that $\mathrm{rng}(\sigma)$ is a subset of the union of $\Phi_{\mathbf{I}}\langle L_0\rangle$ with $\Phi_{\mathbf{I}}\langle L_1\rangle$, but not a subset of either of the sets. Then $M(\sigma)$ is equal to 2^n for some $n \in \mathbb{N}$, and $M(\tau)$ remains equal to 2^n for all $\tau \in \mathrm{SEQ}$ that extend σ and that are initial segments of a text for $\Phi_{\mathbf{I}}\langle L_{2^{n+1}}\rangle$ (the learner cannot change its mind as $L_{2^{n+1}} \subset L_{2^n}$); therefore M fails to learn $\Phi\langle\mathbf{I}\rangle$.

The following theorem gives a characterisation of every translation of an automatic class being strong-monotonically consistently learnable in terms of the class being well-ordered under inclusion. For an ordinal α, we say that an ordered set A is of type α, if its ordering is isomorphic to the ordering of α.

Theorem 24. *Given an automatic class $\mathbf{I} = (L_i)_{i \in I}$, the three conditions below are equivalent.*

1. *Every translation of $\mathbf{I}$ is strong-monotonically consistently learnable.*
2. *Every text-preserving translation of $\mathbf{I}$ is strong-monotonically consistently learnable.*
3. *$\{L_i : i \in I\}$ is $\subset$-well-ordered and of type ω at most.*

Proof. Assume that 3. holds. Given an automatic $\mathbf{I}$-translator Φ, consider a learner that on input $\sigma \in \mathrm{SEQ}$, outputs the index of the $\subseteq$-least member L of $\{L_i : i \in I\}$ with $\mathrm{rng}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L\rangle$, if such an L exists. It is easily verified that this learner is consistent and strong-monotonic and learns **I**. Hence, 3. implies both 1. and 2.

To complete the proof of the theorem, it suffices to show that 2. implies 3. So assume that 2. holds. We first show that for any members i and j of I, L_i and L_j are $\subseteq$-comparable. Suppose otherwise for a contradiction, and fix i, j such that $L_i \nsubseteq L_j$ and $L_j \nsubseteq L_i$. Consider a first-order formula Φ (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which expresses the conjunction of the two conditions that follow:

- If X intersects $Y_i \setminus Y_j$ or $Y_j \setminus Y_i$, then either x is the empty string or x is of the form $0y$ for some $y \in X$.
- If X intersects neither $Y_i \setminus Y_j$ nor $Y_j \setminus Y_i$, then x is of the form $0y$ for some $y \in X$.

It is easily verified that Φ is an automatic **I**-translator and is text-preserving. Let M be a consistent and strong-monotonic learner that learns $\Phi\langle\mathbf{I}\rangle$.

Since M is consistent, M must output, in response to the input sequence consisting of the empty string ε, a member k of I with $\varepsilon \in \Phi_{\mathbf{I}}\langle L_k\rangle$. By definition of Φ, L_k contains an element outside L_i or an element outside L_j. Hence, $\Phi_{\mathbf{I}}\langle L_k\rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_i\rangle$ or $\Phi_{\mathbf{I}}\langle L_k\rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. It follows that M cannot be strong-monotonic as it must be able to switch its hypotheses from k to i or j when it is presented with a text for $\Phi_{\mathbf{I}}\langle L_i\rangle$ or $\Phi_{\mathbf{I}}\langle L_j\rangle$, respectively.

Next we show that for all $i \in I$ such that L_i is not $\subseteq$-minimal, there exists $j \in I$ with $L_j \subset L_i$ such that there exists no $k \in I$ with $L_j \subset L_k \subset L_i$. Assume otherwise for a contradiction, and choose $i \in I$ for which this property does not hold. Consider a first-order formula Φ' (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which express that

$$x \in X \text{ and } (x \notin Y_i \text{ or there exists } j \in I \text{ with } x \in Y_j \text{ and } Y_j \subset Y_i).$$

We now show that for all $j \in I$ and $k \in I$, if $L_j \nsubseteq L_k$ then $\Phi'_{\mathbf{I}}\langle L_j\rangle \nsubseteq \Phi'_{\mathbf{I}}\langle L_k\rangle$. Clearly, it suffices to verify that for all $j \in I$, if $L_j \subset L_i$ then $\Phi'_{\mathbf{I}}\langle L_j\rangle \subset \Phi'_{\mathbf{I}}\langle L_i\rangle$. So let $j \in I$ be such that $L_j \subset L_i$. By the choice of i, there exists $k \in I$ with $L_j \subset L_k \subset L_i$. Let $x \in L_k \setminus L_j$ be given. Then x belongs to $\Phi'_{\mathbf{I}}\langle L_i\rangle$; hence, $x \in \Phi'_{\mathbf{I}}\langle L_i\rangle \setminus \Phi'_{\mathbf{I}}\langle L_j\rangle$. Hence, Φ' is noninclusion preserving for members of **I**; it follows easily that Φ' is an automatic **I**-translator and is text-preserving. However, as $\Phi'_{\mathbf{I}}\langle L_i\rangle$ is the ascending union of the sets of the form $\Phi'_{\mathbf{I}}\langle L_j\rangle$ with $L_j \subset L_i$, $\Phi'\langle\mathbf{I}\rangle$ cannot be learnable [1, 5], a contradiction.

To complete the proof of the theorem, it suffices to show that for all $i \in I$, there exist only finitely many $j \in I$ with $L_j \subset L_i$. For a contradiction, assume otherwise. Let R be the set of all $i \in I$ for which there exist infinitely many $j \in I$ with $L_j \subset L_i$. By assumption, R is not empty and by the previous paragraph, there is no $i \in R$ such that $L_i \subseteq L_j$ for all $j \in R$. Consider a first-order formula Φ'' (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which expresses that

- if there exists $z \in X$ and $j \in R$ such that $z \notin Y_j$, then x is either the empty string or of the form $0y$ for some $y \in X$;
- if for all $z \in X$ and $j \in R$, $z \in Y_j$, then x is of the form $0y$ for some $y \in X$.

It is easily verified that Φ'' is an automatic $\mathbf{I}$-translator and is text-preserving.

Since M is consistent, M must output in response to the input sequence consisting of the empty string ε a member i of I with $\varepsilon \in \Phi''_{\mathbf{I}}\langle L_i \rangle$. So there is a $j \in R$ and $y \in L_i$ with $y \notin L_j$; hence, $L_j \subset L_i$. Furthermore, there is $k \in R$ with $L_k \subset L_j$; so there exists $z \in L_j \setminus L_k$, implying that $\varepsilon \in \Phi''_{\mathbf{I}}\langle L_j \rangle$. We infer that M overgeneralised in response to (ε), in contradiction with the assumption that M is strongly-monotonic and learns $\Phi''\langle \mathbf{I} \rangle$. We conclude that R is empty, completing the proof of the theorem. $\Box$

The following theorem characterises when some translation of an automatic class is consistently and strong-monotonically learnable in terms of the existence of least upper bounds whenever every finite collection of languages in the class are bounded (with respect to inclusion) by a language in the class.

