Increasing the power of uniform inductive learners

Sandra Zilles

Fachbereich Informatik, Universität Kaiserslautern, Postfach 3049, D-67653 Kaiserslautern

Abstract

The analysis of theoretical learning models is basically concerned with the comparison of identification capabilities in different models. Modifications of the formal constraints affect the quality of the corresponding learners on the one hand and regulate the quantity of learnable classes on the other hand.

For many inductive inference models – such as Gold's identification in the limit – the corresponding relationships of learning potential provided by the compatible learners are well-known. Recent work even corroborates the relevance of these relationships by revealing them still in the context of uniform Gold-style learning. Uniform learning is rather concerned with the synthesis of successful learners instead of their mere existence.

The subsequent analysis further strengthens the results regarding uniform learning, particularly aiming at the design of methods for increasing the potential of the relevant learners. This demonstrates *how* to improve given learning strategies instead of just verifying the existence of more powerful uniform learners.

For technical reasons these results are achieved using various formal conditions concerning the learnability of unions of uniformly learnable classes. Therefore numerous sufficient properties for the learnability of such unions are presented and illustrated with several examples.

1 Introduction

The theory of inductive learning is concerned with abstract learning models. In general, such models for learning by induction consist of

- a *learner*,

Email address: `zilles@informatik.uni-kl.de` (Sandra Zilles).

- a class of possible *target objects* to be identified by the learner,
- a method for supplying the learner with *information* about the target objects during the learning process, and
- a set of possible *hypotheses* the learner may output during the learning process.

Each hypothesis is associated to some object, and hence each representation scheme for the target objects can be understood as a *hypothesis space.*[1] The learner is supposed to find a correct representation for an entire target object, that is, a correct hypothesis, from incomplete information about this object. A quite convenient interpretation is to regard each target object as a rule and the information presented to the learner as examples according to this rule. Then the learner can be understood as a mechanism generating rules from given examples and each representation scheme for such rules may serve as a hypothesis space. In this context, every class of rules constitutes a learning problem. Each formal conception following this scheme specifies a different abstract learning model, and it is not conceivable that any simple theoretical learning model can fully describe all phenomena of natural learning behaviour. Still such conceptions can be used to explain at least certain aspects of human learning or to model 'intelligence' with the help of mechanisms.

An approach allowing for a quite formal analysis is to consider the learners as computable devices or *machines*, each defined by a finite program. Of course, a formal learning model within the given scope must be defined by several technical constraints, such as for example the required quality of the hypotheses returned by the learner, the number of guesses allowed, time or space constraints for the learner, etc. Altogether, these constraints describe a successful learning behaviour. Now by varying these constraints we may also vary the classes of target objects which can be identified by a single learner. That means, each specific learning model – resulting from modifications of the given constraints – yields its specific capacities for the corresponding learners. So each learning model is associated with a *collection* of learnable classes of target objects; such a collection will subsequently be called a *learning type* or an *inference type.*[2]

On the one hand, strengthening the technical constraints should in some sense improve the quality of the learning machine (because it may for example compute its hypotheses in less time or provide hypotheses with useful additional properties). But on the other hand, it is conceivable that any increase in the technical constraints reduces the pool of learnable classes of target objects. So

[1] Note that, in general, hypothesis spaces may also represent objects not contained in the target class.

[2] Informally, the terms *learning type* and *inference type* will also refer to the associated learning model.

there may be a trade-off between the quality of the learner and the quantity of learnable object classes. Investigating this trade-off is a common subject in many fields concerning theoretical aspects of artificial intelligence.

If a learning type I is modified to a learning type J by adding some technical constraints to the corresponding inference conditions, we might ask what price has to be paid, i. e. to what extent the gain in quality results in a loss in quantity. If there are I-learnable classes of target objects, which no longer belong to the learning type J, we have witnessed a *separation* of the learning types I and J. But we may also turn the corresponding question around. We know that I results from J by weakening the constraints, and a separation means that there is now an I-learnable class which was not learnable according to the former type J. Thus we might want to know whether the loss in quality is compensated by a gain in quantity. So far this is only an inverted question, but individual practical needs might give rise to more specific questions: imagine that, for some reason, it is important for you to identify at least all objects in a particular J-learnable class C. If all I-learnable supersets of C were even J-learnable, then relaxing the technical constraints would reduce the quality of the learner without raising the quantity of all 'interesting' learnable classes (i. e. there would not be any gain with respect to your learning goal C). Hence the question should no longer be, whether there exists a class learnable in I and not learnable in J, but whether there exists a *superset* of C fulfilling these properties. A *strong separation* of the learning types I and J holds, if

every J-learnable class C has a superset, which is I-learnable, but no longer J-learnable.

The term *strong separation* was chosen by Case, Chen and Jain [7] in the context of inductive inference of recursive functions. Studying the separability of two learning types involves getting an insight into specific structures of classes of target objects learnable according to either type. A way to get these insights might be to keep an eye on the methods of successful learners. Do they use any particular intrinsic knowledge, any preconditions the target classes or the corresponding hypothesis spaces have to fulfil? Speaking in the terminology of machine learning literature, cf. [17,18], the *bias* of a learning system has to be studied. The concept of bias refers to a learning system's restrictions of the search space, most often based on some kind of background knowledge about the structure of the possible target objects. Such a bias is needed to overcome the general problem of logical justification of hypotheses in inductive learning, in particular, it may account for a limitation of the hypothesis space. In practice the bias is often employed, just because the restriction of the search space renders a complexity advantage. So a learner in general is successful, because it has some prior intrinsic knowledge about the class of target objects. Analogously, if a class of target objects is *not* learnable, then the required background knowledge is presumably not expressible adequately to be exploited by a learner.

This point of view can also be expressed in other terms: if intrinsic knowledge of learners about the target classes or hypothesis spaces is assumed, can this knowledge be exploited in a uniform way? That means we ask for common preconditions in learnable classes of target objects, allowing for a common (uniform) method of induction for all these classes. The idea is to aim at some kind of meta-learner M simulating several (perhaps infinitely many) learners for special classes $C_0, C_1, C_2, \ldots$ of target objects. This realizes an approach to merging several intelligent systems into a single machine able to cope with the tasks of any of the systems, which is not a trivial task, if the resulting machine is required to represent a computable device. In other words, a *single* creative learning procedure shall be used for *numerous* learning problems. This approach is referred to by the term *uniform learning*. In summary, an analysis of uniform learning is of interest for several reasons, for example:

- it concerns the general problem of designing learning systems capable of simulating numerous expert learners for special target classes;
- it concerns common principles of solvable learning problems and common principles for possible corresponding successful learners;
- it concerns the general problem of describing and representing learning problems adequately, and thus of appropriately communicating background knowledge on the particular target classes to the learner.

In particular, the latter aspect has to be explained in detail. Recall that the crucial component of uniform learning is supposed to be some kind of meta-learner M simulating several (perhaps infinitely many) learners for special classes $C_0, C_1, C_2, \ldots$ of target objects. As the intrinsic knowledge to be used by the meta-learner M may depend on the class C_i of target objects currently considered, there must be some way to communicate this special knowledge about C_i to M. This is done via some description d_i representing the class C_i of target objects.

For instance, consider the example of inferring compact convex regions of a 2-dimensional plane. Assume there are just two target classes of interest, namely the class C_0 of rectangular regions and the class C_1 of circular regions. Then the description $d_0 = 0$ might represent the class C_0; the description $d_1 = 1$ might represent the class C_1. A meta-learner supplied with the parameter 0 may then simulate a procedure for learning the target class of rectangular regions; analogously, the meta-learner may simulate a procedure for learning the target class of circular regions on input of the parameter 1. Note that this example just serves for illustration; of course, in the general case, when infinitely many target classes are concerned, the choice of descriptions may become a much more delicate affair.

After this example, let us come back to the general aspect of representing additional background knowledge via descriptions of target classes. On the one hand, by assumption, M exploits some common intrinsic knowledge about the classes $C_0, C_1, C_2, \ldots$; on the other hand the information d_i may add to this bias, because it is used to communicate special knowledge about C_i to the meta-learner. Altogether, the resulting knowledge should be sufficient for M to simulate a successful learner for C_i. Thus M learns any of the classes $C_0, C_1, C_2, \ldots$, if knowledge $d_0, d_1, d_2, \ldots$, respectively, is provided. Now if I is any identification model adequate for learning $C_0, C_1, C_2, \ldots$, then the set $D = \{d_0, d_1, d_2, \ldots\}$ is called *uniformly* I-learnable (*UniI*-learnable for short), if there is some meta-learner M simulating I-learners for $C_0, C_1, C_2, \ldots$ according to the method described above.

The present paper lifts the study of separations and strong separations of learning models onto this meta-level referred to by the term *uniform learning*. Given learning types I and J as before, a separation of *UniI* and *UniJ* simply means that there is some description set D, which is uniformly I-learnable, but not uniformly J-learnable. If some description d in D represents a class C, which is learnable under I, but not under J, a separation of *UniI* and *UniJ* is trivially witnessed. Therefore let us additionally demand, that all classes of target objects described via the separating set D should be J-learnable. Now a strong separation of *UniI* and *UniJ* holds if

every UniJ-learnable set D has a superset, which is UniI-learnable, but no longer UniJ-learnable.

Again it is demanded, that the superset of D describes only J-learnable classes of target objects, thus avoiding trivial results. Now we have addressed the study of the trade-off between the quality of learners (achieved by technical constraints) and the quantity of the corresponding identification type (determined by the amount of learnable classes) to the model of uniform learning.

These are basic considerations of interest for learning in any context. Yet we can only study the impact of relaxing or increasing the technical constraints, if a *formal learning model* is provided. A simple and often studied way to model learning behaviour is inductive inference of recursive functions. In this model – introduced by Gold [13] – any set of total recursive functions forms a possible class of target objects. The learner is an algorithmic device, called *inductive inference machine*, or IIM for short. Gradually growing initial segments of a function graph constitute the information the learner receives during the learning process. As a hypothesis space any fixed computable enumeration of at least all target functions may be used. An index of a function in such an enumeration is considered as a program for this function. In particular, the learner returns natural numbers as its hypotheses and each such number represents a computable function via the programming system associated to the hypothesis space. An IIM M identifies a function f *in the limit*, if M is defined on any initial segment of f and the corresponding sequence of hypotheses returned by M converges to a program for f. Of course, the underlying enumeration – serving as a hypothesis space – should be fixed in advance.

