Mitotic Classes in Inductive Inference

Sanjay Jain[⋆1] and Frank Stephan[∗2]

[1] Department of Computer Science,
National University of Singapore, Singapore 117590, Republic of Singapore.
sanjay@comp.nus.edu.sg
[2] Department of Mathematics and Department of Computer Science,
National University of Singapore, Singapore 117543, Republic of Singapore.
fstephan@comp.nus.edu.sg

Abstract. For the natural notion of splitting classes into two disjoint subclasses via a recursive classifier working on texts, the question is addressed how these splittings can look in the case of learnable classes. Here the strength of the classes is compared using the strong and weak reducibility from intrinsic complexity. It is shown that, for explanatorily learnable classes, the complete classes are also mitotic with respect to weak and strong reducibility, respectively. But there is a weakly complete class that cannot be split into two classes which are of the same complexity with respect to strong reducibility. It is shown that for complete classes for behaviourally correct learning, one half of each splitting is complete for this learning notion as well. Furthermore, it is shown that explanatorily learnable and recursively enumerable classes always have a splitting into two incomparable classes; this gives an inductive inference counterpart of Sacks Splitting Theorem from recursion theory.

1 Introduction

A well-known observation is that infinite sets can be split into two parts of the same cardinality as the original set, while finite sets cannot be split in such a way; for example, the integers can be split into the sets of the even and odd numbers while splitting a set of 5 elements would result in subsets of unequal sizes. In this sense, infinite sets are better than finite ones. The corresponding question in complexity and recursion theory is the following: which sets can be split into two sets of the same complexity as the original set [1, 9, 10, 15].

Ambos-Spies [1] defined one of the variants of mitoticity using many-one reducibilities. Here a set A is many-one reducible to a set B iff there is a recursive function f such that $A(x) = B(f(x))$. That is, one translates every input x for A into an input $f(x)$ for B, takes the solution provided by B (in the set or in its complement) and then copies this to obtain the solution for A. Similarly one considers also complexity-theoretic counterparts of many-one reductions; for example one can translate an instance (G_1, G_2) of the Graph-Isomorphism problem into an instance ϕ of the Satisfiability problem in polynomial time, where G_1 is isomorphic to G_2 iff ϕ is satisfiable. Here, one can choose the reduction such that one does not only test membership, but can also

⋆ Supported in part by NUS grant numbers R252-000-212-112 and R252-000-308-112.

translate a solution of ϕ into an isomorphism between G_1 and G_2 whenever such a solution exists for ϕ. Indeed, NP-complete problems are characterized as those into which every NP problem can be translated. This general method of reducing problems and translating solutions (although here the translation of the solution is just the identity) occurs quite frequently in other fields of mathematics. In inductive inference, intrinsic complexity is based on the notion of reducing one learning problem $\mathcal{L}$ to another problem $\mathcal{H}$: first an operator translates a text T for a set L in $\mathcal{L}$ into a text $\Theta(T)$ for a set H in $\mathcal{H}$ and then another operator translates a solution E, which is a sequence converging to an index e of H, into a solution for L given as a sequence converging to an index e' of L. Before explaining this in more detail, some terminology is necessary to make it precise.

- A partial recursive function is a partial function computed by a Turing machine, where the machine does not halt on inputs on which the function is undefined. A recursively enumerable set is the domain (or equivalently the range) of a partial recursive function. There is an acceptable numbering $W_0, W_1, W_2, \ldots$ of all recursively enumerable sets [16, Section II.5]; this numbering will be kept fixed from now on.
- A general recursive operator Θ is a mapping from total functions to total functions such that there is a recursively enumerable set E of triples which satisfies the following: for every total function f and every x, y, $\Theta(f)(x) = y$ iff there is an n such that $(f(0)f(1)\ldots f(n), x, y) \in E$.
- A language is a recursively enumerable subset of the natural numbers.
- A class $\mathcal{L}$ is a set of languages. A family $L_0, L_1, L_2, \ldots$ is an indexing for $\mathcal{L}$ iff $\{(e, x) : x \in L_e\}$ is recursively enumerable and $\mathcal{L} = \{L_0, L_1, L_2, \ldots\}$.
- A text T (see [11]) is a mapping from the set $\mathbb{N}$ of natural numbers to $\mathbb{N} \cup \{\#\}$. Content of a text T, content(T), is the set $\{T(n) \mid n \in \mathbb{N} \wedge T(n) \in \mathbb{N}\}$. T is a text for L iff content$(T) = L$. $T[n]$ denotes the first n elements of the sequence T, that is, $T[n] = T(0)\,T(1)\,\ldots\,T(n-1)$. Furthermore, content$(T[n])$ denotes the set of natural numbers occurring in $T(0)\,T(1)\,\ldots\,T(n-1)$. As an example, $T = \#\#3\#\#8\#8\#\#7\#9\#^\infty$ is a text for the set $\{3, 7, 8, 9\}$, $T(0) = \#$, $T(2) = 3$ and $T[6] = \#\#3\#\#8$.
- A learner is a general recursive operator (see [22]) which translates T into another sequence E. The learner converges on T iff there is a single e such that $E(n) = e$ for almost all n — in this case one says that the learner converges on T to e. The learner explanatorily learns (see [11, 5]) T iff it converges on T to some index e such that $W_e =$ content(T). A learner explanatorily learns L iff it explanatorily learns every text for L. A learner explanatorily learns $\mathcal{L}$ iff it explanatorily learns every $L \in \mathcal{L}$. Note that, in some cases, learning algorithms can also be described such that they use the indices from a given indexing for $\mathcal{L}$; such indices can always be translated into indices of the acceptable numbering $W_0, W_1, \ldots$ of all recursively enumerable sets.
- A classifier [21] is a general recursive operator which translates texts to sequences over $\{0, 1\}$. A classifier C converges on a text T to a iff $C(T[n]) = a$ for almost all n.
- For the learning criteria considered in this paper, one can assume without loss of generality [17] that the learner M computes $E(n)$ from input $T[n]$. $M(T[n])$ denotes this hypothesis. A similar convention holds for classifiers.

As an example, the class $\mathcal{D}$ of all finite sets is explanatorily learnable. The following learner explanatorily learns $\mathcal{D}$: On every finite sequence σ it outputs a canonical index for content(σ), so $M(\#\#3\#\#8\#8)$ outputs an index for $\{3,8\}$. Then it is easy to see that M converges on every text for a finite set to the canonical index for this set. An example for a non explanatorily learnable class was obtained by Gold [11] who showed that $\mathcal{D}\cup\{L\}$ is not explanatorily learnable for any infinite set L. Let $\mathcal{D}_{even}$ and $\mathcal{D}_{odd}$ be the classes of finite sets with even and odd number of elements, respectively. Then one can easily construct a classifier which separates $\mathcal{D}_{odd}$ from $\mathcal{D}_{even}$ by defining that $M(\sigma) = 1$ if content(σ) has an even number of elements and $M(\sigma) = 0$ otherwise.

Freivalds, Kinber and Smith [6] consider reductions between learnability problems for function classes. Jain and Sharma [13] carried this idea over to the field of learning languages from positive data and formalized the following two reducibilities for learnability problems. The main difference between these two reducibilities is that Θ can be one-to-many in the case of the weak reducibility, as different texts for the same language can be mapped to texts for different languages, while for the strong reducibility this is not allowed, at least for languages in the given class.

- A class $\mathcal{L}$ is weakly reducible to $\mathcal{H}$ iff there are general recursive operators Θ and Ψ such that
  - Whenever T is a text for a language in $\mathcal{L}$, $\Theta(T)$ is a text for a language in $\mathcal{H}$;
  - Whenever E is a sequence which converges to an index e with $W_e = \text{content}(\Theta(T))$ for some text T of a language in $\mathcal{L}$, $\Psi(E)$ is a sequence converging to an e' with $W_{e'} = \text{content}(T)$.

  One writes $\mathcal{L} \leq_{weak} \mathcal{H}$ in this case.
- A class $\mathcal{L}$ is strongly reducible to $\mathcal{H}$ iff there are general recursive operators Θ, Ψ as above with the following additional constraint. Whenever T, T' are texts for the same language in $\mathcal{L}$, $\Theta(T), \Theta(T')$ are texts for the same language in $\mathcal{H}$. One writes $\mathcal{L} \leq_{strong} \mathcal{H}$ in this case. Furthermore, $\Theta(L)$ denotes the language content$(\Theta(T))$, where T is a text for L.

For example, the class $\{\{x\} : x \in \mathbb{N}\}$ of all singleton sets is strongly reducible to $\mathcal{D}$ but $\mathcal{D}$ is not weakly reducible to the class of all singleton sets [13]. Furthermore, $\mathcal{D}_{even}$ and $\mathcal{D}_{odd}$ are strongly reducible to each other by the same mapping (Θ, Ψ). The operator Θ translates every text for a set D to a text for $\{0\} \cup \{d+1 : d \in D\}$; the operator Ψ translates an index e, in a sequence of hypotheses, to an index $g(e)$ of the set $W_{g(e)} = \{x : x+1 \in W_e\}$.