Theorem 25. *Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given. Consider the following clause*

($\star$) *For all finite $F \subseteq I$, if there exists $i \in I$ with $L_k \subseteq L_i$ for all $k \in F$, then there exists $i \in I$ such that $L_k \subseteq L_i$ for all $k \in F$, and*

$$\textit{for all } j \in I,\ L_i \subseteq L_j \Leftrightarrow \forall k \in F (L_k \subseteq L_j)$$

which expresses that every finite subset of $\mathbf{I}$ which is $\subseteq$-bounded in $\mathbf{I}$ has a (necessarily unique) $\subseteq$-least upper bound in $\mathbf{I}$. Then statements 1, 2 below hold.

1. *There exists an automatic $\mathbf{I}$-translator Φ such that $\Phi\langle \mathbf{I} \rangle$ is consistently and strongly-monotonically learnable iff ($\star$) holds.*
2. *Suppose that the class $\mathbf{I}$ is strongly-monotonically learnable. Then some text-preserving automatic translation of $\mathbf{I}$ is strongly-monotonically and consistently learnable iff ($\star$) holds.*

Proof. It is convenient to prove both results together. For this, some notation is needed. If $\mathbf{I}$ is not strong-monotonically learnable then for all $i \in I$, let $E_i = L_i$. If $\mathbf{I}$ is strong-monotonically learnable then for all $i \in I$, let $E_i = \emptyset$ if $L_i = \emptyset$; otherwise let $E_i = \{y \in L_i : y \leq_{ll} z\}$ for the $\leq_{ll}$-least $z \in L_i$ such that for all $j \in I$, if $\{y \in L_i : y \leq_{ll} z\} \subseteq L_j$ then $L_i \subseteq L_j$ (such a z exists as any strong-monotonic learner must conjecture i based on a finite input σ for L_i). Note that in both cases above, for all $i \in I$, E_i is first-order definable from L_i.

Now sufficiency is shown. So assume that ($\star$) holds and, in case of 2., that $\mathbf{I}$ is also strong-monotonically learnable and the sets E_i are therefore finite. Consider a first-order formula Φ (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing, where E_i is obtained as above) expressing that E_x is a subset of X. Hence, $\Phi_{\mathbf{I}}\langle L \rangle = \{x \in I : E_x \subseteq L\}$. It is easily verified that Φ is an automatic $\mathbf{I}$-translator. Furthermore, in the case of 2., the sets E_i, $i \in I$, are finite and Φ is text-preserving.

We now show that $\Phi\langle \mathbf{I} \rangle$ is strongly-monotonically and consistently learnable. Define a learner M as follows. Presented with a finite set F of data, M outputs the member i of I such that L_i is the $\subseteq$-least upper bound of the sets L_k with $k \in F$, which exists by ($\star$). Note that in case $F = \varnothing$, M still can output a conjecture as ($\star$) implies that there is a $\subseteq$-least language in $\mathbf{I}$. To see that M learns $\Phi\langle \mathbf{I} \rangle$, note that whenever M is presented with a text for $\Phi_{\mathbf{I}}\langle L_i \rangle$, then i occurs in the text and from that point onwards, M outputs i since L_i is the $\subseteq$-least upper bound of any class of languages which contains L_i and which only contains sets L_j satisfying $E_j \subseteq L_i$. Hence, M learns $\Phi_{\mathbf{I}}\langle L_i\rangle$. Clearly, M is consistent. Furthermore, M is strong-monotonic: if $F \subseteq F'$ then also the $\subseteq$-least upper bound of $\{L_j : j \in F\}$ is a subset of the $\subseteq$-least upper bound of $\{L_j : j \in F'\}$. This completes the proof of sufficiency for the claims given in 1. and 2., respectively.

For necessity, note that by Theorem 22, it is necessary in the case of 2. that $\mathbf{I}$ is strong-monotonically learnable. Hence, it suffices to show $(\star)$ in both cases. Let Φ be an automatic $\mathbf{I}$-translator that in the case of 2., is text-preserving. Let M be a consistent and strong-monotonic learner that learns $\Phi\langle\mathbf{I}\rangle$. Let F be a finite subset of I. For each $j \in F$, there is a finite subset G_j of $\Phi_{\mathbf{I}}\langle L_j\rangle$ such that $\Phi_{\mathbf{I}}\langle L_j\rangle \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ whenever $G_j \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$. Now assume that M is presented with data that include the union of all sets G_j with $j \in F$ — it is a finite set. Then M outputs a conjecture i such that $\Phi_{\mathbf{I}}\langle L_i\rangle$ contains all data seen so far. As M is strong-monotonic, $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ for all k where $\Phi_{\mathbf{I}}\langle L_k\rangle$ contains the data seen so far. Hence, $\Phi_{\mathbf{I}}\langle L_i\rangle$ is a $\subseteq$-least upper bound, in $\Phi\langle\mathbf{I}\rangle$, of the sets $\Phi_{\mathbf{I}}\langle L_j\rangle$ with $j \in F$. As Φ preserves inclusions and noninclusions within $\mathbf{I}$, it follows that L_i is the $\subseteq$-least upper bound of all the L_j with $j \in F$. Hence, $(\star)$ holds. $\square$

Note that a class that contains an infinite ascending chain is not confidently learnable. The following theorem follows from this observation along with the results about strong-monotonic learning shown above.

Theorem 26. *Given an automatic class $\mathbf{I} = (L_i)_{i \in I}$, statements 1–4 below hold.*

1. *Assume that the class $\mathbf{I}$ is strong-monotonically learnable. Now the class $\mathbf{I}$ is confidently learnable iff it contains no infinite ascending chain.*
2. *Every translation of $\mathbf{I}$ is strong-monotonically and confidently learnable iff $\mathbf{I}$ does not contain infinite ascending chains and for all $i \in I$, there exists $b_i \in I$ such that for all $j \in I$, if $L_i \not\subseteq L_j$ then there is $k \leq_{ll} b_i$ with $L_j \subseteq L_k$ and $L_i \not\subseteq L_k$.*
3. *Some translation of $\mathbf{I}$ is strong-monotonically and confidently learnable iff $\mathbf{I}$ contains no infinite ascending chain.*
4. *If some text-preserving translation of $\mathbf{I}$ is strong-monotonically and confidently learnable, then $\mathbf{I}$ itself is strong-monotonically and confidently learnable.*

As an immediate corollary of the above we get the following corollary.