If the constraints in the definition of learning in the limit are modified (e. g. by adding natural demands concerning the intermediate hypotheses), different learning models such as for example *consistent learning* emerge. These learning models have already been compared with respect to the identification potential of the compatible learners, revealing a complete hierarchy of identification types via separations. The basic definitions and comparison results for the corresponding models may be found in [3,4,6,10–13,15,21,22], partly also in the survey paper [1]. Strong separations have especially been studied by Case, Chen and Jain [7].

As explained in general above, Gold-style identification models can also be lifted onto the meta-level of uniform learning. Research on this topic has revealed numerous negative results concerning the learnability of rather simple target classes, see for example [14,16,19]. Very fruitful work has been done by Baliga, Case and Jain [2] in the context of uniform learning of formal languages. Their results – as aimed at in the motivation above – provide much insight into common structures of learnable target classes. In the context of separations of identification types for uniform learning of classes of recursive functions, a complete picture for several types can be found in [24]. The purpose of the present paper is to reveal the corresponding picture of strong separations; as it turns out, all separations achieved before also hold in the strong version. But this analysis is not becoming an end in itself, mostly for two reasons.

Firstly, the results obtained are even stronger than required: given a separated pair $(UniI, UniJ)$ as above, there is a *fixed* description set D^* which – when added to any arbitrary description set D in $UniJ$ – forms a set $D \cup D^*$ suitable for uniform I-learning, but not suitable for uniform J-learning. Hopefully, such a fixed description set contains more information on specific properties disallowing J-learning, since it is complex enough to make uniform J-learning impossible, while still being simple enough always to maintain uniform I-learnability.

Secondly, the proofs of the strong separations provide techniques for changing a J-meta-learner for the original description set D into a meta-learner appropriate for I-identification of the new set $D \cup D^*$. That means methods making use of the possible increase of learning potential are revealed. Previous work by Case, Chen, and Jain, see [7], contains similar results for learning in the non-uniform model.

For strong separation results each description set D has to form a union with some other description set. On the one hand, this union must not be uniformly J-learnable. On the other hand, it must be appropriate for uniform I-learning. So, in order to construct such unions carefully, it might be helpful to find certain properties sufficient for the uniform learnability of the union of two description sets. This will be the purpose of the first main part of technical results, to be found in Section 4. The second main part, presented in Section 5, will deal with the final strong separation results.

A preliminary version of this paper has already appeared, see [25]. The proofs for the strong separations in Section 5 use many methods presented in [26]. Subsequently, details for these methods will only be shown in an example and will be omitted in the general case.

2 Preliminaries

2.1 Notations

Knowledge of basic mathematical concepts and common notions is assumed; for recursion theoretic terms used without definition see [20].

Special notations in the context of set theory are $\subset$ and $\#$, used to indicate proper inclusion and incomparability of sets, respectively. $\emptyset$ is a symbol for the empty set. In order to refer to the cardinality of a set X the notion $card X$ is used, where ∞ is the cardinality of an infinite set. If X is a set of natural numbers, then $\overline{X}$ denotes its complement with respect to the set of all natural numbers. Many of the subsequent results deal with sets of *partial-recursive functions*, cf. [20]. Inputs and outputs of the latter functions are natural numbers; the number of input variables of a particular function will be clear from context on all occasions. e, f, g, and h are used as variables for partial-recursive functions; other variables denoted by lowercase Roman letters (with or without subscripts and superscripts) always range over the natural numbers. $f(n)$ denotes the value of f on input n, where $f(n)\uparrow$ indicates, that f is undefined on input n. The value set $\{f(x) \mid x \geq 0 \text{ and } f(x) \text{ is defined}\}$ of some f will be denoted by $val(f)$. A partial-recursive function which is total, i. e. defined for all inputs, is simply called *recursive function*. Such functions will be the target objects for the learning processes analysed. Sometimes a coding of pairs of natural numbers, i. e. a recursive bijective function mapping pairs of numbers to numbers, is needed. Given numbers n and m, $\langle n, m\rangle$ will denote the corresponding code number.

Via a recursive bijective mapping, finite tuples of natural numbers are identified with natural numbers. Thus, if $f(0), f(1), \ldots, f(n)$ are defined, a code number $f[n]$ is associated with the finite tuple $(f(0), f(1), \ldots, f(n))$, moreover the notions $g(f[n])$ and $g(f(0) \ldots f(n))$ have equal meaning. For convenience, a partial-recursive function may also be regarded as a sequence of output values and 'undefined values' or as a set of input-output pairs. For example let $f(n) = 7$ for $n \leq 2$ and $f(n) \uparrow$ otherwise; $g(n) = 7$ for $n \leq 1$, $g(2) = 6$, and $g(n) \uparrow$ otherwise; $h(n) = 7$ for all n. Then $f = 7^3 \uparrow^\infty = \{(0,7), (1,7), (2,7)\}$, $g = 7^2 6 \uparrow^\infty = \{(0,7), (1,7), (2,6)\}$, $h = 7^\infty = \{(n,7) \mid n \geq 0\}$. The latter representation explains notions like $f \# g$, $g \# h$, $f \subset h$. In this example $f =^* g$ may be written – a notion used, if for all but finitely many n either $f(n)$ and $g(n)$ are both undefined or $f(n) = g(n)$.

Recall that recursive functions serve as the target objects in the inductive inference models considered here. So we need appropriate representation schemes for these functions, to be used as hypothesis spaces later on. The idea is to list, i. e. to enumerate, all possible target objects and represent each object via a number in the resulting list. But to make this list accessible for a computable learner, it should correspond to a computable function, which is not possible for any list of all recursive functions. Therefore we consider partial-recursive enumerations in general: any $(n+1)$-place partial-recursive function ψ enumerates the set $\{\psi_i \mid i \geq 0\}$ of n-place partial-recursive functions, where ψ_i is defined by $\psi_i(x_1, \ldots, x_n) := \psi(i, x_1, \ldots, x_n)$ for all $x_1, \ldots, x_n$. In this context ψ is also called a *numbering*. Given a function f in $\{\psi_i \mid i \geq 0\}$, any index x satisfying $\psi_x = f$ is called a *ψ-number* or a *ψ-program* of f. Note that a numbering may provide more than one program for a single function. A numbering ψ is called *finite*, if $\psi_i = \uparrow^\infty$ for all but finitely many i, i. e. if almost all ψ-programs correspond to the empty function. Frequently, the special term *acceptable numbering* is used. As an example for an acceptable numbering consider any programming system derived from an enumeration of all Turing machines, cf. also [20].

2.2 Inductive inference models

As mentioned in the introduction, some crucial components of a learning model are the learner, the class of possible target objects, as well as a representation scheme to be used as a hypothesis space. The target objects in the inductive inference models considered here are always recursive functions; as a representation scheme some adequate partial-recursive numbering is chosen. It remains to specify the type of learners to be used. Since only computable learners should be investigated, each of these might be considered as some kind of machine, called *inductive inference machine* or IIM for short. An IIM M is an algorithmic device working in time steps. In step n it gets some input $f[n]$ corresponding to an initial segment of a graph of some recursive function

f. If M returns an output on $f[n]$, then this output is a natural number to be interpreted as a program in the given numbering serving as a hypothesis space, cf. [13]. As usual, an IIM which is defined on any input will be called a *total* IIM. Subsequently, the term 'hypothesis space' will always refer to a two-place partial-recursive numbering.

The different inference models defined in this context result from different technical constraints, i. e. from the particular success criteria. In Gold's basic model of *identification in the limit*, cf. [13], also called *explanatory identification*, the IIM working on the graph of some recursive target function f is required to produce guesses converging to a correct program for f. In case M is defined for all inputs $f[n]$, $n \geq 0$, and the sequence $(M(f[n]))_{n\geq 0}$ converges, this will be denoted by $M(f)\downarrow$; moreover $M(f) = i$ then indicates that i is the limit of this sequence. The notion $M(f)\uparrow$ signals the opposite situation.

First the inference type of explanatory identification is defined. Afterwards examples of how to modify the constraints in this model are presented; in particular, three kinds of inference types are considered:

- types resulting from special constraints concerning the success criterion of the sequence of hypotheses;
- types resulting from special constraints concerning the quality of the intermediate hypotheses, independent of the amount of information currently known about the target function;
- types resulting from special constraints concerning the quality of the intermediate hypotheses, depending on the information currently known about the target function.

The inference types defined below are chosen to give at least two representative types for each of these three kinds.

Definition 1 *A class C of recursive functions is identifiable in the limit (Ex-identifiable), iff there is some hypothesis space ψ and an IIM M, such that for any f in C the following conditions are fulfilled:*

(1) $M(f[n])$ is defined for all $n \geq 0$ and $M(f)\downarrow$,
(2) $M(f)$ is a ψ-program for f (i. e. there is some $i \geq 0$, such that $M(f) = i$ and $\psi_i = f$).

Ex denotes the collection of all Ex-learnable classes C.

Each finite class is trivially Ex-learnable. In general, each class C of functions enumerated by a recursive numbering belongs to Ex, cf. Gold's method of *identification by enumeration* [13]. In contrast to that, there is no Ex-learner successful for the whole class of recursive functions, no matter which hypothesis space is used. So it is conceivable that a modification of Definition 1 might yield a learning model allowing for a higher potential of its compatible learners. An approach discussed in [3] is *behaviourally correct identification.* The corresponding model results from learning in the limit, if the demand for convergence of the sequence of hypotheses is loosened. Here the learner is supposed to eventually return correct programs, yet possibly alternating between different correct conjectures.

Definition 2 *A class C of recursive functions is Bc-identifiable, iff there is some hypothesis space ψ and some IIM M, such that for any f in C all values $M(f[n])$ ($n \geq 0$) are defined and all but finitely many of them are ψ-numbers for f. Bc is the collection of all Bc-learnable classes.*

Note that $M(f) \uparrow$ is conceivable for an IIM M which *Bc*-learns a recursive function f. Learners according with the *Bc*-model provide a higher potential than those compatible with the *Ex*-model, in particular, the set *Ex* is a proper subset of *Bc*, cf. [3]. Still the whole class of recursive functions constitutes a learning problem no *Bc*-learner can cope with. A further modification of technical constraints, formally enabling solvability of this problem, follows Case and Smith [10]. Their notion of Bc^*-learning results from permitting 'a few' (i. e. finitely many) errors in each hypothesis.