Jain, Kinber, Sharma and Wiehagen investigated these concepts in several papers [12–14]. They found that there are complete classes with respect to $\leq_{weak}$ and $\leq_{strong}$. Here a class $\mathcal{H}$ is complete with respect to $\leq_{weak}$ ($\leq_{strong}$) iff $\mathcal{H}$ can be explanatorily learned and for every explanatorily learnable class $\mathcal{L}$ it holds that $\mathcal{L} \leq_{weak} \mathcal{H}$ ($\mathcal{L} \leq_{strong} \mathcal{H}$). If $\sqsubset$ is a recursive dense linear ordering on $\mathbb{N}$ without least and greatest element (which makes $\mathbb{N}$ an order-isomorphic copy of the rationals) then

$$\mathcal{Q} = \{\, \{y \in \mathbb{N} \mid y \sqsubseteq x\} \mid x \in \mathbb{N}\}$$

is a class which is complete for both, $\leq_{weak}$ and $\leq_{strong}$ (see [12]). The following classes are complete for $\leq_{weak}$ but not for $\leq_{strong}$ (see [13]):

$$\mathcal{I} = \{\, \{0, 1, \ldots, x\} \mid x \in \mathbb{N}\};$$

$$\mathcal{CS} = \{\, \mathbb{N} - \{x\} \mid x \in \mathbb{N}\}$$

If one looks at $\mathcal{CS}$, one can easily see that it is the disjoint union of two classes of equivalent intrinsic complexity, namely, the class $\{\, \mathbb{N} - \{x\} \mid x \text{ is even}\}$ and $\{\, \mathbb{N} - \{x\} \mid x \text{ is odd}\}$. All three classes can be translated into each other and a classifier can witness the splitting: if T is a text for a member of $\mathcal{CS}$, then the classifier converges in the limit to the remainder of x divided by 2 for the unique $x \notin \text{content}(T)$. This type of splitting can be formalized to the notion of a mitotic class.

Definition 1. Two infinite classes $\mathcal{L}_0$ and $\mathcal{L}_1$ are called a *splitting* of a class $\mathcal{L}$ iff $\mathcal{L}_0 \cup \mathcal{L}_1 = \mathcal{L}$, $\mathcal{L}_0 \cap \mathcal{L}_1 = \emptyset$ and there exists a classifier C such that, for all $a \in \{0,1\}$ and for all texts T with $\text{content}(T) \in \mathcal{L}_a$, C converges on T to a.

A class $\mathcal{L}$ is *strongly mitotic* (*weakly mitotic*) iff there is a splitting $\mathcal{L}_0, \mathcal{L}_1$ of $\mathcal{L}$ such that $\mathcal{L} \equiv_{strong} \mathcal{L}_0 \equiv_{strong} \mathcal{L}_1$ ($\mathcal{L} \equiv_{weak} \mathcal{L}_0 \equiv_{weak} \mathcal{L}_1$).

The study of such notions is motivated by recursion theory [16, 22] where a recursively enumerable set is called mitotic iff it is the disjoint union of two other recursively enumerable sets which have the same Turing degree. The importance of this notion is reflected by Ladner's result that a recursively enumerable set is mitotic iff it is autoreducible, that is, iff there is an oracle Turing machine M such that $A(x) = M^{A \cup \{x\}}(x)$ for all x [15]. Furthermore the notion had been carried over to complexity theory where it is still an important research topic [1, 9, 10]. Thus, we feel it is interesting to explore this notion in the context of inductive inference too. The results show some interesting properties related to mitoticity. In particular classes complete for explanatory learning with respect to $\leq_{strong}$ ($\leq_{weak}$) are strongly (weakly) mitotic. Thus, nonmitotic classes cannot be complete.

Although intrinsic complexity is not the exact counterpart of Turing degrees in recursion theory, it is the only type of complexity which is defined via reducibilities and not via measures such as counting mind changes or the size of long term memory in inductive inference. Therefore, from the viewpoint of inductive inference, the above defined version of mitotic classes is reasonable. Indeed, there are some obvious parallels: in recursion theory, any recursively enumerable cylinder is mitotic, where a cylinder A is a set of the form $\{(x,y) \mid x \in B, y \in \mathbb{N}\}$ for some set $B \subseteq \mathbb{N}$. A corresponding cylindrification of a class $\mathcal{L}$ would be the class

$$\{\, \{(x,y) \mid y \in L\} \mid x \in \mathbb{N}, L \in \mathcal{L}\}.$$

It can easily be seen that this class is strongly mitotic and thus also weakly mitotic. Indeed, two constraints are placed in Definition 1 in order to be as near to the original definition of mitoticity in recursion theory as possible:

- In recursion theoretic setting, if A is split into two recursively enumerable sets A_0, A_1 with $A_0 \equiv_T A_1$, then $A \equiv_T A_0 \equiv_T A_1$. Thus, in the definition of mitoticity for learning theory, it is required that all three classes, class $\mathcal{L}$ and its two halves $\mathcal{L}_1$ and $\mathcal{L}_2$, have the same intrinsic complexity degree.

– If A is recursively enumerable and mitotic and split into A_0, A_1, then there is a partial recursive function with domain A mapping the elements of A_a to a for all $a \in \{0, 1\}$. For mitotic classes of languages, the corresponding function is a classifier which works correctly on all texts for the languages in the class. It is not required that the classifier converges on every text, as then many naturally strongly mitotic classes, such as $\mathcal{CS}$, would no longer be mitotic. This has a parallel in recursion theory: if one splits a maximal set (as defined in Remark 8 below) into two recursively enumerable sets A_0 and A_1, which are both not recursive, then the sets A_0 and A_1 are recursively inseparable.

Besides the reducibilities $\leq_{weak}$ and $\leq_{strong}$ considered here, other reducibilities have also been considered [12, 13]. This paper deals only with $\leq_{weak}$ and $\leq_{strong}$, as these two are the most natural and representative.

One emphasis of the current work is on the search for natural classes which split or do not split. Therefore it is always required that the class under consideration is learnable (under the criterion in consideration). Furthermore, one tries to show properties for complete classes, recursively enumerable classes and indexed families. Angluin [2] defined that $\{L_0, L_1, L_2, \ldots\}$ is an indexed family iff the function $e, x \mapsto L_e(x)$ is recursive. For indexed families $\{L_0, L_1, L_2, \ldots\}$ one can assume, without loss of generality, that $L_n \neq L_m$ whenever $n \neq m$. A learner for this family is called exact iff it converges on every text for L_n to n.

In this paper, it is shown that classes complete for explanatory learning with respect to $\leq_{strong}$ ($\leq_{weak}$) are strongly (weakly) mitotic. Furthermore, there are classes complete for explanatory learning with respect to $\leq_{weak}$ which are not strongly mitotic. Similar phenomena are studied for behaviourally correct learning. A counterpart to Sacks splitting theorem is that every infinite recursively enumerable explanatory learnable class has a splitting into two incomparable subclasses. The relations between autoreducibility and mitoticity are also investigated. However, in contrast to the situation in complexity theory, these two notions turn out to be different.

The following remark is important for several proofs.

Remark 2. One says that a learner M or a classifier C converges on T to a value a iff $M(T[n]) = a$ or $C(T[n]) = a$ for almost all n, respectively. But it does not matter – in the framework of inductive inference – how fast this convergence is; the machine can be slowed down by starting with an arbitrary guess and later repeating hypotheses, if needed. Similarly, if one translates one text of a language L into a text of a language H, it is not important how fast the symbols of H show up in the translated text, it is only important that they show up eventually. Therefore the translator can put into the translated text pause symbol, #, until more data are available or certain simulated computations have terminated.

Therefore, learners, operators translating texts and classifiers can be made primitive recursive by the just mentioned delaying techniques. Thus one can have recursive enumerations $\Theta_0, \Theta_1, \Theta_2, \ldots$ of translators from texts to texts, $M_0, M_1, M_2, \ldots$ of learners and $C_0, C_1, C_2, \ldots$ of classifiers such that, for every given translator, learner or classifier, this list contains an equivalent one. These lists can be used in proofs where diagonalizations are needed.

Given a text T and a number n, one denotes by $\Theta(T[n])$ the initial part $\Theta(T)[m]$ for the largest $m \leq n$ such that $\Theta(T)[m]$ is produced without accessing any datum in T beyond the n-th position. Note that, for every m, there is an n such that $\Theta(T[n])$ extends $\Theta(T)[m]$ and $\Theta(T[n])$ can be computed from $T[n]$.

2 Complete Classes

The two main results are that classes which are complete for $\leq_{strong}$ are strongly mitotic and classes which are complete for $\leq_{weak}$ are weakly mitotic. This stands in contrast to the situation in recursion theory where some Turing-complete recursively enumerable sets are not mitotic [15]. Note that certain classes which are complete only for $\leq_{weak}$ fail to be strongly mitotic; thus the main results cannot be improved.