Corollary 27. *The three statements below hold.*

1. *Every translation of an automatic class $\mathbf{I}$ is consistently, confidently and strong-monotonically learnable iff $\mathbf{I}$ is a finite chain of languages.*
2. *Some translation of an automatic class $\mathbf{I}$ is consistently, confidently and strong-monotonically learnable iff $\mathbf{I}$ has no infinite ascending chain and every $\subseteq$-bounded finite subclass of $\mathbf{I}$ has a $\subseteq$-least upper bound, that is, for all finite $F \subseteq I$, if there is $i \in I$ with $\bigcup_{k \in F} L_k \subseteq L_i$, then there is $i \in I$ with $\bigcup_{k \in F} L_k \subseteq L_i$ and $L_i \subseteq L_h$ for all $h \in I$ with $\bigcup_{k \in F} L_k \subseteq L_h$.*

3. *Some text-preserving translation of an automatic class* **I** *is consistently, confidently and strong-monotonically learnable iff* **I** *satisfies the conditions in 2. and* **I** *itself is strong-monotonically learnable.*

These results give a full characterisation on how confident learnability combines with strong-monotonic learning. We should also observe the fact that every translation of a class being confidently learnable does not imply that the class is strong-monotonically learnable:

Example 28. Consider the automatic class **I** which contains the set $\{0\}^\star$ and for all $n > 0$ the sets $\{0^m : m < n\} \cup \{1^n\}$ and $\{0\}^\star \cup \{1^m : m \geq n\}$. This class is not strong-monotonically learnable as any learner that learns **I** must output the index for $\{0\}^\star$ after seeing a finite sequence of suitable examples. But then, for sufficiently large n, it is necessary to make a mind change to the index for $\{0^m : m < n\} \cup \{1^n\}$ to learn that set from any text for it which extends σ.

Still for every automatic **I**-translator Φ, $\Phi\langle\mathbf{I}\rangle$ is confidently learnable by a learner M that proceeds as follows. As long as the data is consistent with $\Phi_{\mathbf{I}}\langle\{0\}^\star\rangle$, M conjectures the index for $\Phi_{\mathbf{I}}\langle\{0\}^\star\rangle$. If there exists (a necessarily unique) $n > 0$ such that the data seen so far is consistent with the set $\Phi_{\mathbf{I}}\langle\{0^m : m < n\} \cup \{1^n\}\rangle$ but not with $\Phi_{\mathbf{I}}\langle\{0\}^\star \cup \{1^m : m > n\}\rangle$, then M outputs the index for $\Phi_{\mathbf{I}}\langle\{0^m : m < n\} \cup \{1^n\}\rangle$. Otherwise, if presented with some data that is consistent with $\Phi_{\mathbf{I}}\langle 0^\star \cup \{1^m : m \geq n\}\rangle$ for all $n \in \mathbb{N}$, but not with $\Phi_{\mathbf{I}}\langle 0^\star\rangle$, M outputs its previous hypothesis. Otherwise, M conjectures the index for $\Phi_{\mathbf{I}}\langle\{0\}^\star \cup \{1^m : m \geq n\}\rangle$ where $n > 0$ is largest for which the set is consistent with the input data; n might go down as more data are presented, but will eventually stabilise. Hence, $\Phi\langle\mathbf{I}\rangle$ is confidently learnable.

9 Confident learning

A characterisation of classes whose every translation is confidently learnable by a computable learner is open. Theorem 30 deals with the case of general learners.

Theorem 29. *Every translation of an automatic class* **I** *is confidently, conservatively and consistently learnable iff* **I** *is finite.*

Proof. Assume that every translation of **I** is confidently, conservatively and consistently learnable. Confidence implies that **I** contains no infinite ascending chain of languages. Conservativeness and consistency imply that **I** has no infinite descending chain of languages. For a contradiction, assume that **I** contains an infinite antichain.

By arguments similar to those in the proof of Theorem 19, there is an infinite regular set R which consists of indices of an antichain. Consider a first-order formula Φ (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which expresses that

> x is either of the form $0y$ for some $y \in X$ or of the form $1^{|i|+1}01^n$ for some $i \in R$ and $n \in \mathbb{N}$ such that either $Y_i \subseteq X$ or there is a $j \in R$ with $|j| > |i| + 2 + n$ and $Y_j \subseteq X$.

It is easily verified that Φ is an automatic $\mathbf{I}$-translator.

Note that if a language L is a superset of L_j for infinitely many $j \in I$ with $j \in R$, then $\Phi_{\mathbf{I}}\langle L\rangle$ contains all strings of the form $1^{|i|+1}01^n$. Furthermore, for every finite set E of such strings and almost all $j \in R$, $E \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. These two facts will now be used to disprove that $\mathbf{I}$ is confidently, conservatively and consistently learnable.

Present all strings of the form $1^{|i|+1}01^n$ to a consistent learner M that learns $\Phi\langle\mathbf{I}\rangle$. If M converges to an index k, then $\Phi_{\mathbf{I}}\langle L_k\rangle$ contains infinitely many sets of the form $\Phi_{\mathbf{I}}\langle L_i\rangle$ with $i \in R$. As k is output after finitely many data have been received, there exists $i \in R$ such that $L_i \subset L_k$ and k has been output after only data from $\Phi_{\mathbf{I}}\langle L_i\rangle$ have been received; hence, M cannot be conservative. If M does not converge to an index k, then M is not confident. Hence, there is no consistent, conservative and confident learner that learns $\Phi\langle\mathbf{I}\rangle$. Hence, $\mathbf{I}$ does not have an infinite antichain. As every infinite class contains an infinite ascending chain or an infinite descending chain or an infinite antichain, $\mathbf{I}$ must be finite.

For the sufficiency, assume that $\mathbf{I}$ is finite. Then for all automatic $\mathbf{I}$-translators Φ, $\Phi\langle\mathbf{I}\rangle$ is finite, and by Theorem 19, there exists a consistent and conservative learner M that learns $\Phi\langle\mathbf{I}\rangle$. Without loss of generality, this learner never returns to the index for a language that has been conjectured and then abandoned. As there are only finitely many indices, M is also confident. $\square$

The following result is the only one that involves general learners rather than computable learners. Recall the definition of Φ^{nc} from Example 7.

Theorem 30. *Let $\mathbf{I} = (L_i)_{i\in I}$ be an automatic class all of whose translations are learnable. Then both conditions below are equivalent.*

- *Every translation of I is confidently learnable by some general learner.*
- *There exists no nonempty subset J of I such that for all $i \in J$ and finite subsets F of $\Phi^{nc}_{\mathbf{I}}\langle L_i\rangle$, there exists $j \in J$ with $F \cup \{i\} \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_j\rangle$.*

Proof. First one tries to build by induction along the ordinals α below ω_1 a sequence of distinct indices i_α and corresponding bounds b_α such that for each L_{i_α} and $j \in I \setminus \{i_\beta : \beta < \alpha\}$, $\Phi^{nc}_{\mathbf{I}}\langle L_j\rangle$ does not contain the union of $\{i_\alpha\}$ with $\{k \in \Phi^{nc}_{\mathbf{I}}\langle L_{i_\alpha}\rangle : k \leq_{ll} b_\alpha\}$. This induction stops at some ordinal $\gamma < \omega_1$ (that is, i_γ, b_γ do not get defined, but i_α and b_α get defined for all $\alpha < \gamma$). Now let $J = I \setminus \{i_\alpha : \alpha < \gamma\}$. There are two cases:

(a) J is empty. Let Φ be any automatic $\mathbf{I}$-translator. Then one can build the following learner M which is a restriction of the learner M_{ex} defined in Example 14 on the class $\Phi\langle\mathbf{I}\rangle$. After seeing some data, M conjectures i_α iff M_{ex} conjectures i_α on the same data and for each $k \leq_{ll} b_\alpha$ with $k \in \Phi^{nc}_{\mathbf{I}}\langle L_{i_\alpha}\rangle$, some member of $\Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle \setminus \Phi_{\mathbf{I}}\langle L_k\rangle$ has been observed in the data seen so far; otherwise M is undefined. Then M learns every language $\Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle$ as on a text for L_{i_α}, after some finite time M_{ex} has converged to i_α and for all $k \leq_{ll} b_\alpha$ with $k \in \Phi^{nc}_{\mathbf{I}}\langle L_{i_\alpha}\rangle$, some datum in $\Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle \setminus \Phi_{\mathbf{I}}\langle L_k\rangle$ has been observed; hence, M outputs i_α from then onward as well. To see that M is confident, consider any two hypotheses i_α and i_β consecutively output by M on some text. Assume for a contradiction that $\beta > \alpha$. By definition of i_α and b_α, there exists $k \in I$ with $k \leq_{ll} b_\alpha$ such that $k \in \{i_\alpha\} \cup \Phi^{nc}_{\mathbf{I}}\langle L_{i_\alpha}\rangle \setminus \Phi^{nc}_{\mathbf{I}}\langle L_{i_\beta}\rangle$. If $k = i_\alpha$ then $L_{i_\beta} \subseteq L_{i_\alpha}$, and, hence, $\Phi_{\mathbf{I}}\langle L_{i_\beta}\rangle \subseteq \Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle$. This is in contradiction with the following facts taken together:

- i_α and i_β are consecutively output by M_{ex};
- M_{ex} never makes a mind change from a hypothesis for a set to a hypothesis for a proper subset of that set.

Otherwise, if $k \in \Phi^{nc}_{\mathbf{I}}\langle L_{i_\alpha}\rangle$ then $L_{i_\beta} \subseteq L_k$, $\Phi_{\mathbf{I}}\langle L_{i_\beta}\rangle \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ and i_α is output only after observing some data outside $\Phi_{\mathbf{I}}\langle L_k\rangle$; hence, M does not output the hypothesis i_β, which is then inconsistent with the data observed before. From this contradiction it follows that $\beta < \alpha$; hence, the ordinals by which the indices in I are indexed go down with every new hypothesis. It follows that the learner M is confident.

(b) In case J is not empty, one can show that every general learner M that learns $\{\Phi^{nc}_{\mathbf{I}}\langle L_i\rangle : i \in I\}$ is not confident. To see this, consider such a general learner and pick iteratively some members $j_0, j_1, \ldots$ of J as follows. The inductive definition has the following invariants for any fixed $j_0 \in J$ and $n \in \mathbb{N}$:

- $\mathrm{rng}(\sigma_n) \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_{j_n}\rangle$ and $M(\sigma_n) = j_n$;
- $\sigma_n \subseteq \sigma_{n+1}$ and $j_n \in \mathrm{rng}(\sigma_{n+1})$;
- $\mathrm{rng}(\sigma_n) \cup \{j_n\} \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_{j_{n+1}}\rangle$.

Note that j_{n+1} can always be picked from the set J as otherwise one could take $i_\gamma = j_n$ and $b_\gamma = \leq_{ll}$-maximal element of $(\mathrm{rng}(\sigma_n) \cup \{j_n\})$ in the definition at the beginning of the proof, and thus extend the induction, which by assumption could not be extended at γ. As $\Phi^{nc}_{\mathbf{I}}\langle L_{j_{n+1}}\rangle$ contains $\mathrm{rng}(\sigma_n)$ and $j_n \in \Phi^{nc}_{\mathbf{I}}\langle L_{j_{n+1}}\rangle$, there exists a σ_{n+1} extending $\sigma_n \diamond j_n$ such that $\mathrm{rng}(\sigma_{n+1}) \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_{j_{n+1}}\rangle$ and $M(\sigma_{n+1}) = j_{n+1}$. This completes the inductive definition. It follows that M is not a confident learner.

We conclude that either every translation of I is confidently learnable by a general learner, or the subset J of I considered in the proof witnesses that for all $i \in J$ and finite subsets F of $\Phi^{nc}_{\mathbf{I}}\langle L_i\rangle$, there exists $j \in J$ with $F \cup \{i\} \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_j\rangle$.

On the other hand, if the second item in the statement of the theorem does not hold, then $J \neq \emptyset$ in the construction at the beginning of the proof, and thus $\mathbf{I}$ is not confidently learnable by any general learner. This completes the proof of the theorem. □

10 Finite learning

A more restrictive notion of learning is finite learning where the very first conjecture output by the learner has to correctly identify the set to be learnt. Obviously, finitely learnable classes are antichains as otherwise one could see the data for a set L_i and conjecture an index for this set only to find out later that the set to be learnt is actually a superset of L_i. So a key question is to characterise the size of these antichains.

Theorem 31. *Let an automatic class* $\mathbf{I}$ *be given. Statements 1–3 below hold.*

1. *Every text-preserving translation of* $\mathbf{I}$ *is finitely learnable iff* $\mathbf{I}$ *is a finite antichain.*
2. *Some translation of* $\mathbf{I}$ *is finitely learnable iff* $\mathbf{I}$ *is an antichain.*
3. *If* $\mathbf{I}$ *has a finitely learnable text-preserving translation, then* $\mathbf{I}$ *itself is finitely learnable.*

Proof. Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class.

1. Finite antichains are clearly finitely learnable by a learner M that waits until there is a unique $i \in I$ such that for each $j \in I$ distinct from i, some member of $L_i \setminus L_j$ is part of the input, at which point M correctly conjectures i.

For the converse direction, assume that $\mathbf{I}$ is an infinite antichain. Let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. For all $i \in I$, $\Phi^{nc}_{\mathbf{I}}\langle L_i \rangle$ is equal to $I \setminus \{i\}$. Thus, $\Phi^{nc}\langle \mathbf{I} \rangle$ is not finitely learnable.

2. Suppose that $\mathbf{I}$ is an antichain. Consider the first-order formula Φ (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which expresses that $Y_x \subseteq X$. It is easily verified that Φ is an automatic $\mathbf{I}$-translator. Then for all $i \in I$, $\Phi\langle L_i \rangle$ is the singleton $\{i\}$. Obviously, $\Phi\langle \mathbf{I} \rangle$ is finitely learnable.