Definition 3 *A class C of recursive functions is Bc^*-identifiable, iff there is some hypothesis space ψ and some IIM M, such that for any f in C all values $M(f[n])$ ($n \geq 0$) are defined and all but finitely many of them fulfil $\psi_{M(f[n])} =^* f$. Bc^* denotes the collection of all Bc^*-learnable classes.*

According to [10], L. Harrington has verified that each class C of recursive functions – so in particular even the whole class of recursive functions – is Bc^*-identifiable. All in all, weakening the constraints in the definition of learning in the limit yields the hierarchy $Ex \subset Bc \subset Bc^*$ expressing an increase in the learning potential of the corresponding IIMs. In contrast to that, it is conceivable to strengthen the demands of Definition 1 concerning the mind change complexity. Since an IIM *Ex*-learning a recursive function may change its hypothesis in an unbounded finite number of steps, it will never be possible to decide whether the time of convergence is already reached. A restricted learning model with bounds on the number of mind changes is introduced in [10]. In this model the learner is allowed only a certain number of mind changes in its sequence of hypotheses; in particular, whenever this capacity of mind changes is exhausted, the current hypothesis must be correct.

Definition 4 *Let M be an IIM which is permitted to return the auxiliary sign '?'. A class C of recursive functions is Ex_m-identifiable by M, iff C is Ex-learned by M with respect to some hypothesis space ψ, such that for all f in C the following conditions hold:*

(1) there is some $k \geq 0$, such that $M(f[n]) = ?$ iff $n < k$,

(2) $card\{n \mid ? \neq M(f[n]) \neq M(f[n+1])\} \leq m$.

Ex_m is the collection of all classes which are Ex_m-identifiable by some IIM M.

The case $m = 0$ has been introduced in [13] and is very often referred to as *finite learning*. For all bounds m the inclusion $Ex_m \subset Ex_{m+1}$ is verified in [10], thus revealing an infinite hierarchy $Ex_0 \subset Ex_1 \subset \ldots \subset Ex$ of identification potential. An alternative approach to modifying the *Ex*-model by increasing the constraints is to demand special qualities of the intermediate hypotheses. Of course the desired additional properties should in some sense arise from a natural motivation, a few examples known from former studies are collected in Definition 5. To get an insight into the learning potential of IIMs respecting these properties the reader is referred to [1,3,4,6,11–13,15,21,22].

Definition 5 *Let f be any recursive function, M an IIM, ψ any hypothesis space, $n, m \geq 0$. Assume $M(f[n])$ is defined. The hypothesis $M(f[n])$ is called*

- *consistent for $f[m]$ with respect to ψ iff, for all $x \leq m$, $\psi_{M(f[n])}(x)$ is defined and equals $f(x)$;*
- *conform for $f[m]$ with respect to ψ iff, for all $x \leq m$, either $\psi_{M(f[n])}(x)$ is undefined or $\psi_{M(f[n])}(x) = f(x)$;*
- *convergently incorrect for f with respect to ψ iff $\psi_{M(f[n])} \not\subseteq f$;*
- *total with respect to ψ iff $\psi_{M(f[n])}$ is a total function.*

Consistency is a quite natural demand, because any inconsistent hypothesis is in particular incorrect. But as in general a learner cannot detect an inconsistency in an undefined value, the demand for consistency might be considered too hard and loosened to the demand for conformity.

Another aspect motivates the approach of learning with convergently incorrect intermediate hypotheses: if the learner is construed to maintain its current hypothesis, until a fault is detected, then all hypotheses should either be correct or justify a mind change via a convergently incorrect value. This means in particular, that no hypothesis should denote a proper subfunction of the function to be learned. Otherwise a mind change could not be justified convergently.

Finally, it might be natural to demand only total hypotheses, since any guess corresponding to a non-total function must be wrong anyway.

Definition 6 *Let C be a class of recursive functions, M an IIM and ψ some hypothesis space, such that C is Ex-learned by M with respect to ψ. Then C is Cons-learned (Conf-, Cex-, Total-learned, resp.) by M with respect to ψ, iff, for any f in C and $n \geq 0$, $M(f[n])$ is consistent for $f[n]$ (conform for $f[n]$, either correct or convergently incorrect for f, total, resp.) with respect to ψ. The notions Cons, Conf, Cex, Total are defined as usual.*

There are numerous studies analysing consistent identification; most of the important results, including a proof of $Cons \subset Ex$, can be found in [4,6,13,21,22].

Conform learning is defined by a mitigation of the consistency demands, thus increasing the potential of the relevant learners. Yet there are still learning problems solvable in the *Ex*-model, but not in the *Conf*-model; for a proof of $Cons \subset Conf \subset Ex$ the reader is referred to [21]. More details on conform identification are collected in [12].

[11] supplies the main results concerning *Cex*, in particular, the separations $Cex \subset Ex$, $Cex \# Cons$, and $Cex \# Ex_m$ for all $m \geq 1$ are verified there.

Finally, see [8,9,15] for identification with total intermediate hypotheses. Of relevance for the subsequent sections, most of all the result $Total \subset Cons$, cf. [15], must be mentioned.

The reason to consider so many inference types is the purpose to really corroborate the thought of universal dependencies in inductive inference: it will be shown that known relations between inference types (see Theorem 8) still hold in uniform learning, even when strong separations are considered. To give evidence to this fact, it is necessary to regard a few inference types; the ones defined here have been chosen, because they can be classified into the three kinds of inference types mentioned above:

- Ex_m for $m \geq 0$, Bc, and Bc^* are types resulting from special constraints concerning the success criterion of the sequence of hypotheses (where Bc^* also includes modified accuracy demands);
- *Cex* and *Total* are types resulting from special constraints concerning the quality of the intermediate hypotheses, independent of the amount of information currently known about the target function;
- *Cons* and *Conf* are types resulting from special constraints concerning the quality of the intermediate hypotheses, depending on the information currently known about the target function.

For each kind of inference type the strong separation results will be verified.

The set of all inference classes defined above is denoted by $\mathbb{I}$.

$$\mathbb{I} := \{Ex, Bc, Bc^*, Cons, Conf, Cex, Total\} \cup \{Ex_m \mid m \geq 0\} .$$

Note that the term *inference type* formally refers to a class in $\mathbb{I}$, but informally also to the corresponding underlying learning model.

Lemma 7 alludes to two simple and fundamental results commonly used in the literature, also mentioned in [13] and [22].

Lemma 7 *Let $I \in \mathbb{I}$, $C \in I$ and let τ be any acceptable numbering. Then C*

can be I-learned with respect to the hypothesis space τ by some IIM. Moreover, if $I \notin \{Cons, Conf\}$ and ψ is a hypothesis space, such that C is I-learnable with respect to ψ, then there is some total IIM adequate for I-identification of C with respect to ψ.

For a counterexample to the second statement of Lemma 7, respecting *Cons*-identification, see [22]. Moreover, if that statement held for *Conf*, then it would be possible to show its validity also for *Cons* in contradiction to [22]. Details are omitted.

The following theorem is a summary of the known results concerning the potential of learners obeying different formal constraints.

Theorem 8 *[3,4,6,10,11,15,21]*

(1) $Ex_m \subset Ex_{m+1} \subset Ex \subset Bc \subset Bc^*$ *for all* $m \geq 0$, $\{f \mid f \text{ recursive}\} \in Bc^*$,
(2) $Ex_0 \subset Total \subset Cons \subset Conf \subset Ex$,
(3) $Total \subset Cex \subset Ex$,
(4) $Cex \,\#\, Cons$, $Cex \,\#\, Conf$,
(5) $Ex_m \,\#\, I$ *for all* $m \geq 1$ *and all* $I \in \{Total, Cex, Cons, Conf\}$.

For most of the results references have been given above. The verification of $Ex_0 \subset Total$ is straightforward, moreover $Total \subset Cex$ follows from $Total \subseteq Cex$, $Total \subset Cons$, and $Cex \setminus Cons \neq \emptyset$. To obtain $Cex \,\#\, Conf$ similar ideas as in the proof of $Cex \,\#\, Cons$ in [11] can be used. Adapting conceptions provided in [11] additionally yields $Ex_m \,\#\, Cons$ and $Ex_m \,\#\, Conf$ for all bounds $m \geq 1$. Finally $Ex_m \,\#\, Total$ can be verified for $m \geq 1$ with the help of the result $Ex_m \,\#\, Cex$ and its proof.

3 The model of uniform learning

3.1 Definitions

The scope of uniform learning is to view the learning conceptions defined above on a meta-level. The formal analysis is not only concerned with the existence of methods solving specific learning problems, but in particular with the question whether such methods can be synthesized in a universal way. So the focus is on *families* of learning problems (here families of classes of recursive functions). Given a representation or description of any of these learning problems, the aim is to effectively determine a strategy solving the particular problem, i. e. to generate an adequate learner. So, from a description of a class of recursive functions, we want to compute a program for a successful

IIM learning the class.

In order to allow for a formal definition of uniform learning it is first of all necessary to agree on a scheme for describing classes of recursive functions (i. e. a scheme for describing learning problems). For that purpose from now on a fixed three-place acceptable numbering φ is considered. For $d \geq 0$, the numbering φ^d is defined as the function resulting from φ, if the first input is fixed by d. Then any number d corresponds to a two-place numbering φ^d enumerating the set $\{\varphi_i^d \mid i \geq 0\}$ of partial-recursive functions. Now it is conceivable to consider the subset of all total functions in $\{\varphi_i^d \mid i \geq 0\}$ as a learning problem which in particular is uniquely determined by the number d. Thus each number d acts as a description of the set R_d, where

$$R_d := \{\varphi_i^d \mid i \geq 0 \text{ and } \varphi_i^d \text{ is recursive}\} \text{ for any } d \geq 0 .$$

The set R_d is also called the *recursive core* of the numbering φ^d. Moreover each (finite or infinite) set $D = \{d_0, d_1, d_2, \ldots\}$ of natural numbers can be regarded as a set of descriptions and thus as a collection of the learning problems R_{d_0}, R_{d_1}, R_{d_2}, ... In this context, D is called a *description set*.