Theorem 3. *Every class which is complete for $\leq_{strong}$ is also strongly mitotic.*

Proof. Let $\mathcal{L}$ and $\mathcal{H}$ be any classes which are complete for $\leq_{strong}$. Then the class $\mathcal{K}$ consisting of the sets $I = \{1,3,5,7,\ldots\}$, $J = \{0\} \cup I$, $\{2x+3 : x \in H\}$ and $J \cup \{2x+2 : x \in H\}$ for every $H \in \mathcal{H}$ is also complete for $\leq_{strong}$. Since $\mathcal{L}$ is complete for $\leq_{strong}$, there is a translation Θ which maps languages in $\mathcal{K}$ to languages in $\mathcal{L}$ such that proper inclusion is preserved [14]. Thus there is some $e \in \Theta(J) - \Theta(I)$. As $\mathcal{H}$ is complete for $\leq_{strong}$, the subclasses $\{\Theta(\{2x+3 : x \in H\}) : H \in \mathcal{H}\}$ and $\{\Theta(J \cup \{2x+2 : x \in H\}) : H \in \mathcal{H}\}$ of $\mathcal{L}$ are also complete for $\leq_{strong}$. All members of the first class do not contain e while all members of the second class contain e as an element. It follows that the subclasses $\mathcal{L}_0 = \{L \in \mathcal{L} : e \notin L\}$ and $\mathcal{L}_1 = \{L \in \mathcal{L} : e \in L\}$ are disjoint and complete for $\leq_{strong}$. Thus $\mathcal{L}$ can be classified by a classifier C, which conjectures 1 if e has shown up in the text so far and 0 otherwise. Therefore, $\mathcal{L}$ is strongly mitotic. □

The following notion is used to formulate Proposition 6 which is a central ingredient of Theorem 7. Furthermore, learners with certain properties are needed.

Definition 4. For any sequence T of symbols, let $\mathrm{all}(T)$ be the length of the shortest prefix of T containing all symbols which show up in T, that is, let

$$\mathrm{all}(T) = \sup\{n+1 : \mathrm{content}(T[n]) \subset \mathrm{content}(T)\}.$$

Note that $\mathrm{all}(T) < \infty$ iff $\mathrm{content}(T)$ is a finite set.

The following remark combines some ideas of Blum and Blum [4] and Fulk [8].

Remark 5. Let $\mathcal{L}$ be an explanatorily learnable class. Then there is an explanatory learner M for $\mathcal{L}$ with the following properties:

- M is prudent [17], that is, whenever M outputs an index e on some input data, then M explanatorily learns W_e;
- M is order independent [4], that is, for every set L, either M diverges on all texts for L or M converges on all texts for L to the same index;
- for every text T and index e, if $M(T[n]) = e$ for infinitely many n, then $M(T[n]) = e$ for almost all n.

Proposition 6. *Suppose $\mathcal{I} \leq_{weak} \mathcal{L}$ and M is a explanatory learner for $\mathcal{L}$ which satisfies the conditions in Remark 5. Then there is a reduction (Θ, Ψ) from $\mathcal{I}$ to $\mathcal{L}$ such that, for all texts T for a language in $\mathcal{I}$, M converges on $\Theta(T)$ to an index $e >$ all(T).*

Proof. Assume that (Θ'', Ψ'') witnesses that $\mathcal{I} \leq_{weak} \mathcal{L}$. Now a reduction (Θ', Ψ') from $\mathcal{I}$ to $\mathcal{I}$ is constructed such that Θ can be taken to be the composition of Θ' and Θ''.

The key idea for this is the following: one constructs (Θ', Ψ') from $\mathcal{I}$ to $\mathcal{I}$ such that, for every text T of a set in $\mathcal{I}$, M converges on $\Theta''(\Theta'(T))$ to an index $e > \text{all}(T)$. Note that all(T) is finite for all texts for members of $\mathcal{I}$. By Remark 2, assume without loss of generality that Θ'' is primitive recursive. The idea is that Θ' translates a text T for $I_n = \{0, 1, \ldots, n\}$ to a text for $I_{2^n(1+2m)}$ for some m; Ψ' translates any sequence converging to an index of the set $I_{2^n(1+2m)}$ into a sequence converging to an index of I_n.

Given a sequence E of indices, $\Psi'(E)(s)$ is computed as follows. Let k be the least number such that $W_{E(s),s} \subseteq I_k$. Choose m, n such that $2^n(1+2m) = k$ and output the canonical index for I_n. It is easy to see that this translation works whenever E converges to an index of some set in $\mathcal{I}$.

Construction of $\mathbf{\Theta'}$. The construction of Θ' is more involved. For the construction, the special properties of M from Remark 5 are important. The most adequate way to describe $\Theta'(T)$ is to build longer and longer finite prefixes $\tau_0, \tau_1, \tau_2, \ldots$ of this target $\Theta'(T)$. The construction starts with $\tau_0 = 0\#$ and in stage s, the extension τ_{s+1} of τ_s is defined according to the first case which applies:

- Case 1: $M(\Theta''(\tau_s)) \leq \text{all}(T[s])$. Then let τ_{s+1} be the first extension of τ_s found such that $M(\Theta''(\tau_{s+1})) \neq M(\Theta''(\tau_s))$;
- Case 2: Case 1 does not hold but content$(\tau_s) \neq I_{2^n(1+2m)}$ for all m, where n is the least number with content$(T[s]) \subseteq I_n$. Then let $\tau_{s+1} = \tau_s a$ for the least nonelement a of content(τ_s);
- Case 3: Case 1 and Case 2 do not hold. Then $\tau_{s+1} = \tau_s\#$.

Here $\Theta''(\tau_s)$ and $\Theta''(\tau_{s+1})$ are defined as in Remark 2 and can be computed from τ_s and τ_{s+1}, respectively.

Verification. For the verification, assume that a set $I_n = \{0, 1, 2, \ldots, n\} \in \mathcal{I}$ and a text T for I_n are given.

First note that, in Case 1 of the construction, the extension τ_{s+1} of τ_s can always be found. To see this, note that there are two texts T_1, T_2 extending τ_s for different sets in $\mathcal{I}$. It follows that $\Theta''(T_1)$ and $\Theta''(T_2)$ are texts of different sets and thus M converges on them to different indices. Thus one can take a sufficiently long prefix of one of T_1, T_2 in order to get the desired τ_{s+1}.

Second, it can be shown by induction that $|\tau_s| > s$ at all stages s; this guarantees that Θ' is indeed a general recursive operator.

Third, one shows that M does not converge on $\Theta''(\Theta'(T))$ to any index less than or equal to all(T). By Case 1 in the construction, M cannot converge on $\Theta''(\Theta'(T))$ to an index $e \leq \text{all}(T)$.

Thus, by Remark 5, there is a stage s_0 such that Case 1 of the construction is never taken after stage s_0 and content$(T[s_0]) = I_n$.

Fourth, one shows that $\Theta'(T)$ is a text for some language in $\mathcal{I}$. There is a least m such that content$(\tau_{s_0}) \subseteq I_{2^n(1+2m)}$. For all stages $s > s_0$, if content$(\tau_s) \subset I_{2^n(1+2m)}$, then τ_{s+1} is chosen by Case 2, else τ_{s+1} is chosen by Case 3. One can easily see that the resulting text $\Theta'(T) = \lim_{s\to\infty} \tau_s$ is a text for $I_{2^n(1+2m)}$. Indeed, $\Theta'(T) = \tau_{s_1}\#^\infty$ for $s_1 = s_0 + 2^n(1+2m) + 2$.

So it follows that Θ' maps every text of a set I_n to a text for some set $I_{2^n(1+2m)}$, as desired. So, for all texts T of sets in $\mathcal{I}$, M converges on $\Theta''(\Theta'(T))$ to some index $e >$ all(T). □

Theorem 7. *Let $\mathcal{L}$ be a explanatorily learnable class which is complete for $\leq_{weak}$. Then $\mathcal{L}$ is weakly mitotic.*

Proof. Let $I_n = \{0, 1, \ldots, n\}$. By Proposition 6 there is a reduction (Θ, Ψ) from $\mathcal{I}$ to $\mathcal{L}$ and a learner M such that M satisfies the conditions outlined in Remark 5 and, for every text T of a member of $\mathcal{I}$, M converges on $\Theta(T)$ to an index $e >$ all(T). For this reason, using oracle K for the halting problem, one can check, for every index e, whether there is a text T for a language in $\mathcal{I}$ such that M on $\Theta(T)$ converges to e. This can be seen as follows: One can assume, without loss of generality, that, besides #, no data-item in a text is repeated. Also, among the texts for sets in $\mathcal{I}$, only the texts of the sets $\{0, 1, 2, \ldots, d\}$ with $d < e - 1$ can satisfy all$(T) < e$. Thus, one has just to check the behaviour of the given explanatory learner M for the class $\mathcal{L}$ on the texts $\Theta(T')$, for T' in the class

$$\mathcal{T}_e = \{T' \mid T' \in \{d, \# : d < e - 1\}^e \cdot \#^\infty \wedge \text{content}(T) \in \mathcal{I}\}.$$

Now, define a classifier C such that on a text T, the n-th guess of C is 1 iff there is an odd number $m \leq M(T[n])$ and a text $T'' \in \mathcal{T}_{M(T[n])}$ for I_m such that $M(\Theta(T'')[n]) = M(T[n])$.

For the verification that C is a classifier, assume that M converges on T to some index e. Then C converges on T to 1 iff there is an odd number m and a text T' for I_m in $\mathcal{T}_e$ such that M converges on the texts T and $\Theta(T')$ to the same number. Otherwise C converges on T to 0. If M does not converge on T, then T is not a text for a set in $\mathcal{L}$ and the behaviour of C on T is irrelevant. Thus C is a classifier which splits $\mathcal{L}$ into two classes $\mathcal{L}_0$ and $\mathcal{L}_1$. These classes $\mathcal{L}_0$ and $\mathcal{L}_1$ contain the images of repetition-free texts of sets in the classes $\{I_0, I_2, I_4, \ldots\}$ and $\{I_1, I_3, I_5, \ldots\}$, respectively. Thus both classes are complete for $\leq_{weak}$ and the splitting of $\mathcal{L}$ into $\mathcal{L}_0$ and $\mathcal{L}_1$ witnesses that $\mathcal{L}$ is weakly mitotic. □

As several proofs use known properties of the maximal sets, the following remark summarizes some of these properties.