Conversely, let Φ be an automatic $\mathbf{I}$-translator and let M be a learner such that M finitely learns $\Phi\langle \mathbf{I} \rangle$. Then for all $i \in I$, there exists a finite sequence σ of members of $\Phi_{\mathbf{I}}\langle L_i \rangle \cup \{\#\}$ such that $M(\sigma) = i$, and $\mathrm{rng}(\sigma)$ is not contained in $\Phi_{\mathbf{I}}\langle L_j \rangle$ for any $j \in I$ satisfying $\Phi_{\mathbf{I}}\langle L_j \rangle \neq \Phi_{\mathbf{I}}\langle L_i \rangle$. Hence, $\Phi_{\mathbf{I}}\langle L_i \rangle \not\subset \Phi_{\mathbf{I}}\langle L_j \rangle$ for all $j \in I$. Thus, $L_i \not\subset L_j$ for all $j \in I$.

3. Assume that Φ is a text-preserving automatic $\mathbf{I}$-translator which maps $\mathbf{I}$ to a finitely learnable class. Let M be a learner that finitely learns $\Phi\langle \mathbf{I} \rangle$. Then for every $i \in I$, there is a finite subset E_i of $\Phi_{\mathbf{I}}\langle L_i \rangle$ such that M outputs i on some finite input sequence containing only members of E_i. Let $i \in I$ be given. Then there exists a finite subset F_i of L_i with $E_i \subseteq \Phi_{\mathbf{I}}\langle F_i \rangle$. Now $F_i \nsubseteq L_j$ for all $j \in I \setminus \{i\}$. One can give the following first-order definition of a finite set G_i with the same property:

$$G_i = \big\{ x \in L_i : \exists j \in I \setminus \{i\}\, \forall y <_{ll} x\, [y \in L_i \Rightarrow y \in L_j] \big\}.$$

Hence, there is a finite learner which outputs i iff i is the $\leq_{ll}$-least member of I such that G_i is contained in the data observed. Thus $\mathbf{I}$ is finitely learnable. $\square$

11 Learning from queries

Whereas learnability in the limit offers a model of passive learning, learning from queries allows agents to play an active role by questioning an oracle on some properties that the target language might have, and sometimes getting further clues [2]. Four kinds of queries are usually used, alone or in combination. In a given context, the selection of queries that the learner is allowed to make is dictated by the desire to obtain natural, elegant and insightful characterisations. Some studies have compared the passive and active models of learning, which usually turn out to be different [19]. Interestingly, in our model both families of paradigms bind tightly as learnability of a class of languages from superset queries is equivalent to learnability from positive data of every translation of the class (Corollary 42 below).

Definition 32. Let T be the set of queries of at least one of the following types:

Membership query: is $x \in L$?	Subset query: is $L_e \subseteq L$?
Superset query: is $L_e \supseteq L$?	Equivalence query: is $L_e = L$?

Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given.

An $\mathbf{I}$-*query learner of type* T is a machine M such that, for all $i \in I$, when learning L_i, M makes finitely many queries from T, possibly taking into account the answers to earlier queries, with all queries answered correctly w.r.t. $L = L_i$, and eventually outputs a member of I.

An $\mathbf{I}$-query learner of type T *learns* $\mathbf{I}$ iff for all $i \in I$, i is the member of I that M eventually outputs when learning L_i. A *query learner of type* T *for* $\mathbf{I}$ is an $\mathbf{I}$-query learner of type T that learns $\mathbf{I}$.

$\mathbf{I}$ is *learnable from queries of type* T iff a query learner of type T for $\mathbf{I}$ exists.

When T is clear from the context, we omit to mention "of type T."

Remark 33. *Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ and an automatic $\mathbf{I}$-translator Φ be given. Note that Φ preserves $\subseteq$ relation, that is, $L_i \subseteq L_j$ iff $\Phi_{\mathbf{I}}(L_i) \subseteq \Phi_{\mathbf{I}}(L_j)$. As our queries do not involve counterexamples, for query types T consisting only of subset, superset and equivalence queries, it immediately follows that "learnability of $\mathbf{I}$ from queries of type T" and "learnability of $\Phi\langle\mathbf{I}\rangle$ from queries of type T" have the same answer. Observe that subset, superset and equivalence queries are only with reference to languages in $\mathbf{I}$, or $\Phi\langle\mathbf{I}\rangle$, respectively.*

We immediately have the following result.

Theorem 34. *Every automatic class is learnable from equivalence queries.*

We illustrate query learning with a few examples of automatic classes.

Example 35. All translations of the classes below are learnable from membership queries:

- $\big\{\{x \in \{0,1\}^\star : x \text{ is a prefix of } y \vee y \text{ is a prefix of } x\} : y \in \{0,1\}^\star\big\}$.
- $\big\{\{0\}^\star \cup \{1^m : m \leq n\} : n > 0\big\} \cup \big\{\{0^m : m \geq n\} : n > 0\big\}$.
- Any finite class.

Example 36. Given an automatic class $\mathbf{I}$, let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. Then $\Phi^{nc}\langle\mathbf{I}\rangle$ is learnable from membership and subset queries by searching for the unique $i \in I$ which satisfies that $i \notin \Phi^{nc}_{\mathbf{I}}\langle L\rangle \wedge \Phi^{nc}_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi^{nc}_{\mathbf{I}}\langle L\rangle$. Indeed, a negative answer to the membership query for i implies $\Phi^{nc}_{\mathbf{I}}\langle L\rangle \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_i\rangle$ and so $\Phi^{nc}_{\mathbf{I}}\langle L_i\rangle = \Phi^{nc}_{\mathbf{I}}\langle L\rangle$.

Example 37. Let an automatic class $\mathbf{I}$ be given and let Φ be a, not necessarily text-preserving, automatic $\mathbf{I}$-translator satisfying $\Phi_{\mathbf{I}}\langle L\rangle = \{i \in I : L_i \subseteq L\}$ for all languages L. Then $\Phi\langle\mathbf{I}\rangle$ is learnable from membership and superset queries: a $\Phi\langle\mathbf{I}\rangle$-query learner can search for the unique $i \in I \cap \Phi_{\mathbf{I}}\langle L\rangle$ with $\Phi_{\mathbf{I}}\langle L\rangle \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle$. This i satisfies $\Phi_{\mathbf{I}}\langle L_i\rangle = \Phi_{\mathbf{I}}\langle L\rangle$ and can be found when the learner is allowed both kinds of queries.

Example 38. Consider the automatic class $\mathbf{I}$ consisting of $\{0,1\}^\star$ and all co-singletons of the form $\{0,1\}^\star \setminus \{x\}$ with $x \in \{0,1\}^\star$. Then none of $\mathbf{I}$'s text-preserving translations is learnable from superset and membership queries. Let Φ be a text-preserving $\mathbf{I}$-translator, and assume for a contradiction that a query learner M for $\Phi\langle\mathbf{I}\rangle$ outputs an index for $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$ after finitely many superset and membership queries on x_1, x_2, …, x_n. Here, the superset query "is $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle \supseteq$

L?" receives the answer "yes", and for all $i \in I$ with $L_i \neq \{0,1\}^\star$, the superset query "is $\Phi_{\mathbf{I}}\langle L_i\rangle \supseteq L$?" receives the answer "no". Furthermore, the membership queries "is $x_k \in L$?" receives the answer based on whether $x_k \in \Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$. Now for each $x_k \in \Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$, there is a finite subset E_k of $\{0,1\}^\star$ with $x_k \in \Phi_{\mathbf{I}}\langle E_k\rangle$. Consider any $y \in \{0,1\}^\star$ satisfying the following conditions:

- for all $k \in I$ such that M has queried the membership of x_k to the target language when learning $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$, $y \notin E_k$;
- the superset query "is $\Phi_{\mathbf{I}}\langle L\rangle \subseteq \Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{y\}\rangle$?" has not been asked by M when learning $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$.