A *meta-IIM* is an IIM expecting two inputs: firstly, a natural number d interpreted as a description of some recursive core, and secondly, a coding $f[n]$ of an initial segment of some recursive function f. If M is a meta-IIM and d any description, then M_d denotes the IIM resulting from M, when the first input is fixed by d. So a meta-IIM M can also be regarded as some kind of 'computable function' mapping descriptions d to IIMs M_d (usually of course the value set of a computable function does not consist of IIMs, but perhaps of programs for IIMs). Now, if D is any set of natural numbers (i. e. descriptions), M is a uniform learner successful on D, in case M_d learns R_d for all $d \in D$. That means M is supposed to develop a suitable learner from each description in the set D. Following Lemma 7 an acceptable numbering is chosen as a hypothesis space.

Definition 9 *Let $I \in \mathbb{I}$ and let D be a set of natural numbers. Fix an acceptable numbering τ. D is uniformly I-learnable iff there is a meta-IIM M, such that, for any description $d \in D$, the IIM M_d is an I-learner for the class R_d with respect to τ. $UniI$ denotes the collection of all uniformly I-learnable description sets.*

Fixing an acceptable numbering as a hypothesis space is a straightforward idea, because all learnability results will remain valid in any acceptable numbering, i. e. Definition 9 is independent of the choice of τ. But this is not the only suggestive notion of hypothesis spaces in uniform learning. Note that each numbering φ^d enumerates at least all functions in R_d, so a meta-IIM might also try to use φ^d as a hypothesis space when learning R_d. This is just a special case of learning with respect to Definition 9, because φ^d-programs can be uniformly translated into programs in a fixed acceptable numbering. As it goes along with stricter demands than in the $Uni I$-model, this conception is referred to by the term *restricted uniform learning*.

Definition 10 *Let $I \in \mathbb{I}$ and let D be a set of natural numbers. D is uniformly I-learnable with restricted choice of hypothesis spaces iff there is a meta-IIM M, such that, for any description d in D, the IIM M_d is an I-learner for the class R_d with respect to φ^d. $res\,Uni I$ denotes the collection of all description sets which are uniformly I-learnable in this restricted sense.*

Just as there exists a reasonable restriction of the $Uni I$-model, it is also possible to extend the view of uniform learning by weakening the formal constraints of Definition 9. If the meta-IIM is no longer required to synthesize suitable hypothesis spaces for the particular learning problems, $Uni I$-learning is generalized to the model of *extended uniform learning*. Here the meta-IIM must only develop learners from descriptions; it is sufficient if the corresponding adequate hypothesis spaces *exist*.

Definition 11 *Let $I \in \mathbb{I}$ and let D be a set of natural numbers. D is uniformly I-learnable with extended choice of hypothesis spaces iff there is a meta-IIM M, such that, for any description d in D, the IIM M_d is an I-learner for the class R_d with respect to some arbitrary hypothesis space ψ. $ext\,Uni I$ denotes the collection of all description sets which are uniformly I-learnable in this extended sense.*

It is not hard to verify that

$$res\,Uni I \subseteq Uni I \subseteq ext\,Uni I$$

for all $I \in \mathbb{I}$. In general, equality of these classes does not hold, as will be shown implicitly in the subsequent examples and theorems. Hopefully, the following examples give a bit more insight into the concept of uniform learning.

Firstly, consider the whole set of natural numbers as a description set D. As there are numbers d, such that R_d equals the class of all recursive functions (which is not Bc-learnable), this set D cannot be uniformly Bc-learnable. This even holds in the extended model. Therefore $D \notin ext\,Uni I$ for all $I \in \mathbb{I}\backslash\{Bc^*\}$. As there is an IIM M^* which Bc^*-learns the whole class of recursive functions with respect to a given acceptable numbering τ, the meta-IIM M, satisfying $M_d = M^*$ for all d, witnesses $D \in Uni Bc^*$. In contrast to that, D is not $res\,Uni Bc^*$-identifiable (see [23]).

Secondly, define the description set D by

$$D := \{d \mid \varphi^d \text{ is a total recursive numbering}\}\,.$$

This implies that R_d equals the set $\{\varphi_i^d \mid i \geq 0\}$ for all d in D. Now let a meta-

IIM M on input $(d, f[n])$ return the least i satisfying $\varphi_i^d[n] = f[n]$.[3] Then M is a meta-IIM witnessing $D \in \mathit{res\,Uni\,Cons}$ and $D \in \mathit{res\,Uni\,Total}$, and thus also $D \in \mathit{res\,Uni\,I}$ for all $I \in \{\mathit{Conf}, \mathit{Cex}, \mathit{Ex}, \mathit{Bc}, \mathit{Bc}^*\}$. But, given $m \geq 0$, the set D is not $\mathit{ext\,Uni\,Ex}_m$-learnable: obviously, there is some description d in D, such that the corresponding recursive core R_d equals the class $\{f \mid f \text{ is recursive and } f =^* 0^\infty\}$ of all recursive functions of finite support. This class is not Ex_m-learnable and therefore no description set containing d can be uniformly Ex_m-learnable – even in the extended sense.

Thirdly, consider the description set $D := \{d \mid \mathit{card}\,R_d = 1\}$ representing all singleton recursive cores. Defining a meta-IIM M by $M(d, f[n]) = 0$ for all d, n and all recursive functions f shows that $D \in \mathit{ext\,Uni\,I}$ for all $I \in \mathbb{I}$: for any recursive function f there is a numbering ψ, such that $\psi_0 = f$. So, if $d \in D$, $R_d = \{f\}$, then some ψ fulfils $\psi_{M_d(f[n])} = f$ for all $n \geq 0$. But, as can be found in [23], there is no appropriate IIM for uniform Bc-learning or restricted uniform Bc^*-learning of the set D, i. e. $D \notin \mathit{Uni\,Bc}$ and $D \notin \mathit{res\,Uni\,Bc}^*$. An idea for the corresponding proofs can also be found in Example 16 below.

3.2 Helpful results

As in the non-uniform model, identifiability implies the existence of appropriate total meta-IIMs, if neither consistency nor conformity of the intermediate hypotheses is required, cf. Lemma 7. In most cases such total meta-IIMs can be constructed uniformly.

Proposition 12 *Let $I \in \mathbb{I} \setminus \{\mathit{Cons}, \mathit{Conf}\}$. There is a family $(MT^i)_{i \geq 0}$ of meta-IIMs satisfying the following properties.*

(1) A program for MT^i can be uniformly computed from i.
(2) All machines MT^i, $i \geq 0$, are total.
(3) If $D \in \mathit{Uni\,I}$ (or $\mathit{ext\,Uni\,I}$), then there is some $i \geq 0$, such that the machine MT^i learns D with respect to the model $\mathit{Uni\,I}$ ($\mathit{ext\,Uni\,I}$, respectively).
(4) If $I \notin \{\mathit{Cex}, \mathit{Total}\}$ and $D \in \mathit{res\,Uni\,I}$, then there is some $i \geq 0$, such that the machine MT^i learns D with respect to the model $\mathit{res\,Uni\,I}$.

Note that the family $(MT^i)_{i \geq 0}$ may depend on the inference criterion $I \in \mathbb{I} \setminus \{\mathit{Cons}, \mathit{Conf}\}$ chosen in advance. The benefit of Proposition 12 shows in its applications in the proofs of various non-learnability results. If the purpose is to verify that a certain description set is not suitable for uniform learning in some specified model, it will in some cases be sufficient to defeat all recursive IIMs in an indirect proof.

[3] This is the uniform version of Gold's *identification by enumeration*, cf. [13].

The following examples illustrate learning problems for which Proposition 12 does not remain valid.

Example 13 *Let $I \in \{Cons, Conf, Cex, Total\}$; fix a description set D by*

$$D := \{d \mid R_d = \{0^\infty\} \text{ and there is exactly one index } i \text{ such that } \varphi_i^d(0) = 0\}\ .$$

Then D belongs to $res\,Uni I$, but D is not $res\,Uni I$-identifiable by any total meta-IIM.

Example 14 *Let $I \in \{Cons, Conf\}$ and define a description set D by*

$$D := \{d \mid \varphi^d \text{ is a recursive function}\}\ .$$

Then D belongs to $res\,Uni I$, but D is not $ext\,Uni I$-identifiable by any total meta-IIM.

Proofs for both examples are included in [26].

4 Results on learning of unions

As has been alluded to in the introduction, the investigation of unions of uniformly learnable description sets is of particular interest with the prospect of strong separation results. The analysis below concerns the question, what properties regarding two arbitrary description sets D_1, D_2 in $Uni I$ (or in $res\,Uni I$, $ext\,Uni I$) for some $I \in \mathbb{I}$ are sufficient for uniform learnability of the union $D_1 \cup D_2$. Since the whole class of recursive functions is Bc^*-learnable, the classes $Uni Bc^*$ and $ext\,Uni Bc^*$ are closed with respect to the union of sets, whereas in the general case even unions of rather 'simple' description sets yield negative results. The following examples illustrate this fact.

Example 15 *Define two description sets by $D_1 := \{d \mid R_d = \{\varphi_0^d\} = \{0^\infty\}\}$ and $D_2 := \{d \mid$ there is some m such that $R_d = \{\varphi_1^d\} = \{0^m 1^\infty\}\}$. Then both D_1 and D_2 belong to $res\,Uni Ex_0$, but the union $D_1 \cup D_2$ is not contained in $Uni Ex_0$.*

Proof. Let $i \in \{1, 2\}$. Obviously the meta-IIM returning $i - 1$ on any input is successful for D_i according to the $res\,Uni Ex_0$-model. Now assume $D_1 \cup D_2 \in Uni Ex_0$. This implies the existence of some total meta-IIM M and some acceptable numbering τ, such that R_d is Ex_0-learned by M_d with respect to τ, whenever d belongs to $D_1 \cup D_2$. A contradiction can be revealed by constructing a description $d^* \in D_1 \cup D_2$, such that M_{d^*} fails to Ex_0-identify R_{d^*} in the hypothesis space τ. The construction of d^* proceeds as follows.

First for each number d a numbering ψ is defined, such that its recursive core R fulfils

(1) $R = \{\psi_0\} = \{0^\infty\}$ or $R = \{\psi_1\} = \{0^m 1^\infty\}$ for some m,
(2) M_d does not learn R in τ according to the Ex_0-model.

The recursion theorem (see [20]) then supplies some fixed point value d^* which serves as a description for exactly the numbering ψ constructed from d^* (i. e. $\varphi^{d^*} = \psi$; in particular, R_{d^*} equals the recursive core R of ψ). The properties (1) and (2) then imply $d^* \in D_1 \cup D_2$ and M_{d^*} does not learn R_{d^*} in τ with respect to the Ex_0-model. This provides the desired contradiction.