Remark 8. A set A is *maximal* (see [16]) iff (a) A is recursively enumerable, (b) A has an infinite complement and (c) every recursively enumerable set B satisfies that either $B - A$ is finite or the complement of $A \cup B$ is finite.

A maximal set is dense simple, that is, if $a_0, a_1, a_2, \ldots$ gives the complement in the ascending order and f is a recursive function, then $f(a_n) < a_{n+1}$ for almost all n.

For any partial recursive function ψ and any maximal set A, the following statements hold.

- Either $\psi(x)$ is defined for almost all $x \in \overline{A}$ or for only finitely many $x \in \overline{A}$.
- The set $\{x \notin A \mid \psi(x) \in A\}$ is either finite or contains almost all elements of $\overline{A}$.
- If, for every x, there is some $y > x$ such that $y \notin A$, $\psi(y)$ is defined, $\psi(y) > x$ and $\psi(y) \notin A$, then $\psi(z)$ is defined and $\psi(z) = z$ for almost all $z \in \overline{A}$.

These basic facts about the maximal sets will be used in several proofs. Odifreddi [16, Pages 288–294] provides more information on the maximal sets including the proof of their existence by Friedberg [7].

Theorem 9. *There exists an indexed family $\{L_0, L_1, L_2, \ldots\}$ which is weakly mitotic and complete for $\leq_{weak}$, but not strongly mitotic.*

Proof. Let A be a maximal set with complement $\{a_0, a_1, \ldots\}$ where $a_n < a_{n+1}$ for all n. Now let $\mathcal{L}$ consist of the sets

- $\{x, x+1, x+2, \ldots, x+y\}$ for all $x \in A$ and $y \in \mathbb{N}$;
- $\{x, x+1, x+2, \ldots\}$ for all $x \notin A$.

As A is recursively enumerable, $\mathcal{L}$ can be represented as an indexed family. Explanatory learnability is also clear as the learner, on input σ, first determines $x = \min(\text{content}(\sigma))$ and then conjectures content(σ), if $x \in A_{|\sigma|}$, and conjectures $\{x, x+1, x+2, \ldots\}$, otherwise. Without loss of generality, it can be assumed that $a_0 > 0$ and thus $\mathcal{L}$ is a superclass of $\mathcal{I}$ and therefore $\mathcal{L}$ is complete for $\leq_{weak}$. By Theorem 7, $\mathcal{L}$ is weakly mitotic.

Let $\mathcal{L}_0$ and $\mathcal{L}_1$ be two disjoint classes with union $\mathcal{L}$. Without loss of generality, $\{a_0, a_0+1, a_0+2, \ldots\} \in \mathcal{L}_1$. Assume now by way of contradiction that $\mathcal{L} \leq_{strong} \mathcal{L}_0$ as witnessed by (Θ, Ψ). As Θ has to preserve the proper subset relation on the content of the texts while translating, every text of a set of the form $\{a_n, a_n+1, a_n+2, \ldots\}$ has to be translated into a text for a set of the form $\{a_m, a_m+1, a_m+2, \ldots\}$ (to preserve the property that translation of $\{a_n, a_n+1, a_n+2, \ldots\}$ has infinitely many subsets in the class).

Now consider the function f which outputs, on input x, the first element found to be in the range of the image $\Theta(\sigma)$ for some σ with $x = \min(\text{content}(\sigma))$. The function f is recursive, but by Remark 8 and A being a maximal set, the relation $f(a_n) < a_{n+1}$ holds for almost all n. It follows that if n is sufficiently large, then some text of $\{a_n, a_n+1, a_n+2, \ldots\}$ is translated to a text of one of the sets $\{a_k, a_k+1, a_k+2, \ldots\}$ with $k \leq n$. Now fix a text T for $\{a_n, a_n+1, a_n+2, \ldots\}$. One can then inductively define a sequence of strings $\sigma_n, \sigma_{n-1}, \ldots, \sigma_0$ such that each sequence $\sigma_n \sigma_{n-1} \ldots \sigma_m T$ is a text for $\{a_m, a_m+1, a_m+2, \ldots\}$ and

$$\text{content}(\Theta(\sigma_n \sigma_{n-1} \ldots \sigma_m \sigma_{m-1})) \not\subseteq \text{content}(\Theta(\sigma_n \sigma_{n-1} \ldots \sigma_m T))$$

for each $m \leq n$. As Θ maps texts of infinite sets in $\mathcal{L}$ to texts of infinite sets in $\mathcal{L}$, one can conclude that

$$\text{content}(\Theta(\sigma_n \sigma_{n-1} \ldots \sigma_m T)) = \{a_m, a_m+1, a_m+2, \ldots\}.$$

Thus, for every m, some text of the set $\{a_m, a_m+1, a_m+2, \ldots\}$ is mapped to a text for the same set contradicting the assumption that Θ does not have $\{a_0, a_0+1, a_0+2, \ldots\}$ in its range. Therefore $\mathcal{L}$ is not strongly mitotic. $\square$

3 Incomplete Learnable Classes

Finite classes are not mitotic and thus every nonempty class has a subclass which is not mitotic. For infinite classes, one can get that the corresponding subclass is also infinite. The proof is a standard application of Ramsey's Theorem: Given classifiers $C_0, C_1, C_2, \ldots$, one selects a subclass $\{H_0, H_1, H_2, \ldots\}$ of $\{L_0, L_1, L_2, \ldots\}$ such that each classifier C_n classifies $H_n, H_{n+1}, H_{n+2}, \ldots$ in the same way. The class $\{H_0, H_1, H_2, \ldots\}$ may not be an indexed family but be a very thin class in the sense that the indices of H_n with respect to $L_0, L_1, L_2, \ldots$ are growing very fast. Alternatively, one can also take the $H_0, H_1, H_2, \ldots$ such that, for a given enumeration of primitive recursive operators, the text $\Theta_n(T_m)$ of the ascending text T_m of H_m is not a text for any H_k with $k > \max(\{n, m\})$. The latter method gives the following result.

Theorem 10. *Every infinite class $\mathcal{L}$ has an infinite subclass $\mathcal{H}$ such that $\mathcal{H}$ is not weakly reducible to any proper subclass of $\mathcal{H}$. In particular, $\mathcal{H}$ is not weakly mitotic.*

There is an easier example of a class which is not weakly mitotic. It is even an indexed family consisting only of finite sets.

Example 11. *Assume that $\{L_0, L_1, L_2, \ldots\}$ is given as $L_0 = \{0, 1\}$ and $L_n = \{n\}$ for all $n \in \mathbb{N} - \{0\}$. Then $\{L_0, L_1, L_2, \ldots\}$ is not weakly mitotic.*

Proof. Given any splitting $\mathcal{L}_0, \mathcal{L}_1$ of $\{L_0, L_1, L_2, \ldots\}$, one of these classes, say $\mathcal{L}_0$, contains at most one of the sets L_0, L_1. Then, for any given reduction (Θ, Ψ) from $\{L_0, L_1, L_2, \ldots\}$ to $\mathcal{L}_0$, $\Theta(\sigma)$ produces some string of nonempty content for some $\sigma \in 1\#^*$. Thus there are texts T_0, T_1 extending σ for L_0 and L_1, respectively, such that $\Theta(T_0)$ and $\Theta(T_1)$ are texts for different sets in $\mathcal{L}_0$ with a nonempty intersection. However, such sets do not exist, by choice of $\mathcal{L}_0$. $\square$

Note that the class

$$\{\{0, 1, 2\}, \{1, 2\}, \{2\}, \{3\}, \{4\}, \{5\}, \ldots, \{n\}, \ldots\}$$

compared with the class from Example 11 has the following slight improvement. For any splitting $\mathcal{L}_0, \mathcal{L}_1$ of the class, one half of the splitting contains an ascending chain of two or three sets, while the other half contains only disjoint sets. Thus the two halves are not equivalent with respect to $\leq_{weak}$.

As these two examples show, it is more adequate to study the splitting of more restrictive classes like the inclusion-free classes. A special case of such classes are the finitely learnable classes. Here a class is finitely learnable [11] iff there is a learner which, on every text for a language to be learnt, outputs only one hypothesis, which must be correct. For technical reasons, the learner keeps outputting a special symbol denoting the absence of a reasonable conjecture until it outputs its only hypothesis.

Theorem 12. $\{L_0, L_1, L_2, \ldots\} \equiv_{strong} \{H_0, H_1, H_2, \ldots\}$ *whenever both classes are infinite indexed families which are finitely learnable. In particular, every such class is strongly mitotic.*

Proof. As $\{L_0, L_1, L_2, \ldots\}$ and $\{H_0, H_1, H_2, \ldots\}$ are infinite, one can without loss of generality assume that the underlying enumerations are one-to-one. Furthermore, they have exact finite learners M and N, respectively, which use the corresponding indexing. Now one translates $\{L_0, L_1, L_2, \ldots\}$ to $\{H_0, H_1, H_2, \ldots\}$ by mapping L_n to H_n; thus Ψ is the identity mapping, where in the domain the n stands for H_n and in the range the n stands for L_n. $\Theta(T) = \#^k T_n$ where k is the least number such that M outputs a hypothesis n on input $T[k]$ (that is, first position where M conjectures a hypothesis) and T_n is the ascending text of H_n. This completes the proof of the first statement.