Then all queries would have received the same answer if the language L to be learnt was $\Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{y\}\rangle$; therefore M cannot distinguish $\Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{y\}\rangle$ from $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$. Hence, M is incorrect and $\Phi\langle\mathbf{I}\rangle$ is not learnable from superset and membership queries.

Theorem 39. *Every automatic class has a translation learnable using membership queries.*

Proof. Let $\mathbf{I} = (L_i)_{i\in I}$ be an automatic class. Consider a first-order formula Φ (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which expresses that either x is of the form $i0$ for some $i \in I$ with $X \nsubseteq Y_i$, or x is of the form $i1$ for some $i \in I$ with $Y_i \subseteq X$. It is easy to verify that Φ is an automatic $\mathbf{I}$-translator; note that Φ is not text-preserving. In order to learn $\Phi\langle\mathbf{I}\rangle$, a $\Phi\langle\mathbf{I}\rangle$-query learner can search for the first $i \in I$ such that $i0 \notin \Phi_{\mathbf{I}}\langle L\rangle \wedge i1 \in \Phi_{\mathbf{I}}\langle L\rangle$. Since $i0 \notin \Phi_{\mathbf{I}}\langle L\rangle$, $L \subseteq L_i$. Since $i1 \in \Phi_{\mathbf{I}}\langle L\rangle$, $L_i \subseteq L$. Hence, i is uniquely determined and is such that $\Phi_{\mathbf{I}}\langle L\rangle = \Phi_{\mathbf{I}}\langle L_i\rangle$. □

The theorem and corollary that follow characterise learnability from subset and superset queries. These results have a similar flavour as Theorems 4, 5 and 10 in [16], obtained in the context of indexable classes of r.e. languages and a broader class of queries.

Theorem 40. *Let an automatic class $\mathbf{I} = (L_i)_{i\in I}$ be given. Then $\mathbf{I}$ is learnable from subset queries iff for all $i \in I$, there exists $b_i \in I$ such that for all $j \in I$ with $L_i \subset L_j$, there exists $k \in I$ with $k \leq_{ll} b_i$ and $L_k \subseteq L_j \wedge L_k \nsubseteq L_i$.*

Proof. Suppose that for all $i \in I$, there exists $b_i \in I$ that satisfies the condition of the theorem. Note that there exists a computable function that maps any $i \in I$ to a member b_i of I that satisfies the condition of the theorem. Hence, an $\mathbf{I}$-query learner can, using subset queries $L_j \subseteq L$ where L is the language to be learnt, find and output the first $i \in I$ such that $L_i \subseteq L$ and for all $k \in I$ with $k \leq_{ll} b_i$, $L_k \subseteq L$ iff $L_k \subseteq L_i$. Obviously $L_i = L$. Note that testing whether $L_k \subseteq L_i$ is recursive as the structure is automatic.

Conversely, assume that there exists $i \in I$ such that no $b_i \in I$ satisfies the condition of the theorem. For a contradiction, suppose that M is a query learner for $\mathbf{I}$ that uses subset queries. Then there exists $b_i \in I$ such that M outputs i after asking subset queries of the form $L_k \subseteq L$ only for L_k with $k \leq_{ll} b_i$, answered w.r.t. $L = L_i$. By the choice of i, there exists $j \in I$ such that $L_i \subset L_j$ and there exists no member k of I with $k \leq_{ll} b_i$, $L_k \subseteq L_j$ and $L_k \nsubseteq L_i$. Hence, all queries involving indices $k \leq_{ll} b_i$ are answered in the same way when learning $L = L_i$ and when learning $L = L_j$. Hence, the algorithm would give the same answer i when learning L_j and thus cannot be correct. $\square$

A similar result can be obtained when using superset queries only:

Corollary 41. *Let an automatic class* $\mathbf{I} = (L_i)_{i \in I}$ *be given. Then* $\mathbf{I}$ *is learnable from superset queries iff for all* $i \in I$*, there exists* $b_i \in I$ *such that for all* $j \in I$ *with* $L_i \supset L_j$*, there exists* $k \in I$ *with* $k \leq_{ll} b_i$ *and* $L_k \supseteq L_j \wedge L_k \not\supseteq L_i$.

The following corollary is a consequence of Theorem 17.

Corollary 42. *An automatic class* $\mathbf{I}$ *is learnable from superset queries iff every translation of* $\mathbf{I}$ *is learnable from positive data.*

Given an automatic class $\mathbf{I}$ of languages all of whose text-preserving translations are learnable from superset and membership queries, $\mathbf{I}$-query learners that ask superset queries do not benefit from also asking membership queries:

Theorem 43. *If every text-preserving translation of an automatic class* $\mathbf{I}$ *is learnable from membership and superset queries, then* $\mathbf{I}$ *itself is learnable from superset queries.*

Proof. Suppose $\mathbf{I} = (L_i)_{i \in I}$. Let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. When learning $\Phi^{nc}\langle\mathbf{I}\rangle$, a $\Phi^{nc}\langle\mathbf{I}\rangle$-query learner can replace every membership query of the form "is $i \in \Phi^{nc}\langle L\rangle$?" by the superset query "is $\Phi^{nc}\langle L\rangle \subseteq \Phi^{nc}\langle L_i\rangle$?" and reverse the answer. Hence, membership queries can be simulated and $\Phi^{nc}\langle\mathbf{I}\rangle$ can be learnt by using superset queries alone. As learnability from superset queries is invariant under translations, $\mathbf{I}$ can also be learnt from superset queries alone. $\square$

One has an analogous result for subset queries, but considering all translations rather than all text-preserving translations of the class, thanks to a (non text-preserving) automatic $\mathbf{I}$-translator Φ that satisfies $\Phi_{\mathbf{I}}\langle L\rangle = \{i \in I : L_i \subseteq L\}$ for all languages L. Indeed a membership query of the form "is $i \in \Phi_{\mathbf{I}}\langle L\rangle$?" is then equivalent to the subset query "is $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L\rangle$?":

Theorem 44. *If every translation of an automatic class* $\mathbf{I}$ *is learnable from membership and subset queries, then* $\mathbf{I}$ *itself is learnable from subset queries only.*

In the previous result, restriction to text-preserving translations is impossible:

Theorem 45. *Let* $\mathbf{I}$ *be the automatic class* $\big\{\varnothing\big\} \cup \big\{\{0,1\}^\star \setminus \{x\} : x \in \{0,1\}^\star\big\}$.