More formally: for any number d a partial-recursive numbering ψ is constructed as follows. Start defining $\psi_0(x) = 0$ for gradually growing x; in parallel look for some number $x \geq 1$ satisfying

$$M_d(0^x) \neq ? \text{ and } \tau_{M_d(0^x)}[x] = 0^{x+1} .$$

Case A. Such a number x exists. Then let m be the first such number x found; stop defining ψ_0 any further, i. e. $\psi_0 = 0^z \uparrow^\infty$ for some z. Instead let $\psi_1 = 0^m 1^\infty$. Moreover $\psi_i = \uparrow^\infty$ for all $i \geq 2$.

Remark. If no such number x exists, then the search for m does not terminate. Hence $\psi_0 = 0^\infty$ and $\psi_i = \uparrow^\infty$ for $i \geq 1$. *End construction of ψ.*

As the whole construction is uniformly effective in d, the recursion theorem supplies some number d^*, such that φ^{d^*} equals the numbering ψ constructed from d^*.
If case A does not occur in the definition of ψ, then, by the remark above, $R_{d^*} = \{\varphi_0^{d^*}\} = \{0^\infty\}$, so d^* belongs to D_1. In this case there is no $x \geq 1$, such that $M_{d^*}(0^x) \neq ?$ and $\tau_{M_{d^*}(0^x)} = 0^\infty$. Therefore R_{d^*} is not Ex_0-learned by M_{d^*} in τ.
If case A occurs in the definition of ψ, then $R_{d^*} = \{\varphi_1^{d^*}\} = \{0^m 1^\infty\}$, where m is the first number found in the corresponding construction. In particular, d^* belongs to D_2. Moreover $M_{d^*}(0^m) \neq ?$ and $\tau_{M_{d^*}(0^m)}[m] = 0^{m+1}$. This implies $\tau_{M_{d^*}(0^m)} \neq 0^m 1^\infty$, so M_{d^*} makes a wrong guess on $0^m 1^\infty$ with respect to τ. Therefore again R_{d^*} is not Ex_0-learned by M_{d^*} in τ.

Since both cases result in a contradiction, the assumption $D_1 \cup D_2 \in UniEx_0$ must be wrong. This completes the proof. □

For the next example recall that in finite numberings almost all programs correspond to the empty function $\uparrow^\infty$.

Example 16 *Define description sets D, D_1, D_2 by $D := \{d \mid R_d = \{\varphi_0^d\}\}$, $D_1 := \{d \mid card R_d = card\{i \mid \varphi_i^d$ is recursive$\} = 2$ and φ^d is finite$\}$, as well* *as* $D_2 := \{d \mid card\{i \mid \varphi_i^d \text{ is recursive}\} = 1 \text{ and } \varphi^d \text{ is finite}\}$*. Then* $D \in$ *resUniEx*$_0$*, moreover* D_1 *and* D_2 *belong to resUniCons. In contrast to that* $D \cup D_1 \notin$ *extUniEx,* $D \cup D_2 \notin$ *UniBc, and* $D \cup D_2 \notin$ *resUniBc**.

Proof. D is *resUniEx*$_0$-learned by the meta-IIM constantly zero. To identify functions in some R_d, where d belongs to D_1 or D_2, let a meta-IIM simulate a learner that – given $f[n]$ – returns some arbitrary program consistent for $f[n]$ with respect to φ^d. By definition of D_1 and D_2 this method must be successful according to the *resUniCons*-constraints.

Next assume $D \cup D_1 \in$ *extUniEx* via some total meta-IIM M. As in the proof of Example 15, from any number d a numbering ψ is constructed uniformly, such that some fixed point value $d^* \in D \cup D_1$ provides a contradiction, namely that R_{d^*} is not *Ex*-learned by M_{d^*} (in any hypothesis space). Since the argumentation is similar to that in the proof of Example 15, just the crucial ideas are presented.

For each number d construct a two-place function ψ in stages. In stage 0 let $\psi_0(0) = 0$, $n_1 = 0$, and go to stage 1. In each stage k, for $k \geq 1$, proceed as follows.

Let $\psi_{2k-1}[n_k+1] = \psi_0[n_k]0$, $\psi_{2k}[n_k+1] = \psi_0[n_k]1$; moreover consider the guess $M_d(\psi_0[n_k])$ as the current hypothesis of M_d. Then extend both ψ_{2k-1} and ψ_{2k} with a sequence of the value 0, until M_d changes its mind on at least one of the two extensions.

Case A. After m steps M_d changes its mind on some extension. Let $n_{k+1} = n_k + m + 2$.

Case A.1. $M_d(\psi_0[n_k]00^m) \neq M_d(\psi_0[n_k])$.
Then define $\psi_0[n_{k+1}] = \psi_0[n_k]00^m2$. Moreover suspend the definition of the functions ψ_{2k-1} and ψ_{2k} forever and go to stage $k+1$.

Case A.2. $M_d(\psi_0[n_k]00^m) = M_d(\psi_0[n_k])$.
Then $M_d(\psi_0[n_k]10^m) \neq M_d(\psi_0[n_k])$. Let $\psi_0[n_{k+1}] = \psi_0[n_k]10^m2$ and let both functions ψ_{2k-1} and ψ_{2k} remain initial; go to stage $k+1$.

(* In case A the IIM M_d changes its mind on the extension of ψ_0 constructed in stage k. *)

Remark 1. If M_d never changes its mind on any of the extensions (i. e. case A does not occur), then stage k does not terminate. In this case ψ_0 is not defined any further, i. e. $\psi_0 = \psi_0[n_k] \uparrow^\infty$, $\psi_{2k-1} = \psi_0[n_k]00^\infty$, $\psi_{2k} = \psi_0[n_k]10^\infty$, and $\psi_i = \uparrow^\infty$ for all $i > 2k$. In particular, the recursive core of ψ equals the set $\{\psi_{2k-1}, \psi_{2k}\}$. As $M_d(\psi_{2k-1}) = M_d(\psi_{2k}) = M_d(\psi_0[n_k])$, the IIM M_d is not *Ex*-successful for this recursive core. Moreover note that $2k-1$ and $2k$ are the only ψ-programs of total functions in this case. *End stage k.*

Now M_d does not *Ex*-learn the recursive core R of ψ in any hypothesis space: if case A never occurs in any stage of the definition of ψ, see the corresponding remark 1. If case A always occurs, then all stages $k \geq 1$ are reached. This implies $R = \{\psi_0\}$ and, by the note in case A, M_d changes its mind on ψ_0 infinitely often. In particular, M_d cannot *Ex*-identify R.

Next, recall that there must be some d^*, such that φ^{d^*} equals the numbering ψ constructed from d^*. With the help of the above remarks it is not hard to verify that d^* belongs to $D \cup D_1$ and that R_{d^*} is not *Ex*-learned by M_{d^*}. So $D \cup D_1 \notin \mathit{ext\,UniEx}$.

A similar construction verifies $D \cup D_2 \notin \mathit{UniBc}$. Given a total meta-IIM M and some acceptable numbering τ, for each number d again some numbering ψ is constructed in stages. In stage 0, let $\psi_0(0) = 0$, $n_1 = 0$, and go to stage 1. In stage k, $k \geq 1$, let $\psi_k[n_k] = \psi_0[n_k]$ and extend ψ_k with zeros, until some number m is found, such that

$$\tau_{M_d(\psi_0[n_k]0^m)}(n_k + m + 1) \text{ is defined and equals } 0\,.$$

If such an m is found, suspend the definition of ψ_k forever. In this case let $n_{k+1} = n_k + m + 1$, $\psi_0[n_{k+1}] = \psi_0[n_k]0^m1$ and go to stage $k+1$. Note that now $M_d(\psi_0[n_k]0^m)$ is a wrong guess for ψ_0 with respect to τ. In case no such number m exists, stage k is never left and $\psi_k = \psi_0[n_k]0^\infty$, whereas $\psi_i =\uparrow^\infty$ for all $i > k$. This implies that none of the hypotheses $M_d(\psi_k[n])$, for $n \geq n_k$, is a τ-program for ψ_k.

Via the usual argumentation this construction will prove $D \cup D_2 \notin \mathit{UniBc}$.

Finally, if $D \cup D_2$ was $\mathit{res\,UniBc}^*$-identifiable, then also $D \cup D_2 \in \mathit{UniBc}$ would hold (cf. a proof in [23]) – a contradiction. □

These two examples first of all show that in general the classes UniI, $\mathit{res\,UniI}$, and $\mathit{ext\,UniI}$, resulting from different models of uniform learning, are not closed with respect to the union of description sets. That means, considering two learnable description sets, we cannot be sure that their union is also uniformly learnable. This explains a need for additional conditions concerning the two description sets, such that learnability of the union can be guaranteed. Theorem 17 proposes a first conception in view of that purpose.

Theorem 17 *Let $I \in \mathbb{I}$ and assume D_1 and D_2 are description sets in UniI (or in $\mathit{res\,UniI}$, $\mathit{ext\,UniI}$, respectively).*

(1) If D_1 is recursive, then the union $D_1 \cup D_2$ belongs to UniI (or $\mathit{res\,UniI}$, $\mathit{ext\,UniI}$, respectively).

(2) *If both* D_1 *and* D_2 *are r. e., then the union* $D_1 \cup D_2$ *belongs to* $UniI$ *(or* $res\,UniI$*,* $ext\,UniI$*, respectively).*
(3) *If* $I \in \{Ex, Bc, Bc^*\}$ *and at least one of the sets* D_1*,* $\overline{D_2}$*,* $D_1 \setminus D_2$ *is r. e., then the union* $D_1 \cup D_2$ *belongs to* $UniI$ *(or* $res\,UniI$*,* $ext\,UniI$*, respectively).*

Proof. Fix $I \in \mathbb{I}$ and $D_1, D_2 \in UniI$. For the restricted and the extended models all proofs proceed analogously. Choose meta-IIMs M^1 and M^2 appropriate for $UniI$-learning of D_1 and D_2, respectively.

ad (1). Assume D_1 is recursive. Given any number d, let a meta-IIM M simulate M^1, if d belongs to D_1, and M^2, otherwise. Obviously M witnesses assertion 1. *qed (1).*

ad (2). Assume that both D_1 and D_2 are r. e. Given d, let a meta-IIM M search in the value sets of some enumerations of D_1 and D_2, until d is found to belong to some D_i ($i \in \{1, 2\}$). Afterwards M can simulate M^i, so M is a $UniI$-learner for $D_1 \cup D_2$. *qed (2).*

ad (3). Let $I \in \{Ex, Bc, Bc^*\}$ and assume that D_1 ($\overline{D_2}$, $D_1 \setminus D_2$) is r. e. via some effective enumeration π. Let d be a natural number. The desired meta-IIM just has to simulate M^2, as long as d is not found in the value set of π. Afterwards M^1 can be simulated. *qed (3).* □

Note that the third assertion of Theorem 17 is not stated for all of the inference types in $\mathbb{I}$. As will be shown in Theorem 20, this assertion really does not hold in the general case. Still, if learning with consistent intermediate hypotheses is considered, at least parts of the statement remain valid.