Given now an infinite finitely learnable class $\{L_0, L_1, L_2, \ldots\}$, one can split it into $\{L_0, L_2, L_4, \ldots\}$ and $\{L_1, L_3, L_5, \ldots\}$, which are the subclasses of languages with even and odd index, respectively. Both classes are also infinite indexed families which are finitely learnable. Thus they are all equivalent by the above result. Furthermore, a classifier for splitting can be obtained by simulating the exact finite learner M for $\{L_0, L_1, L_2, \ldots\}$ on the input text, and then converging to 0 if the (only) grammar output by M on the input text is even, and to 1 if the (only) grammar output by M on the input text is odd. $\square$

4 Further Splitting Theorems

Another question is whether classes can be split into incomparable classes. So one would ask whether there is a parallel result to Sacks Splitting Theorem [20]: Every recursively enumerable but nonrecursive set A is the disjoint union of two recursively enumerable sets A_0 and A_1 such that the Turing degrees of A_0 and A_1 are incomparable and strictly below the one of A. The next example shows that there are classes where every splitting is of this form; so these classes are not weakly mitotic. Furthermore, splittings exist, so the result is not making use of a pathological diagonalization against all classifiers.

Example 13. *Let A be a maximal set. If $a \notin A$, then let $L_a = \{a\}$, else let $L_a = A$. Then $\{L_0, L_1, L_2, \ldots\}$ is recursively enumerable and finitely learnable, but any splitting $\mathcal{L}_0, \mathcal{L}_1$ of $\{L_0, L_1, L_2, \ldots\}$ satisfies $\mathcal{L}_0 \not\leq_{weak} \mathcal{L}_1$ and $\mathcal{L}_1 \not\leq_{weak} \mathcal{L}_0$.*

Proof. Let $T_0, T_1, T_2, \ldots$ be a recursive enumeration of recursive texts for $L_0, L_1, L_2, \ldots$, respectively. Let $F(a)$ be the cardinality of $\{b < a \mid b \notin A\}$. It is easy to see that one can split $\{L_0, L_1, L_2, \ldots\}$ into $\{L_a \mid a \in A \vee F(a) \text{ is even}\}$ and $\{L_a \mid a \notin A \wedge F(a) \text{ is odd}\}$. Thus this class has a splitting; in fact there are infinitely many of them. Furthermore, $\{L_0, L_1, L_2, \ldots\}$ is finitely learnable by outputting an index for L_a for the first a occurring in a given text.

Assume now by way of contradiction that there is a splitting $\mathcal{L}_0, \mathcal{L}_1$ with $\mathcal{L}_0 \leq_{weak} \mathcal{L}_1$ via a reduction (Θ, Ψ). Now one defines the partial recursive function f which outputs on input a the first number occurring in $\Theta(T_a)$; if there occurs no number then $f(a)$ is undefined. As $\mathcal{L}_0$ is infinite, there are infinitely many $a \notin A$ with $L_a \in \mathcal{L}_0$. For all but one of these, $\Theta(T_a)$ has to be a text for some set $L_b \neq A$ in $\mathcal{L}_1$. Then $L_b = \{b\}$ and $f(a) = b \notin A$ for these a's. It follows that, for every x, there is an $a > x$ with $a \notin A \wedge f(a) \notin A \wedge f(a) > x$. Then, by Remark 8, $f(a) = a$ for almost all $a \notin A$. As infinitely many of these a's belong to an $L_a \in \mathcal{L}_0$, one has that $\Theta(T_a)$ is a text for L_a and Θ translates some text for a set in $\mathcal{L}_0$ into a text for a set in $\mathcal{L}_0$ and not into a text for a set in $\mathcal{L}_1$. Thus $\mathcal{L}_0 \not\leq_{weak} \mathcal{L}_1$. By symmetry of the argument, $\mathcal{L}_1 \not\leq_{weak} \mathcal{L}_0$. □

While Example 13 showed that there are classes for which every splitting is a Sacks splitting, the next result shows that every explanatorily learnable recursively enumerable class has a Sacks splitting; but it might also have other splittings.

Theorem 14. *Every infinite recursively enumerable and explanatorily learnable class* $\{L_0, L_1, L_2, \ldots\}$ *has a splitting into two infinite subclasses* $\mathcal{L}_0, \mathcal{L}_1$ *such that* $\mathcal{L}_0 \not\leq_{weak} \mathcal{L}_1$ *and* $\mathcal{L}_1 \not\leq_{weak} \mathcal{L}_0$.

Proof. Let M be a explanatory learner for $\{L_0, L_1, L_2, \ldots\}$ which satisfies the three conditions from Remark 5. Now one defines the following function F from $\mathbb{N}$ to $\{0, 1\}$, inductively using the oracle K for the halting problem.

Let $\Theta_0, \Theta_1, \Theta_2, \ldots$ be the enumeration of operators as given in Remark 2. Let T_n be a text for L_n, constructed effectively from n (as $\{L_0, L_1, L_2, \ldots\}$ is recursively enumerable, this can be done). Let U be the set of all minimal indices of languages in $\{L_0, L_1, L_2, \ldots\}$, that is, $n \in U$ iff for all $m < n$, $L_m \neq L_n$. Note that one can decide U relative to K, since $L_m = L_n$ iff M converges to the same index on both T_m and T_n. Let $F^*(n, a)$ be the number of $k \in U$ satisfying $k < n$ and $F(k) = a$. Note that $F^*(n, a)$ can be computed using oracle K (assuming F can be computed relative to K, as will be shown later).

The value of $F(n)$ is defined by applying the case below, with highest priority (reflected by the least number), which applies:

- Case with priority 0: $n \notin U$. Then there is an $m < n$ such that $L_m = L_n$. Let $F(n) = F(m)$ for the least such m.
- Case with priority $4e + 1$: There is an $m < n$ such that (a) M converges on $\Theta_e(T_m)$ and T_n to the same value, (b) $F(m) = 0$ and (c) there are no $i, j < n$ such that $F(i) = 0$, $F(j) = 0$ and M converges on $\Theta_e(T_i)$ to the same value as on T_j. Then let $F(n) = 0$.
- Case with priority $4e + 2$: There is an $m < n$ such that (a) M converges on $\Theta_e(T_m)$ and T_n to the same value, (b) $F(m) = 1$ and (c) there are no $i, j < n$ such that $F(i) = 1$, $F(j) = 1$ and M converges on $\Theta_e(T_i)$ to the same value as on T_j. Then let $F(n) = 1$.
- Case with priority $4e + 3$: $F^*(m, 0) < F^*(m, 1) + e$ for all $m \leq n$. Then let $F(n) = 0$.
- Case with priority $4e + 4$: $F^*(m, 1) < F^*(m, 0) + e$ for all $m \leq n$. Then let $F(n) = 1$.

More precisely, when defining $F(n)$ one searches for the least number k such that the case with priority k applies, and then defines F as described in that case. Note that the conditions for case with priorities $4e + 3$ and $4e + 4$ apply for all $e > n$, and thus there is always some case which applies.

Next it is shown that $F \leq_T K$. As M converges on T_m, for every m, the test whether condition of the case with priority 0 applies for computing $F(n)$ (and the corresponding computation of $F(n)$) can be done relative to K: $L_m = L_n$ iff M converges to the same value on both T_m and T_n. For the test whether condition of the case with priority $4e+1$ applies, it needs to be checked whether M converges to the same value on both $\Theta_e(T_m)$ and T_n. This can be done using oracle for K as follows. First compute the value d to which M converges on T_n. Then, as M satisfies the constraints in Remark 5, M on $\Theta_e(T_m)$ either converges to d or outputs d only finitely often. Thus one can check, using the oracle for K, whether M converges on $\Theta_e(T_m)$ to d. Similarly one can check, for any two numbers i and $j < n$, whether M converges to the same value on $\Theta_e(T_i)$ and T_j. Thus, one can test, using oracle for K, whether the case with priority $4e+1$ applies. Similarly one can test whether the case with priority $4e+2$ applies. The tests for the case with priorities $4e+3$ and $4e+4$ are obviously doable relative to K as $U \leq_T K$ and the conditions only refer to statistics of previous values of F at places where the argument is in U.

So F can be computed in the limit. Having an approximation F_s to F, one defines a classifier C as $C(\sigma) = F_{|\sigma|}(m)$ for the least m such that $m = |\sigma| \vee M(T_m[|\sigma|]) = M(\sigma)$. Assume now a text T of a language in $\{L_0, L_1, L_2, \ldots\}$ is given and n is the least index such that $L_n = \mathrm{content}(T)$. Then, for all sufficiently large s, $C(T[s]) = F_s(n)$ (as M converges on T and T_n to the same index of L_n, but, for $m < n$, M converges on T_m to an index for the language L_m which is not equal to L_n). Thus C converges on T to $F(n)$.

Clearly the case with priority 0 is applied for computing $F(n)$ iff $n \notin U$ (this happens for infinitely many n). Now it is shown by induction that the case with priority $4e + c > 0$ (where $c \in \{1, 2, 3, 4\}$) applies for computing only finitely many $F(n)$. Assume by induction that, for some ℓ, all cases with priorities k strictly between 0 and $4e + c$ do not apply for computation of $F(n)$, $n \geq \ell$, and the case with priority $4e + c$ applies for computing $F(\ell)$. Now one makes a case-distinction depending on which of the priorities $4e + c$ is applied for computing $F(\ell)$.