1. *Every text-preserving translation of* $\mathbf{I}$ *is learnable using membership and subset queries.*
2. *Some translation of* $\mathbf{I}$ *is not learnable using membership queries only.*
3. $\mathbf{I}$ *is not learnable using subset queries only.*

Proof. Given an automatic **I**-translator Φ, the translation $\Phi\langle\mathbf{I}\rangle$ can be learnt from membership queries and subset queries as follows. There is a finite subset S of $\{0,1\}^\star \setminus \{0\}$ such that $\Phi_{\mathbf{I}}\langle S\rangle$ contains an element y outside $\Phi_{\mathbf{I}}\langle\varnothing\rangle$. Now for every $x \notin S$, y belongs to $\Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{x\}\rangle$. Hence, a $\Phi\langle\mathbf{I}\rangle$-query learner can first use the membership query "is $y \in \Phi_{\mathbf{I}}\langle L\rangle$?". If the answer is "yes", then the query learner goes on querying whether $\Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{x\}\rangle \subseteq L$ until the answer is again "yes" for some x, and then the correct language is found. If the answer is "no" then the query learner knows that $\Phi_{\mathbf{I}}\langle L\rangle$ is either $\Phi_{\mathbf{I}}\langle\varnothing\rangle$ or $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\setminus\{x\}\rangle$ for one of the finitely many members x of S, and these finitely many cases can be distinguished using membership queries.

For the second item, consider an automatic **I**-translator Φ such that $\Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{x\}\rangle = \{x\}$ and $\Phi_{\mathbf{I}}\langle\varnothing\rangle = \varnothing$. In order to learn $\varnothing$ from queries only, a $\Phi\langle\mathbf{I}\rangle$-query learner M can make only finitely many membership queries before it concludes that $\varnothing$ is the language L to be learnt. The answers to these queries are consistent with L being one of infinitely many singletons (the individuals whose membership to the target language has been queried are excluded) rather than $\varnothing$. Hence, M cannot learn $\Phi\langle\mathbf{I}\rangle$.

Finally, **I** cannot be learnt from subset queries only: if finitely many queries of the form "is $\{0,1\}^\star \setminus \{x\} \subseteq L$?" have all been answered negatively, then an **I**-query learner still does not know whether $L = \varnothing$ or whether $L = \{0,1\}^\star \setminus \{y\}$ for some y such that no corresponding query has been made yet. $\square$

We end this section with a characterisation of the automatic classes all of whose translations are learnable from membership queries.

Theorem 46. *Given automatic class $\mathbf{I} = (L_i)_{i\in I}$, every translation of $\mathbf{I}$ is learnable from membership queries iff*

$$(\forall i)(\exists b_i)(\forall j \neq i)(\exists k \leq_{ll} b_i)\big[(L_j \subseteq L_k \wedge L_i \nsubseteq L_k) \vee (L_k \subseteq L_j \wedge L_k \nsubseteq L_i)\big].$$

Proof. Assume that the condition of the theorem holds. We exhibit a query learner M for **I** that uses membership queries. Since translations of automatic classes preserve inclusion between languages, we have that for all **I**-translators Φ, the condition of the theorem also holds for $\Phi\langle\mathbf{I}\rangle$, and M can be modified into a query learner for $\Phi\langle\mathbf{I}\rangle$ that uses membership queries.

Let M ask membership queries for individuals taken in length lexicographic order until it finds the $\leq_{ll}$-minimal $i \in I$ such that L_i is consistent with the answers to the queries asked so far and for all $k \leq_{ll} b_i$, both the following conditions hold:

- If $L_i \nsubseteq L_k$ then M got the answer "Yes" to some query of the form "Is $x \in L$" with $x \notin L_k$;
- If $L_k \nsubseteq L_i$ then M got the answer "No" to some query of the form "Is $x \in L$" with $x \in L_k$.

Then, M outputs i. By the assumed condition, M is well defined. To see that M learns **I**, let $i_0 \in I$ be given and assume that L_{i_0} is the language to be learnt. Consider any $j \in I \setminus \{i_0\}$. Let $k \in I$, with $k \leq b_j$, be such that

$$\big[(L_{i_0} \subseteq L_k \wedge L_j \nsubseteq L_k) \vee (L_k \subseteq L_{i_0} \wedge L_k \nsubseteq L_j)\big].$$

If $L_{i_0} \subseteq L_k$ and $L_j \nsubseteq L_k$, then M cannot find an $x \in L$ such that $x \notin L_k$; if $L_k \subseteq L_{i_0}$ and $L_k \nsubseteq L_j$, then M cannot find an $x \notin L$, with $x \in L_k$. Thus, based on the definition of M above, M will not eventually conjecture j. Furthermore, as the requirements above are eventually satisfied for $i = i_0$, M eventually does conjecture i_0. Thus, M learns $\mathbf{I}$.

For the converse, suppose that the condition of the theorem does not hold. Let $i \in I$ be such that for all $b_i \in I$, it is not true that

$$\forall j \neq i\, \exists k \leq_{ll} b_i \left[(L_j \subseteq L_k \wedge L_i \nsubseteq L_k) \vee (L_k \subseteq L_j \wedge L_k \nsubseteq L_i)\right].$$

Consider a first-order formula Φ (with x as unique free variable and parameters X for the input language and Y_r for the r-th element L_r of the indexing) which expresses that one of the following conditions holds:

1. x is of the form (α, β) for $\alpha, \beta \in I$ with $Y_\alpha \subseteq X$ and $\alpha \leq_{ll} \beta$;
2. there exists a $\leq_{ll}$-least member j of I such that $X \subseteq Y_j$ and $Y_i \setminus Y_j \neq \varnothing$, and x is of the form (α, β) for $\alpha, \beta \in I$ with $\alpha \leq_{ll} \beta \leq_{ll} j$ and $Y_\alpha \subseteq Y_i$;
3. there exists no member j of I such that $X \subseteq Y_j$ and $Y_i \setminus Y_j \neq \varnothing$, and x is of the form (α, β) for $\alpha, \beta \in I$ with $\alpha \leq_{ll} \beta$ and $Y_\alpha \subseteq Y_i$.

Note that for all members j, k of I, if $L_j \nsubseteq L_k$ then $\Phi_{\mathbf{I}}\langle L_k \rangle$ contains no pair of the form (j, β) with $\beta >_{ll} k$, and, hence, L_k contains only finitely many elements of the form (j, β). This implies that the second item in Definition 4 holds, and the first item easily follows from the definition of Φ.