Theorem 18 *Let* D_1 *and* D_2 *be description sets, such that* D_1 *or* $\overline{D_2}$ *or* $D_1 \setminus D_2$ *is r. e. Moreover assume* $D_2 \in Uni\,Cons$*.*

(1) *If* $D_1 \in ext\,Uni\,Cons$*, then* $D_1 \cup D_2 \in ext\,Uni\,Cons$*.*
(2) *If* $D_1 \in Uni\,Cons$*, then* $D_1 \cup D_2 \in Uni\,Cons$*.*

Proof. Let π be a partial-recursive function with the value set D_1 (or $\overline{D_2}$ or $D_1 \setminus D_2$). As D_2 belongs to $Uni\,Cons$, there must be an acceptable numbering τ as well as a meta-IIM M^2, such that M^2_d learns the recursive core R_d consistently with respect to τ, whenever $d \in D_2$.

ad (1). Assume $D_1 \in ext\,Uni\,Cons$. This implies the existence of numberings $\psi^{[d]}$ ($d \in D_1$) and a meta-IIM M^1 appropriate for uniform *Cons*-learning of D_1 in the following sense: if d belongs to D_1, then M^1_d learns the recursive core R_d with consistent intermediate hypotheses with respect to $\psi^{[d]}$. Now it remains to define new hypothesis spaces $\eta^{[d]}$ for $d \in D_1 \cup D_2$ and a new meta-IIM M suitable for uniform *Cons*-learning of $D_1 \cup D_2$.

The new hypothesis spaces just result from 'mixing' the numberings $\psi^{[d]}$ with the acceptable numbering τ. Thus the hypotheses of both M^1 and M^2 can be translated into the new numberings, if necessary. Formally let

$$\eta^{[d]}_{2i} := \begin{cases} \psi^{[d]}_i\,, & \text{if } d \in val(\pi)\,, \\ 4^\infty\,, & \text{otherwise}\,, \end{cases}$$
$$\eta^{[d]}_{2i+1} := \tau_i\,,$$

for all i.

The idea for the new meta-learner, given a number d, is to translate the hypotheses of M^1 into the numbering $\eta^{[d]}$, as soon as d has been found in the value set of π. In parallel with the membership test for d with respect to $val(\pi)$ the learner just checks the hypothesis of M^2 for consistency in the hypothesis space τ. A positive consistency test allows to translate the corresponding hypothesis of M^2 into $\eta^{[d]}$. So define

$$M_d(f[n]) := \begin{cases} 2M^1_d(f[n])\,, & \text{if test A stops within } n \text{ steps or} \\ & \text{if test A stops before test B}\,, \\ 2M^2_d(f[n]) + 1\,, & \text{otherwise}\,, \end{cases}$$

where test A and test B work as follows.

Test A. Enumerate the value set of π and stop as soon as d is listed.

Test B. Compute $M^2_d(f[n]) = j$, then $\tau_j(0), \ldots, \tau_j(n)$. Stop in case j is consistent for $f[n]$ with respect to τ.

Now, if d belongs to $D_2 \setminus D_1$ and f is any element of R_d, test A will never stop for any n, whereas test B will always stop. So $M_d(f[n]) = 2M^2_d(f[n]) + 1$ for all n. By choice of M^2 and definition of $\eta^{[d]}$ the IIM M_d then learns f with consistent hypotheses with respect to $\eta^{[d]}$.

If d belongs to the value set of π, then M^1_d is a *Cons*-learner for R_d with respect to $\psi^{[d]}$. Since test A will stop within a fixed number of steps, the first case in the definition of M_d will be relevant in the limit for all functions in R_d, i. e. M_d learns R_d with respect to $\eta^{[d]}$ in the limit. All intermediate hypotheses must also be consistent: in each step either the hypothesis of the suitable *Cons*-learner M^1_d is translated, or the hypothesis of M^2_d is translated after a positive consistency check in test B.

All in all M is a meta-IIM appropriate for uniform *Cons*-learning of the set $D_1 \cup D_2$ with respect to the hypothesis spaces $\eta^{[d]}$ defined above. *qed (1).*

ad (2). Assume $D_1 \in \mathit{UniCons}$. Without loss of generality the numbering τ is suitable as a hypothesis space for *UniCons*-identification of D_1, say via some meta-IIM M^1.

The same idea as in assertion (1) shows that the meta-IIM M, defined by

$$M_d(f[n]) := \begin{cases} M_d^1(f[n])\,, & \text{if test A stops within } n \text{ steps or} \\ & \text{if test A stops before test B}\,, \\ M_d^2(f[n])\,, & \text{otherwise}\,, \end{cases}$$

is appropriate for *UniCons*-identification of $D_1 \cup D_2$ with respect to τ. Here test A and test B are defined as above. *qed (2).* □

So the results of Theorem 17.(3) concerning *Ex*-, *Bc*-, and Bc^*-identification can be transferred to consistent learning in many cases. Considering identification with a bounded number of mind changes, the results are not that straightforward. Still, a simple positive result is obtained, if the demands concerning the number of mind changes allowed for learning the union of two sets are loosened.

Theorem 19 *Let D_1 and D_2 be description sets and fix $m_1, m_2 \geq 0$. If $D_1 \in \mathit{extUniEx}_{m_1}$ and $D_2 \in \mathit{extUniEx}_{m_2}$, then the union $D_1 \cup D_2$ is an element of $\mathit{extUniEx}_{m_1+m_2+1}$.*

Proof. Without loss of generality choose total meta-IIMs M^i ($i \in \{1,2\}$), such that M^i learns D_i according to $\mathit{extUniEx}_{m_i}$ and M_d^i *never* changes its mind more than m_i times (no matter what d and what graph M^i is fed with). In particular, for each $i \in \{1,2\}$ and $d \in D_i$ there is some hypothesis space $\psi^{[d]}$ which is suitable for Ex_{m_i}-learning of R_d via M_d^i.

For each number d now define a numbering $\eta^{[d]}$ by

$$\eta^{[d]}_{\langle y,z\rangle} := \begin{cases} \psi^{[d]}_y\,, & \text{if } d \in D_1\,, \\ \psi^{[d]}_z\,, & \text{if } d \in D_2 \setminus D_1\,, \\ 4^\infty\,, & \text{otherwise}\,, \end{cases}$$

for all y, z. These numberings $\eta^{[d]}$ can be used as hypothesis spaces for a new meta-IIM M which, given d and $f[n]$, works as follows:

$$M_d(f[n]) := \begin{cases} ?\,, & \text{if } M_d^1(f[n]) = M_d^2(f[n]) = ?\,, \\ \langle M_d^1(f[n]), 0\rangle\,, & \text{if } M_d^1(f[n]) \neq ? \text{ and } M_d^2(f[n]) = ?\,, \\ \langle 0, M_d^2(f[n])\rangle\,, & \text{if } M_d^1(f[n]) = ? \text{ and } M_d^2(f[n]) \neq ?\,, \\ \langle M_d^1(f[n]), M_d^2(f[n])\rangle\,, & \text{if } M_d^1(f[n]) \neq ? \text{ and } M_d^2(f[n]) \neq ?\,. \end{cases}$$

Since for any $i \in \{1, 2\}$ and any description d the IIM M_d^i changes its mind at most m_i times on any input sequence, M_d must be appropriate for $Ex_{m_1+m_2+1}$-identification of R_d with respect to the hypothesis space $\eta^{[d]}$. This implies $D_1 \cup D_2 \in ext\,Uni\,Ex_{m_1+m_2+1}$. $\square$

Theorem 19 straightly raises the question whether the upper bound m_1+m_2+1 for the number of mind changes constitutes the optimal result. The following theorem now provides two insights: firstly, the bound m_1+m_2+1 is tight, and secondly, the third assertion of Theorem 17 does not remain valid for uniform identification in the Ex_m-models.

Theorem 20 *Let $m_1, m_2 \geq 0$. There exist description sets D_1 and D_2 such that*

(1) D_i belongs to $res\,Uni\,Ex_{m_i}$ for $i \in \{1, 2\}$,
(2) for all d in $D_1 \cup D_2$ the recursive core R_d consists of at most two functions (at most one function),
(3) D_1 is r. e., but
(4) the union $D_1 \cup D_2$ does not belong to $ext\,Uni\,Ex_{m_1+m_2}$ ($Uni\,Ex_{m_1+m_2}$, respectively).

Proof. Let $(MT^i)_{i \geq 0}$ be the enumeration of total meta-IIMs according to Proposition 12. We only prove the statement for the *ext Uni*-model, the proof for the *Uni*-model uses similar ideas, see also [26].

First a description set D ($= D_1 \cup D_2$) is defined via the uniform construction of partial-recursive functions ψ, each depending on an index i of some MT^i and a description d. By construction, MT_d^i will be inappropriate for learning the recursive core of ψ. The recursion theorem will yield a recursive function fp, such that $\varphi^{fp(i)}$ always equals the numbering ψ constructed from i and $d = fp(i)$. D will be the value set of fp, D_1 some suitable r. e. subset of D, and $D_2 = D \setminus D_1$. In particular, any total meta-IIM MT^i – given the description $fp(i)$ – will fail to identify $R_{fp(i)}$ with no more than $m_1 + m_2$ mind changes. This implies $D_1 \cup D_2 = D \notin ext\,Uni\,Ex_{m_1+m_2}$.

Given i and d, construct a numbering ψ in stages as follows.

In stage 0 let $\psi_0(0) = 0$ and extend ψ_0 with a sequence of zeros, until MT_d^i returns some value different from '?' on a sequence $\psi_0[n_1] = 0^{n_1+1}$ constructed so far.