In the case of priority $4e + 1$, for $n = \ell$, there is an $m < n$ such that (a) $F(m) = 0$, (b) $F(n) = 0$ and (c) M converges on $\Theta_e(T_m)$ and T_n to the same value. Now let $i = m$ and $j = \ell$. Then, for $n > \ell$, these values i, j are below n, and ensure that the case with priority $4e+1$ does not apply. So $n = \ell$ is the maximal n where $F(n)$ is defined according to this case.

In the case of priority $4e + 3$, consider the set $\{n_0, n_1, n_2, \ldots, n_k\}$ of the least $k+1$ elements of $U \cap \{\ell, \ell+1, \ell+2, \ldots\}$, where $k = e + F^*(\ell, 1) - F^*(\ell, 0)$. One can now prove by induction for $u = 0, 1, 2, \ldots$ that

- for n_u with $u < k$, the case with priority $4e + 3$ applies and $F(n_u) = 0$;
- for n_u with $u \leq k$, $F^*(n_u, 0) = F^*(\ell, 0) + u$ and $F^*(n_u, 1) = F^*(\ell, 1)$.

Thus, $F^*(n_k, 0) = F^*(n_k, 1) + e$, and for $n \geq n_k$ the case with priority $4e + 3$ does not apply for computing $F(n)$.

The other two cases of priority $4e + 2$ and $4e + 4$ are symmetric to the two previous cases. Thus, one can conclude that there are only finitely many n where the case with priority $4e + c$ applies in the computation of $F(n)$. This completes the inductive step.

Let $\{L_0, L_1, L_2, \ldots\}$ be split into two classes as $\mathcal{L}_a = \{L_i : F(i) = a\}$. Now assume by way of contradiction that there is a reduction (Θ_e, Ψ) witnessing that $\mathcal{L}_0 \leq_{weak} \mathcal{L}_1$. Let ℓ be so large that the cases with priorities $1, 2, \ldots, 4e$ are not used to define any $F(n)$ with $n \geq \ell$. Due to the case with priority $4\ell + 7$, there is an $\ell' \in U$ with $F^*(\ell', 0) \geq F^*(\ell', 1) + \ell + 1$; note that $\ell' > \ell$. So more sets in $\mathcal{L}_0$ than in $\mathcal{L}_1$ have an index below ℓ'. Thus, there is an $m \leq \ell'$ such that $L_m \in \mathcal{L}_0$ and $\Theta_e(T_m)$ is not the text of any of the sets $L_0, L_1, \ldots, L_{\ell'}$. Let n be the minimal index of $\mathrm{content}(\Theta_e(T_m))$; this index exists as Θ_e maps texts of languages in $\mathcal{L}_0$ to texts of languages in $\mathcal{L}_1$. It follows from the construction that either $F(n) = F(m) = 0$ and $L_n \in \mathcal{L}_0$ or there are

$i, j < n$ with $F(i) = F(j) = 0$, $L_i, L_j \in \mathcal{L}_0$ and $\Theta_e(T_i)$ being a text for L_j. This contradicts the assumption that (Θ_e, Ψ) reduces $\mathcal{L}_0$ to $\mathcal{L}_1$. Hence $\mathcal{L}_0 \not\leq_{weak} \mathcal{L}_1$. Similarly one can show that $\mathcal{L}_1 \not\leq_{weak} \mathcal{L}_0$. $\square$

For this reason, one cannot give a recursively enumerable class where all splittings $\mathcal{L}_0, \mathcal{L}_1$ satisfy either $\mathcal{L}_0 \leq_{strong} \mathcal{L}_1$ or $\mathcal{L}_1 \leq_{strong} \mathcal{L}_0$. Furthermore, complete classes have comparable splittings like before as they are mitotic and have even equivalent splittings. The next example gives a class where halves of some splittings are comparable but are never equivalent.

Example 15. *Let A be a maximal set. For all $a \in \mathbb{N}$ and $b \in \{0, 1, 2\}$, let $L_{3a+b} = \{3a + b\}$ if $a \notin A$ and $L_{3a+b} = \{3c + b \mid c \in A\}$ if $a \in A$. Then $\{L_0, L_1, L_2, \ldots\}$ is not weakly mitotic as no halves of any splitting are equivalent with respect to $\leq_{weak}$; but $\{L_0, L_1, L_2, \ldots\}$ has a splitting $\mathcal{L}_0, \mathcal{L}_1$ with $\mathcal{L}_0 \leq_{strong} \mathcal{L}_1$.*

Proof. If one takes the splitting $\mathcal{L}_0 = \{L_0, L_3, L_6, \ldots\}$ and $\mathcal{L}_1 = \{L_1, L_2, L_4, L_5, L_7, L_8, \ldots\}$ then it is easy to see that $\mathcal{L}_0 \leq_{strong} \mathcal{L}_1$ via (Θ, Ψ) such that Θ is based on translating every datum $3x$ to $3x + 1$ and Ψ is based on transforming every index e into an index for $\{3x \mid 3x + 1 \in W_e\}$. The details are left to the reader.

Given now a further splitting $\mathcal{L}_2, \mathcal{L}_3$ of $\{L_0, L_1, L_2, \ldots\}$, one of these two classes, say $\mathcal{L}_2$, must contain at least two of the sets $L_{3a}, L_{3a+1}, L_{3a+2}$ for infinitely many $a \notin A$. Assume by way of contradiction that (Θ, Ψ) would witness $\mathcal{L}_2 \leq_{weak} \mathcal{L}_3$. Now one defines the following functions f_b for $b = 0, 1, 2$ by letting $f_b(a)$ to be the number x such that $3x$ or $3x + 1$ or $3x + 2$ occurs earliest in the text $\Theta((3a + b)^\infty)$. Now choose two different $b, b' \in \{0, 1, 2\}$ such that there are infinitely many $a \in \mathbb{N} - A$ with $L_{3a+b}, L_{3a+b'} \in \mathcal{L}_2$. Then one knows that, for every bound c, there are infinitely many $a \in \mathbb{N} - A$ such that $L_{3a+b} \in \mathcal{L}_2$ and $\Theta((3a + b)^\infty)$ is a text for some language in $\mathcal{L}_3 - \{L_0, L_1, L_2, \ldots, L_c\}$. It follows by Remark 8 that $f_b(a) = a$ for almost all $a \notin A$. The same applies to $f_{b'}$. So there is an $a \notin A$ such that $L_{3a+b}, L_{3a+b'}$ are both in $\mathcal{L}_2$ and that Θ maps texts of both languages to texts of the sets L_{3a}, L_{3a+1}, or L_{3a+2}. As only one of these sets can be in $\mathcal{L}_3$, Θ has to map texts of different languages to texts of the same language, a contradiction. Thus $\mathcal{L}_2 \not\leq_{weak} \mathcal{L}_3$ and the class cannot be weakly mitotic. $\square$

While in recursion theory a splitting A_0, A_1 of a recursively enumerable set A satisfies $A_0 \equiv_T A_1 \Rightarrow A_0 \equiv_T A$, the next example shows that the corresponding connection does not hold in inductive inference. The proof is very similar to Example 15 and thus omitted.

Example 16. *Let A be a maximal set. For all $a \in \mathbb{N}$ and $b \in \{0, 1\}$, let $L_{2a+b} = \{2a + b\}$ if $a \notin A$ and $L_{2a+b} = \{2c + b \mid c \in A\}$ if $a \in A$. Then $\{L_0, L_1, L_2, \ldots\}$ is not weakly mitotic but the halves of the splitting $\{L_0, L_2, L_4, \ldots\}, \{L_1, L_3, L_5, \ldots\}$ are equivalent with respect to $\leq_{strong}$.*

5 Beyond Explanatory Learning

One could, besides classes which are complete for explanatory learning, also consider classes which are complete for behaviourally correct learning [3, 5, 18] with respect to $\leq_{strong}$. Note that such a class $\mathcal{L}$ may no longer be explanatorily learnable, but $\mathcal{L}$ satisfies the following two properties:

– The class $\mathcal{L}$ is behaviourally correct learnable, that is, there is a learner which outputs, on every text T for a language in $\mathcal{L}$, an infinite sequence $e_0, e_1, e_2, \ldots$ of hypotheses such that $W_{e_n} = \text{content}(T)$ for almost all n;
– Every behaviourally correct learnable class $\mathcal{H}$ satisfies $\mathcal{H} \leq_{strong} \mathcal{L}$.

Note that the reduction $\leq_{strong}$ considered in this paper is always the same as defined for explanatory learning; reducibilities more adapted to behaviourally correct learning had also been studied [12, 13]. Completeness with respect to $\leq_{weak}$ is not considered in this section, so "complete" means "complete for $\leq_{strong}$" in this section.

It is easy to show that such complete classes exist. Consider as an example the class $\mathcal{L}$ of all sets $\{x\} \cup \{x + y + 1 \mid y \in L\}$, where the x-th learner behaviourally correct learns the set L. So given any behaviourally correct learnable class and an index x of its learner, the translation $L \mapsto \{x\} \cup \{x + y + 1 \mid y \in L\}$ would translate all the sets behaviourally correct learnt by this learner into sets in $\mathcal{L}$.

In the following let $\mathcal{L}$ be any class which is complete for behaviourally correct learning with respect to $\leq_{strong}$. Note that methods similar to those in Theorem 3 show that $\mathcal{L}$ is strongly mitotic. The next result shows that for any splitting $\mathcal{L}_0, \mathcal{L}_1$ of $\mathcal{L}$, one of these two classes is complete for behaviourally correct learning as well and therefore this class cannot be split into two incomparable subclasses.