For a contradiction, assume that M is a query learner for $\Phi\langle \mathbf{I} \rangle$. Let $b_i \in I$ be such that M makes membership queries only about elements (α, β) with α, β $\leq_{ll}$-smaller than b_i when learning $\Phi_{\mathbf{I}}\langle L_i \rangle$. Let $j \in I \setminus \{i\}$ be such that

$$\forall k \leq_{ll} b_i \left[(L_j \nsubseteq L_k \vee L_i \subseteq L_k) \wedge (L_k \nsubseteq L_j \vee L_k \subseteq L_i)\right]$$

holds. We claim that $\Phi_{\mathbf{I}}\langle L_j \rangle$ agrees with $\Phi_{\mathbf{I}}\langle L_i \rangle$ on all elements $\leq_{ll}$-smaller than b_i, and, hence, cannot be distinguished from $\Phi_{\mathbf{I}}\langle L_i \rangle$ by M, contrary to the assumption that M learns $\Phi\langle \mathbf{I} \rangle$. First, $\Phi_{\mathbf{I}}\langle L_j \rangle \setminus \Phi_{\mathbf{I}}\langle L_i \rangle$ contains no element of the form (α, β) with $\alpha \leq_{ll} b_i$ and $\beta \leq_{ll} b_i$: indeed, only 1. above in the definition of Φ could introduce such an element; however, no $\alpha \leq_{ll} b_i$ satisfies $L_\alpha \subseteq L_j$ but $L_\alpha \nsubseteq L_i$, and thus there is no such element. Second, consider a member (α, β) of $\Phi_{\mathbf{I}}\langle L_i \rangle$ with α, β $\leq_{ll}$-smaller than b_i. By 2. above in the definition of Φ, $\Phi_{\mathbf{I}}\langle L_j \rangle$ also contains (α, β) as otherwise, there would exist a $k \in I$ with $k \leq_{ll} b_i$ such that $L_j \subseteq L_k$ and $L_i \nsubseteq L_k$. Hence, $\Phi_{\mathbf{I}}\langle L_j \rangle$ agrees with $\Phi_{\mathbf{I}}\langle L_i \rangle$ on all elements $\leq_{ll}$-smaller than b_i, as needed. □

12 Conclusion

A notion of learnability is robust if it is immune to natural transformations of the class of objects to be learned. The associated notion of transformation of languages has been defined as a function, called a translator, that maps languages to languages and preserves the inclusion structure of the languages in the original class. Our study has focused on automatic classes of languages, as automaticity is invariant under translation and as this restriction allows one to obtain appealing characterisations of robust learning under many classical learning criteria, namely the following: consistent and conservative learning, strong-monotonic learning, strong-monotonic consistent learning, finite learning, learning from subset queries, learning from superset queries and learning from membership queries. The characterisations are natural as they express a particular constraint on the inclusion structure of the original class. In many cases, they are especially strong as they also deal with learnability under those translations that are text-preserving, in that they can be generated from an enumeration of a language without necessitating the latter to be "seen as a whole." In some of the characterisations, learning from every translation turned out to be equivalent to learning from every text-preserving translation: Theorem 17 (standard learnability), Theorem 20 (strong-monotonic learnability) and Theorem 24 (strong-monotonic and consistent learnability). Though there are some similarities in the proofs, we do not know of a general characterisation of learning criteria for which such a result applies. A further open question is in relation to confident learning: we found a characterisation for nonrecursive learners, but none for recursive ones. Also, it would be interesting for further work to address complexity issues, in particular in the context of learning from queries.

References

1. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2), pp. 117–135, 1980.
2. Dana Angluin. Learning regular sets from queries and counterexamples. *Information and Computation*, 75, pp. 87–106, 1987.
3. Janis Bārzdiņš. Inductive inference of automata, functions and programs. In *Proceedings of the 20th International Congress of Mathematicians, Vancouver*, pages 455–560, 1974. In Russian. English translation in American Mathematical Society Translations: Series 2, 109:107-112, 1977.
4. Mark Fulk. Robust separations in inductive inference. *Proceedings of the 31st Annual Symposium on Foundations of Computer Science* (FOCS 1990), pp. 405–410, 1990.
5. E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5), pp. 447–474, 1967.
6. Bernard R. Hodgson. Décidabilité par automate fini. *Annales des sciences mathématiques du Québec*, 7(1), pp. 39–57, 1983.
7. Sanjay Jain, Qinglong Luo and Frank Stephan. Learnability of automatic classes. *Journal of Computer and System Sciences*, 78(6), pp. 1910–1927, 2012.
8. Sanjay Jain, Yuh Shin Ong, Shi Pu and Frank Stephan. *On automatic families.* In T. Arai, Q. Feng, B. Kim, G. Wu and Y. Yang, *Proceedings of the 11th Asian Logic Conference*, (ALC 2009), pp. 94–113. World Scientific, 2011.
9. Sanjay Jain, Daniel Osherson, James S. Royer and Arun Sharma. *Systems That Learn*, 2nd Edition. MIT Press, 1999.

10. Sanjay Jain and Frank Stephan. A tour of robust learning. In *Computability and Models.* Perspectives East and West. Kluwer Academic / Plenum Publishers, pp. 215–247, 2003.
11. Sanjay Jain, Carl H. Smith and Rolf Wiehagen. Robust learning is rich. *Journal of Computer and System Sciences*, 62(1), pp. 178–212, 2001.
12. Klaus P. Jantke. Monotonic and non-monotonic inductive inference. *New Generation Computing* 8, pp. 349–360, 1991.
13. Daniel N. Osherson and Scott Weinstein. Criteria of language learning. *Information and Control* 52, pp. 123–138, 1982.
14. Bakhadyr Khoussainov and Anil Nerode. Automatic presentations of structures. In Daniel Leivant, editor. *Selected Papers from Logic and Computational Complexity* (LCC 1994), pp. 367–392, 1994.
15. Bakhadyr Khoussainov and Sasha Rubin. Automatic structures: Overview and future directions. *Journal of Automata, Languages and Combinatorics*, 8(2), pp. 287–301, 2003.
16. Steffen Lange, Jochen Nessel and Sandra Zilles. Learning languages with queries. Proceedings of the FGML Meeting 2002, pp. 92–99, 2002.
17. Steffen Lange and Thomas Zeugmann. Language learning in dependence on the space of hypotheses. *Proceedings of the Sixth Annual Conference on Computational Learning Theory* (COLT 1993), pp. 127–136. ACM Press, 1993.
18. Steffen Lange, Thomas Zeugmann and Sandra Zilles. Learning indexed families of recursive languages from positive data: a survey. *Theoretical Computer Science*, 397, pp. 194–232, 2008.
19. Steffen Lange and Sandra Zilles. Formal language identification: Query learning vs. Gold-style learning. *Information Processing Letters*, 91(6), pp. 285–292, 2004.
20. Matthias Ott and Frank Stephan. Avoiding coding tricks by hyperrobust learning. *Theoretical Computer Science*, 284(1), pp. 161–180, 2002.
21. Michael Rabin. *Automata on Infinite Objects and Church's Problem.* AMS, 1972.
22. Herman Weyl. *Symmetry.* Princeton University Press, 1952.
23. Rolf Wiehagen and Thomas Zeugmann. Learning and consistency. *Algorithmic Learning for Knowledge-Based Systems*, LNAI 961, pp. 1–24, Springer, 1995.
24. Thomas Zeugmann. On Bārzdiņš' Conjecture. Analogical and Inductive Inference (AII 1986), Proceedings of the International Workshop, LNCS 265, pp. 220–227, Springer, 1986.