Case A. n_1 exists (so $MT_d^i(0^{n_1+1}) \neq ?$).
Then let $\psi_0 = 0^{n_1+1}2\uparrow^\infty$, $\alpha_1 = 0^{n_1+1}$, and go to stage 1.

Remark. If $MT_d^i(0^x) = ?$ for all x (i. e. case A does not occur), then stage 0 does not terminate. Hence $\psi_0 = 0^\infty$ and MT_d^i does not *Ex*-identify ψ_0. Moreover

$\psi_i = \uparrow^\infty$ for all $i \geq 0$.

In stage k, $1 \leq k \leq m_1 + m_2$, let $\psi_{2k-1}[n_k + 1] = \alpha_k 0$, $\psi_{2k}[n_k + 1] = \alpha_k 1$. Extend both ψ_{2k-1} and ψ_{2k} with zeros, until MT^i_d changes its mind on either extension (i. e. on some segment $\alpha_k 00^y$ or $\alpha_k 10^y$).

Case A. y exists.
Then let $\psi_{2k-1} = \alpha_k 00^y 2 \uparrow^\infty$ and $\psi_{2k} = \alpha_k 10^y 2 \uparrow^\infty$; define $n_{k+1} = n_k + y + 1$ and

$$\alpha_{k+1} := \begin{cases} \alpha_k 00^y\,, & \text{if } MT^i_d(\alpha_k 00^y) \neq MT^i_d(\alpha_k)\,, \\ \alpha_k 10^y\,, & \text{if } MT^i_d(\alpha_k 00^y) = MT^i_d(\alpha_k)\,, \\ & (\text{so } MT^i_d(\alpha_k 10^y) \neq MT^i_d(\alpha_k))\,. \end{cases}$$

Go to stage $k + 1$.

Remark. If MT^i_d never changes its mind on any of these extensions (i. e. case A does not occur), then stage k does not terminate. Hence $\psi_{2k-1} = \alpha_k 00^\infty$, $\psi_{2k} = \alpha_k 10^\infty$, but

$$MT^i_d(\psi_{2k-1}) = MT^i_d(\psi_{2k}) = MT^i_d(\alpha_k)\,,$$

that means MT^i_d does not *Ex*-learn $\{\psi_{2k-1}, \psi_{2k}\}$. Moreover ψ_i is initial for all $i \notin \{2k - 1, 2k\}$.

In stage $m_1 + m_2 + 1$ finally define $\psi_{2m_1+2m_2+1} = \alpha_{m_1+m_2+1}00^\infty$ as well as $\psi_{2m_1+2m_2+2} = \alpha_{m_1+m_2+1}10^\infty$. In addition let $\psi_x = \uparrow^\infty$ remain undefined for $x > 2m_1 + 2m_2 + 2$. *End construction of* ψ.

By the recursion theorem there is a recursive function *fp*, such that $\varphi^{fp(i)}$ equals the numbering ψ constructed from i and $fp(i)$, whenever $i \geq 0$. Now define $D := \{fp(i) \mid i \geq 0\}$. Note that even D is r. e. Moreover let

$$D_1 := \{fp(i) \mid i \geq 0 \text{ and } 2 \in val(\varphi^{fp(i)}_{2m_2})\}\,,\ D_2 := D \setminus D_1\,.$$

Obviously, D_1 is r. e. and for all $d \in D_1 \cup D_2$ the recursive core R_d consists of at most two functions. A possible strategy for *Ex*-learning such a recursive core is to look for the value 2 in the value set of the numbering φ^d and always to return the minimal φ^d-program, for which the value 2 has not yet been found in the corresponding function.

If $d \in D_2$, then stage $m_2 + 1$ is not reached. By construction, any IIM using the strategy above will be successful for R_d in the sense of Ex_{m_2}.

If $d \in D_1$, then stage $m_2 + 1$ is reached. As there are at most m_1 stages left, an IIM applying the method explained above to all programs greater than $2m_2$ will Ex_{m_1}-identify the class R_d.

Finally, if $D_1 \cup D_2$ was in $ext\,Uni\,Ex_{m_1+m_2}$, there would be some number i, such that MT^i is an appropriate meta-IIM for $D_1 \cup D_2$ according to $ext\,Uni\,Ex_{m_1+m_2}$. Now let $d = fp(i)$. The construction of ψ then implies that R_d is not $Ex_{m_1+m_2}$-learned by MT^i_d (with respect to any hypothesis space). To verify this consider the following argumentation.

Firstly, if stage $m_1 + m_2 + 1$ is reached in the corresponding construction, then MT^i_d must change its hypothesis at least $m_1 + m_2 + 1$ times on one of the functions $\varphi^d_{2m_1+2m_2+1}$, $\varphi^d_{2m_1+2m_2+2}$ in order to identify both. The reason for this is that the current segment $\alpha_{m_1+m_2+1}$ is constructed to force MT^i_d into $m_1 + m_2$ mind changes. Note that the two distinct functions $\varphi^d_{2m_1+2m_2+1}$ and $\varphi^d_{2m_1+2m_2+2}$ both have the initial segment $\alpha_{m_1+m_2+1}$ in common.

Secondly, if stage m_1+m_2+1 is not reached in the corresponding construction, then case A was not fulfilled at some stage before, so MT^i_d does not Ex-learn the recursive core R_d.

Thus $D_1 \cup D_2 \notin ext\,Uni\,Ex_{m_1+m_2}$. □

Corollary 21 *Let $D := \{d \mid R_d = \{\varphi^d_0\}\}$. Then D belongs to $res\,Uni\,Ex_0$ and the following assertions hold.*

(1) If $I \in \mathbb{I} \setminus \{Bc, Bc^*\}$, there is some D' in $res\,Uni\,I$ satisfying $D \cup D' \notin ext\,Uni\,I$. Moreover, D' can be chosen such that, for any $d \in D'$, R_d consists of at most two functions.
(2) If $I \in \mathbb{I} \setminus \{Bc^*\}$, there is some D' in $res\,Uni\,I$ satisfying $D \cup D' \notin Uni\,I$. Moreover, D' can be chosen such that, for any $d \in D'$, R_d is a singleton.
(3) For all $I \in \mathbb{I}$ there is some D' in $res\,Uni\,I$ satisfying $D \cup D' \notin res\,Uni\,I$. Moreover, D' can be chosen such that, for any $d \in D'$, R_d is a singleton.

Proof. Given $I = Ex_m$ for some $m \geq 0$ the statement is verified as Theorem 20, where $m_1 = 0$ and $m_2 = m$. In all other cases the assertions are immediate consequences of Example 16. □

Up to now we have discussed several conditions sufficient for the uniform learnability of unions of description sets. In particular, the condition proposed in Theorem 17.(3) for the inference types Ex, Bc, and Bc^* turned out not binding for some other inference types. This might at first suggest that the conditions in Theorem 17 are quite demanding in the sense that it is not easy to formulate much weaker sufficient conditions. But a second glance reveals some kind of inaccuracy of Theorem 17: the properties required there always concern only the structure of the description sets without alluding to the structure of the corresponding recursive cores. Thus it is easy to find two simple description sets D_1, D_2, which are *not* r. e. (so the condition in Theorem 17.(3) is *not* fulfilled), but still the union $D_1 \cup D_2$ belongs to $res\,Uni\,Ex_0$. For illustration consider the following example: let X be any set of natural numbers such that neither X nor $\overline{X}$ is r. e. Moreover define

$$D_1 := \{d \mid R_d = \{\varphi_0^d\} = \{0i^\infty\} \text{ for some } i \in X\}\,,$$
$$D_2 := \{d \mid R_d = \{\varphi_1^d\} = \{1i^\infty\} \text{ for some } i \in \overline{X}\}\,.$$

The sets D_1, D_2 are not r. e., because otherwise X, $\overline{X}$ were r. e. But the meta-IIM M constructed to make $M_d(f[n])$ always return $f(0)$ witnesses to $D_1 \cup D_2 \in \mathit{resUniEx}_0$. The reason is that the specific form of the functions in the recursive cores indicates which of the two sets the current description belongs to. So a successful meta-IIM does not need to exploit any special structures of the description sets.

These observations propose the choice of sufficient conditions much weaker than those in Theorem 17, based on the aim to use *both* the specific structure of the description sets *and* the specific information provided by the functions in the corresponding recursive cores. These two parts form the information presented in the learning process, so they can both be exploited by successful meta-IIMs. Definition 22 suggests some notions useful in that sense.

Definition 22 *Let D_1 and D_2 be two description sets of natural numbers. Then a computable function e is said to fulfil*

- *Property α, iff*

(1) $e(d, f(0)) = 0$ for all descriptions $d \in D_1$ and all functions $f \in R_d$,
(2) $e(d, f(0)) = 1$ for all descriptions $d \in D_2 \setminus D_1$ and all functions $f \in R_d$;

- *Property β, iff*

(1) for all descriptions $d \in D_1$ and all functions $f \in R_d$ there is some number n, such that $e(d, f[n]) = 0$ and $e(d, f[n']) =?$ whenever $n' < n$,
(2) for all descriptions $d \in D_2 \setminus D_1$ and all functions $f \in R_d$ there is some number n, such that $e(d, f[n]) = 1$ and $e(d, f[n']) =?$ whenever $n' < n$;

- *Property γ, iff*

(1) for all descriptions $d \in D_1$ and all functions $f \in R_d$ there is some number n, such that $e(d, f[n']) = 0$ whenever $n' \geq n$,
(2) for all descriptions $d \in D_2 \setminus D_1$ and all functions $f \in R_d$ there is some number n, such that $e(d, f[n']) = 1$ whenever $n' \geq n$.

D_1 and D_2 possess Property α (Property β, Property γ), iff there is a computable function e satisfying Property α (Property β, Property γ, respectively) for D_1 and D_2.

Note that Property α implies Property β and if some computable function satisfies Property β for D_1 and D_2, this implies the existence of some computable function satisfying Property γ for D_1 and D_2.

Moreover, these properties are associated to the demands of Theorem 17 as follows:

- if D_1 is recursive and D_2 is any arbitrary description set, then D_1 and D_2 possess Property α.
- if D_1 and $D_2 \setminus D_1$ are r. e., then D_1 and D_2 possess Property β.
- if D_1 or $\overline{D_1}$ or $D_2 \setminus D_1$ is r. e., then D_1 and D_2 possess Property γ.