Theorem 17. *If $\mathcal{L}_0, \mathcal{L}_1$ are a splitting of a class which is complete for behaviourally correct learning with respect to $\leq_{strong}$, then either $\mathcal{L}_0 \equiv_{strong} \mathcal{L}_0 \cup \mathcal{L}_1$ or $\mathcal{L}_1 \equiv_{strong} \mathcal{L}_0 \cup \mathcal{L}_1$.*

Proof. First the theorem is shown for a special class $\mathcal{S}$ defined below. Let $\mathcal{H}$ be a class which is complete for behaviourally correct learning with respect to $\leq_{strong}$. Furthermore, let $C_0, C_1, \ldots$ be a list of all primitive recursive classifiers. One builds, for each x, a sequence $\tau_{x,0}, \tau_{x,1}, \ldots$ starting with $\tau_{x,0} = x$ as follows. If $\tau_{x,y}$ has been defined, then one takes $\tau_{x,y+1}$ to be the first extension of $\tau_{x,y}$ found, if any, such that

$$\{x, x + 1, x + 2, \ldots, x + y\} \subseteq \text{content}(\tau_{x,y+1}) \subseteq \{x, x + 1, x + 2, \ldots\}$$

and $C_x(\tau_{x,y+1}) \neq C_x(\tau_{x,y})$.

In the case that the above process terminates at some y, that is, if $\tau_{x,y+1}$ does not exist, then let $z = \max(\text{content}(\tau_{x,y}))$ and place in $\mathcal{S}$ the set

$$\{x, x + 1, x + 2, \ldots, z\} \cup \{z + u + 1 \mid u \in H\},$$

for each $H \in \mathcal{H}$.

On the other hand, if the process does not terminate at any y, then $T = \lim_{y \in \mathbb{N}} \tau_{x,y}$ is a text for $\{x, x + 1, x + 2, \ldots\}$ on which C_x does not converge; in this case, one places the set

$$\{x, x + 1, x + 2, \ldots\}$$

into $\mathcal{S}$. Note that, in this case, C_x does not split $\mathcal{S}$ into two subclasses, as it diverges on a text for $\{x, x + 1, x + 2, \ldots\}$.

Now it is shown that $\mathcal{S}$ is behaviourally correct learnable. Suppose M is a behaviourally correct learner for $\mathcal{H}$. The new learner N for $\mathcal{S}$ works as follows. On input $T[n]$, it determines $x_n = \min(\text{content}(T[n]))$, $y_n =$ largest value of y such that $\tau_{x_n,y}$ is defined within n steps in the process given above and $z_n = \max(\text{content}(\tau_{x_n,y_n}))$. The learner N then constructs a new string η, where it replaces every u in $T[n]$ by $u - z_n - 1$, if $u > z_n$, and by #, if $u < z_n$. Then N conjectures the following set:

$$W_{N(\sigma)} = \begin{cases} \{x_n, x_n+1, x_n+2, \ldots, z_n\} & \\ \quad \cup \ \{u + z_n + 1 \mid u \in W_{M(\eta)}\} & \text{if } \tau_{x_n,y_n+1} \text{ does not get defined;} \\ \{x_n, x_n+1, x_n+2, \ldots\} & \text{if } \tau_{x_n,y_n+1} \text{ is defined.} \end{cases}$$

Note that the definition above is valid, as $\{x_n, x_n{+}1, x_n{+}2, \ldots, z_n\} \cup \{u{+}z_n{+}1 \mid u \in W_{M(\eta)}\} \subseteq \{x_n, x_n{+}1, x_n{+}2, \ldots\}$ and thus one can always switch to the second case when τ_{x_n,y_n+1} is defined. The verification that N indeed behaviourally correct learns $\mathcal{S}$ is straightforward; therefore the verification is omitted.

Now consider any classifier C_x which converges on every text for a language in $\mathcal{S}$. Then there is a maximal y such that $\tau_{x,y}$ is defined (since otherwise C_x does not converge on $T = \bigcup_{y\in\mathbb{N}} \tau_{x,y}$, which is a text for the language in $\mathcal{S}$). Let $z = \max(\text{content}(\tau_{x,y}))$. Therefore, the class $\mathcal{H}$ is strongly reducible to the subclass

$$\{\, \{x, x+1, x+2, \ldots, z\} \cup \{z + u + 1 \mid u \in H\} \mid H \in \mathcal{H}\}$$

of $\mathcal{S}$. Furthermore, every set in this subclass has a text starting with $\tau_{x,y}$ and C_x converges on all such texts to $C_x(\tau_{x,y})$. Therefore this complete class is contained in one member of the splitting of $\mathcal{S}$ defined by C_x. Thus, one of these members is complete for behaviourally correct learning with respect to $\leq_{strong}$.

After dealing with this special class $\mathcal{S}$, consider any splitting $\mathcal{L}_0, \mathcal{L}_1$ of a class $\mathcal{L}$ which is complete for behaviourally correct learning with respect to $\leq_{strong}$. There is a reduction (Θ, Γ) from $\mathcal{S}$ to $\mathcal{L}_0 \cup \mathcal{L}_1$ due to completeness. Let C be the classifier which splits $\mathcal{L}$ into $\mathcal{L}_0, \mathcal{L}_1$. Let $\mathcal{S}_a$ consist of those L in $\mathcal{S}$ which are mapped to $\mathcal{L}_a$ by Θ. Note that this splitting of $\mathcal{S}$ can be effectively obtained by using Θ and C. As one of $\mathcal{S}_0$, $\mathcal{S}_1$ is complete for behaviourally correct learning (by proof above), one of $\mathcal{L}_0$, $\mathcal{L}_1$ is complete for behaviourally correct learning with respect to $\leq_{strong}$. $\square$

As just seen, any splitting $\mathcal{L}_0, \mathcal{L}_1$ of a class which is complete for behaviourally correct learning satisfies either $\mathcal{L}_0 \equiv_{strong} \mathcal{L}_1$ or $\mathcal{L}_0 <_{strong} \mathcal{L}_1$ or $\mathcal{L}_1 <_{strong} \mathcal{L}_0$. As the class is strongly mitotic, it can happen that the two halves of a splitting are equivalent (although this is not always the case). The next result gives a class where the two halves of a splitting are always comparable but never equivalent.

Theorem 18. *There is a recursively enumerable and behaviourally correct learnable class, which is not weakly mitotic, such that every splitting $\mathcal{L}_0, \mathcal{L}_1$ of the class satisfies either $\mathcal{L}_0 \leq_{strong} \mathcal{L}_1$ or $\mathcal{L}_1 \leq_{strong} \mathcal{L}_0$, but not $\mathcal{L}_0 \equiv_{weak} \mathcal{L}_1$.*

Proof. A minor modification of the construction of Post's simple set [19] gives the following: There is a recursive partition of the odd natural numbers into sets $I_0, I_1, I_2, \ldots$ and a recursively enumerable set L_1 of odd natural numbers such that (a) $1 \in L_1$, (b) $I_n \not\subseteq L_1$ for all n, (c) L_1 intersects every recursively enumerable set which contains infinitely many odd natural numbers.

For $n \neq 1$, if n is even, then let $L_n = \{n\}$, else let $L_n = L_1 \cup \{n\}$.

Clearly $\{L_0, L_1, L_2, \ldots\}$ is behaviourally correct learnable: on input σ, the learner conjectures content(σ), if σ does not contain an odd number; otherwise it conjectures content(σ) $\cup L_1$.

Let $\mathcal{L}_0, \mathcal{L}_1$ be a splitting of $\{L_0, L_1, L_2, \ldots\}$. L_1 is in one of these classes, say in $\mathcal{L}_1$. Let C be the classifier witnessing the split. Then there is a stabilizing sequence σ for C on L_1, such that content(σ) $\subseteq L_1$ and $C(\tau) = 1$ for all extensions τ of σ with content(τ) $\subseteq L_1$. Let T be a text of L_1. Now for every $a \in L_1$ and every n, $C(\sigma a T[n]) = 1$. Since L_1 is simple, it follows that the set

$$D = \{a \mid a \text{ is odd and } \exists n\, [M(\sigma a T[n]) = 0]\}$$

is finite and thus $L_a \in \mathcal{L}_1$ for all odd $a \notin D$. Let n be such that $D \cap I_m = \emptyset$ for all $m \geq n$. Let $a_0, a_1, a_2, \ldots$ be an ascending enumeration of all even numbers plus members of D. Note that $\mathcal{L}_0 \subseteq \{L_{a_0}, L_{a_1}, L_{a_2}, \ldots\}$.

Let $b_m = \min(I_{n+m} - L_1)$ — note that b_m is well defined since $I_{n+m} \not\subseteq L_1$ for all m. Note that $L_1 \cup \{b_m\}$ is in $\mathcal{L}_1$ for all m.

Now it is shown that $\mathcal{L}_0 \leq_{strong} \mathcal{L}_1$. Let Θ be an operator which maps texts for L_{a_m} to texts for $L_1 \cup \{b_m\}$. Note that such a Θ can easily be constructed. For Ψ, given a sequence E, one can determine, in the limit, the grammar e to which E converges (if any), and then determine, in the limit, the m such that e is a grammar for $L_1 \cup \{b_m\}$ (if there is any such m). Thus, one can construct a sequence which converges to a grammar for L_{a_m}, whenever such an m as above exists.

On the other hand, if an operator translates texts of sets in $\mathcal{L}_1$ into texts of sets in $\mathcal{L}_0$, then it has to map some text T of L_1 to some text of some L_{a_m}. There is an initial segment σ of T such that a_m appears on the output when σ is fed into the operator. There is only the set L_{a_m} in $\mathcal{L}_0$ containing a_m, but infinitely many sets in $\mathcal{L}_1$ that have a text starting with σ. Therefore the translation maps some texts of different sets in $\mathcal{L}_1$ to texts of L_{a_m}. Thus, the translation cannot be used for a weak reduction from $\mathcal{L}_1$ to $\mathcal{L}_0$. Hence $\mathcal{L}_1 \not\leq_{weak} \mathcal{L}_0$. $\square$

6 Autoreducibility

Trakhtenbrot [23] defined that a set A is autoreducible iff one can reduce A to A such that $A(x)$ is obtained by accessing A only at places different than x. Ladner [15] showed that a recursively enumerable set is mitotic iff it is autoreducible. Ambos-Spies pointed this result out to the authors and asked whether the same holds in the setting of inductive inference. Unfortunately, this characterisation fails for both of the major variants of autoreducibility. These variants are the ones corresponding to strong and weak reducibility.

Definition 19. A class $\mathcal{L}$ is strongly (weakly) autoreducible iff there is a strong (weak) reduction (Θ, Ψ) from $\mathcal{L}$ to $\mathcal{L}$ such that, for all sets $L \in \mathcal{L}$ and all texts T for L, $\Theta(T)$ is a text for a language in $\mathcal{L} - \{L\}$.

Example 20. *Let A be a maximal set and $\mathcal{L}$ contain the following sets:*

- $\{3x\}, \{3x+1\}, \{3x+2\}$ *for all* $x \notin A$*;*
- $\{3y : y \in A\}$*,* $\{3y+1 : y \in A\}$*,* $\{3y+2 : y \in A\}$*.*

Then the class $\mathcal{L}$ is neither strongly mitotic nor weakly mitotic. But $\mathcal{L}$ is autoreducible via some (Θ, Ψ) where Θ maps any text T to a text T' such that all elements of the form $3y$ in T have the form $3y+1$ in T', all elements of the form $3y+1$ in T have the form $3y+2$ in T' and all elements of the form $3y+2$ have the form $3y$ in T'.

So even the implication "strongly autoreducible $\Rightarrow$ weakly mitotic" fails. The remaining question is whether at least the converse direction is true in inductive inference. This is still unknown, but there is some preliminary result on sets which are complete for $\leq_{weak}$.

Theorem 21. *If a class $\mathcal{L}$ is complete for $\leq_{weak}$, then it is weakly autoreducible.*

Proof. Let $\mathcal{L}$ be complete for $\leq_{weak}$ and M be an explanatory learner for $\mathcal{L}$ which satisfies the conditions from Remark 5. As $\mathcal{L}$ is complete for $\leq_{weak}$, by Proposition 6, there is a reduction (Θ, Ψ) from the class $\mathcal{I}$ to $\mathcal{L}$ such that, for any set $I_x = \{0, 1, \ldots, x\} \in \mathcal{I}$ and any text T for I_x, $\Theta(T)$ is a text for a set on which M does not converge to an index less than or equal to x. Now, an autoreduction (Θ', Ψ') is constructed.

For this, one first defines Θ'' as follows and then concatenates it with Θ. The operator Θ'' translates every text T for a set L into a text for $I_{2^n(1+2m)}$ where m, n are chosen such that n is the value to which M converges on T and m is so large that all the elements put into $\Theta''(T)$, when following intermediate hypotheses of M on T, are contained in the set $I_{2^n(1+2m)}$. It is easy to verify that this can be done. Then Θ' is given as $\Theta'(T) = \Theta(\Theta''(T))$. The sequence $\Theta'(T)$ is a text for a set in $\mathcal{L}$ with the additional property that M converges on it to an index larger than $2^n(1+2m)$; this index is therefore different from n and $\mathrm{content}(\Theta'(T)) \neq \mathrm{content}(T)$.

The reverse operator Ψ' can easily be generated from Ψ. If E converges to an index for $\mathrm{content}(\Theta'(T))$, then $\Psi(E)$ converges to some index for $I_{2^n(1+2m)}$. The number $2^n(1+2m)$ can be determined in the limit from this index by enumerating the corresponding finite set; thus Ψ' can translate E via $\Psi(E)$ to a sequence which converges to n. $\square$

Example 22. *The class $\mathcal{L}$ from Theorem 9 is complete for $\leq_{weak}$ and weakly autoreducible but not strongly autoreducible.*

Proof. Let $\mathcal{L}$ and $a_0, a_1, a_2, \ldots$ be as in Theorem 9. Assume that (Θ, Ψ) witness that $\mathcal{L}$ is strongly autoreducible. Then Θ has to preserve inclusions and therefore map infinite sets in $\mathcal{L}$ to infinite sets. So, $\mathrm{content}(\Theta(a_0\,(a_0+1)\,(a_0+2)\,\ldots))$ is an infinite set in $\mathcal{L}$ different from $\{a_0, a_0+1, a_0+2, \ldots\}$. By induction, one can show that

$$\mathrm{content}(\Theta(a_n\,(a_n+1)\,(a_n+2)\,\ldots)) \subseteq \{a_{n+1}, a_{n+1}+1, a_{n+1}+2, \ldots\} \text{ and}$$
$$\mathrm{content}(\Theta(a_n\,(a_n+1)\,(a_n+2)\,\ldots)) \subset \{a_n, a_n+1, a_n+2, \ldots\}.$$

But in Theorem 9 it was shown that no recursive operator has these properties. That $\mathcal{L}$ is complete for $\leq_{weak}$ was shown in Theorem 9 and that $\mathcal{L}$ is weakly autoreducible follows from Theorem 21. □

Acknowledgements. We would like to thank the anonymous referees for several helpful comments.

References

1. Klaus Ambos-Spies. P-mitotic sets. *Logic and Machines*. Springer LNCS 177:1–23, 1984.
2. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
3. Janis Bārzdiņš. Two theorems on the limiting synthesis of functions. In *Theory of Algorithms and Programs*, Volume 1, Latvian State University, Riga, 210:82–88, 1974.
4. Lenore Blum and Manuel Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.
5. John Case and Christopher Lynes. Inductive inference and language identification. *Ninth International Colloquium on Automata, Languages and Programming* (ICALP), Aarhus, Denmark, July 12-16, 1982, Proceedings. Springer, *Lecture Notes in Computer Science*, 140:107–115, 1982.
6. Rusins Freivalds, Efim Kinber and Carl Smith. On the intrinsic complexity of learning. *Information and Computation*, 123:64–71, 1995.
7. Richard Friedberg. Three theorems on recursive enumeration. *The Journal of Symbolic Logic*, 23:309–316, 1958.
8. Mark Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.
9. Christian Glaßer, Aduri Pavan, Alan Selman and Liyu Zhang. Mitosis in computational complexity. *Theory and Applications of Models of Computation*: Third International Conference, TAMC 2006, Beijing, China, May 15-20, 2006. Proceedings. Springer LNCS, 3959:61–67, 2006.
10. Christian Glaßer, Mitsunori Ogihara, Aduri Pavan, Alan Selman and Liyu Zhang. Autoreducibility, mitoticity and immunity. *Mathematical Foundations of Computer Science*: Thirtieth International Symposium, MFCS 2005, Gdansk, Poland, August 29 - September 2, 2005, Proceedings. Springer LNCS, 3618:387–398, 2005.
11. E. Mark Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
12. Sanjay Jain, Efim Kinber and Rolf Wiehagen. Language learning from texts: degrees of intrinsic complexity and their characterizations. *Journal of Computer and System Sciences*, 63:305–354, 2001.
13. Sanjay Jain and Arun Sharma. The intrinsic complexity of language identification. *Journal of Computer and System Sciences*, 52:393–402, 1996.
14. Sanjay Jain and Arun Sharma. The Structure of Intrinsic Complexity of Learning. *The Journal of Symbolic Logic*, 62:1187–1201, 1997.

15. Richard Ladner. Mitotic recursively enumerable sets. *The Journal of Symbolic Logic*, 38:199–211, 1973.
16. Piergiorgio Odifreddi. *Classical Recursion Theory.* North-Holland, Amsterdam, 1989.
17. Daniel N. Osherson, Michael Stob and Scott Weinstein. *Systems That Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists.* Bradford — The MIT Press, Cambridge, Massachusetts, 1986.
18. Daniel N. Osherson and Scott Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.
19. Emil Post. Recursively enumerable sets of positive integers and their decision problems, *Bulletin of the American Mathematical Society*, 50:284–316, 1944.
20. Gerald E. Sacks. On the degrees less than $\mathbf{0}'$. *Annals of Mathematics*, 77:211–231, 1963.
21. Carl H. Smith, Rolf Wiehagen and Thomas Zeugmann. Classifying predicates and languages. *International Journal of Foundations of Computer Science*, 8:15–41, 1997.
22. Robert Soare. *Recursively Enumerable Sets and Degrees.* Springer-Verlag, Heidelberg, 1987.
23. Boris A. Trakhtenbrot. On autoreducibility. *Soviet Mathematics, Doklady* (*Doklady Akademii Nauk SSSR*), 11:814–817, 1970.