Now let us return to the example above, where X is a set of numbers, X, $\overline{X}$ are not r. e., D_1 consists of all descriptions d, such that $R_d = \{\varphi_0^d\} = \{0i^\infty\}$ for some $i \in X$, and D_2 consists of all descriptions d, such that $R_d = \{\varphi_1^d\} = \{1i^\infty\}$ for some $i \in \overline{X}$. This example reveals that the properties of Definition 22 definitely weaken the conditions in Theorem 17. The function e given by $e(d, f(0)) = f(0)$ witnesses to Property α for D_1 and D_2, whereas none of the conditions in Theorem 17 are fulfilled. Now the sufficiency of Properties α, β, γ for uniform learning of unions of description sets holds as follows.

Theorem 23 *Let D_1 and D_2 be two description sets of natural numbers and $I \in \mathbb{I}$.*

(1) If both D_1 and D_2 belong to $resUniI$ and D_1, D_2 possess Property α, then $D_1 \cup D_2$ belongs to $resUniI$.
(2) If both D_1 and D_2 belong to $UniI$ (or $extUniI$) and D_1, D_2 possess Property β, then $D_1 \cup D_2$ belongs to $UniI$ (or $extUniI$, respectively).
(3) Let $I \in \{Ex, Bc, Bc^*\}$. If both D_1 and D_2 belong to $UniI$ (or $resUniI$, $extUniI$, respectively) and D_1, D_2 possess Property γ, then $D_1 \cup D_2$ belongs to $UniI$ (or $resUniI$, $extUniI$, respectively).

The proof is straightforward and hence omitted. The third assertion of Theorem 23 does not hold for $I = Ex_0$, as Example 15 shows: here both D_1 and D_2 belong to $resUniEx_0$. Defining $e(d, 0^n) = 0$ and $e(d, 0^n1^m) = 1$ for all d, n, m yields a computable function satisfying Property γ for D_1 and D_2. Still the union $D_1 \cup D_2$ does not even belong to $UniEx_0$. Following the first two assertions of Theorem 23, this illustrates the existence of pairs of description sets possessing Property γ, but possessing neither Property β nor Property α.

A final remark in this section is to be made on the relevance of Properties α, β, and γ. Similar to the demonstration below Corollary 21 it is also possible to construct two simple description sets D_1 and D_2, such that $D_1 \cup D_2$ belongs to $resUniEx_0$, although no computable function e can satisfy Property γ (or β or α). For that purpose let X be any set of natural numbers which is Σ_4-complete in the arithmetical hierarchy (cf. [20]). Moreover define

$$D_1 := \{d \mid R_d = \{\varphi_0^d\} = \{i^\infty\} \text{ for some } i \in X\}\,,$$
$$D_2 := \{d \mid R_d = \{\varphi_0^d\} = \{i^\infty\} \text{ for some } i \in \overline{X}\}\,.$$

The proof of $D_1 \cup D_2 \in \mathit{res\,UniEx}_0$ is straightforward. If there was a computable function e satisfying Property γ for D_1 and D_2, then

- $e(d, i^n) = 0$ for all but finitely many n, if $d \in D_1$ and $\varphi_0^d = i^\infty$ (in particular, $i \in X$);
- $e(d, i^n) = 1$ for all but finitely many n, if $d \in D_2$ and $\varphi_0^d = i^\infty$ (in particular, $i \in \overline{X}$).

Choose a total recursive function g such that

$$\varphi_0^{g(i)} = i^\infty \text{ and } \varphi_{j+1}^{g(i)} = \uparrow^\infty$$

for all i, j. This implies that $g(i)$ belongs to D_1 if and only if $i \in X$; analogously $g(i)$ belongs to D_2 if and only if $i \in \overline{X}$. Therefore

$$\begin{aligned} i \in X &\iff e(g(i), i^n) = 0 \text{ for all but finitely many } n \\ &\iff \text{there is an } n_0 \text{ such that for all } n \geq n_0 \text{ there is an } s \text{ such} \\ &\qquad \text{that } e(g(i), i^n) \text{ is computed in } s \text{ steps and } e(g(i), i^n) = 0\,. \end{aligned}$$

Hence X belongs to Σ_3 in the arithmetical hierarchy, in contradiction to the choice of X. So there is no computable function e satisfying Property γ (or β or α) for D_1 and D_2.

Despite this example, the properties in Definition 22 are of importance, because they will be applicable in the proofs of the strong separations in the following section.

5 Strong separation results

This section provides the desired strong separations and thus the main results of this paper. Informally, the statements of the subsequent theorems can be summarized as follows:

(1) almost all pairs of inference types are strongly separable, but there are definitely also pairs, which cannot be separated;
(2) all strong separations can even be witnessed by a fixed r. e. description set D^*;
(3) almost all separations in the original and the restricted model of uniform learning are achieved, if D^* describes only singletons – but there are also exceptions;

(4) all separations in the extended model of uniform learning are achieved, if D^* describes only recursive cores of up to two elements.

In particular, the results depend on the restrictions concerning the choice of hypothesis spaces in uniform learning.

Theorem 24 concerns the strong separations in the $Uni\,I$-model and even in the restricted cases, except for the separation of the inference types *Total* and Ex_m. The latter exception is handled in Theorem 25; proofs are discussed below at the end of this section.

Theorem 24 *Let $I, J \in \mathbb{I}$ fulfil $I \setminus J \neq \emptyset$. Then there is a description set D^* satisfying*

(1) D^* is r. e.,
(2) for any $d \in D^*$ the recursive core R_d is a singleton,
(3) if D belongs to $Uni\,J \cap Uni\,I$, then $D \cup D^* \in Uni\,I \setminus Uni\,J$,
(4) if D belongs to $res\,Uni\,J \cap res\,Uni\,I$ and $(I, J) \neq (Total, Ex_m)$ for $m \geq 0$, then $D \cup D^* \in res\,Uni\,I \setminus Uni\,J$.

Theorem 25 *Let $m \geq 0$. There is a description set D^* satisfying*

(1) D^* is r. e.,
(2) for any $d \in D^*$ the recursive core R_d consists of at most $m+2$ functions,
(3) if D belongs to $res\,Uni\,Ex_m \cap res\,Uni\,Total$, then $D \cup D^* \in res\,Uni\,Total \setminus ext\,Uni\,Ex_m$.

The following fact shows that Theorem 25 provides the best result possible for the separation of *Total* and Ex_m. For the corresponding proof see [26].

Fact 26 *If $D \in res\,Uni\,Total$ and each recursive core R_d, $d \in D$, described by D consists of at most $m + 1$ functions, then $D \in res\,Uni\,Ex_m$.*

Dually to Theorem 24, there are also strong separation results for the extended model of uniform learning, again with a few exceptions handled below.

Theorem 27 *Let $I, J \in \mathbb{I}$ fulfil $I \setminus J \neq \emptyset$, but $(I, J) \neq (Bc^*, Bc)$ and $J \notin \{Cex, Total\}$. Then there is a description set D^* satisfying*

(1) D^* is r. e.,
(2) for any $d \in D^*$ the recursive core R_d consists of at most two functions,
(3) if D belongs to $ext\,Uni\,J \cap ext\,Uni\,I$, then $D \cup D^* \in ext\,Uni\,I \setminus ext\,Uni\,J$.

A separation as in Theorem 27 cannot be achieved for $(I, J) = (Bc^*, Bc)$ or $J \in \{Cex, Total\}$, because the learning potentials of the admissible meta-IIMs coincide in the relevant cases, if the description set represents only finite recursive cores. For a proof of Fact 28 see [23,24]. In particular, this fact shows that the inference types resulting from constraints concerning the quality of the intermediate hypotheses (independent of the amount of information currently known about the target function) constitute an exception within the separation results presented.

Fact 28 *Let D be a description set such that each recursive core R_d for $d \in D$ is a finite set. Then the following assertions hold.*

(1) $D \in \mathit{ext\,Uni\,Bc}$.
(2) $D \in \mathit{ext\,Uni\,Ex}$ *iff* $D \in \mathit{ext\,Uni\,Cex}$ *iff* $D \in \mathit{ext\,Uni\,Total}$.

A non-trivial separation of $\mathit{ext\,Uni\,Bc}^*$ and $\mathit{ext\,Uni\,Bc}$ is even impossible, if descriptions of infinite recursive cores are admitted. The reason is that *every* set D describing only Bc-learnable classes is uniformly Bc-learnable with extended choice of hypothesis spaces, cf. [23]. Whether or not a non-trivial separation of $\mathit{ext\,Uni\,Ex}$, $\mathit{ext\,Uni\,Cex}$, and $\mathit{ext\,Uni\,Total}$ can be witnessed by a description set representing infinite Cex- or $Total$-learnable classes, is not known yet.

The idea for the proofs of Theorem 24, Theorem 25, and Theorem 27 is to carefully construct a recursively enumerable set D^* in $\mathit{Uni\,I} \setminus \mathit{Uni\,J}$ (and correspondingly for the restricted and extended models). By Theorem 17, if $I \in \{Ex, Bc, Bc^*\}$, these properties are already sufficient to obtain $D \cup D^* \in \mathit{Uni\,I} \setminus \mathit{Uni\,J}$ for all $D \in \mathit{Uni\,J} \cap \mathit{Uni\,I}$ (analogously for $\mathit{res\,Uni}$- and $\mathit{ext\,Uni}$-learning). Unfortunately, this does not yet help in case $I \notin \{Ex, Bc, Bc^*\}$. But, as it turns out, it is possible to define D^* such that some computable function e fulfils Property α from Definition 22, where D_1 equals D^* and D_2 is any description set. Then Theorem 23 yields the desired consequences.

For example, if $I = \mathit{Conf}$ and $J = \mathit{Cons}$, Theorem 24 can be verified by constructing a set D^* satisfying

- D^* is r. e.,
- for any $d \in D^*$ the recursive core R_d is a singleton,
- $D^* \in \mathit{res\,Uni\,Conf} \setminus \mathit{Uni\,Cons}$,
- for any description set D there is some computable function e which fulfils Property α with $D_1 = D^*$ and $D_2 = D$.

The corresponding idea also works for Theorem 27.

Proof of Theorem 24 for $I = \mathit{Conf}$ and $J = \mathit{Cons}$. First the set D^* is defined via the construction of a partial-recursive function ψ and a recursive function fp such that $\varphi^{fp(i)} = \psi^{(i,fp(i))}$ for all numbers i. The function fp assigns to each number i some fixed point value according to the recursion theorem. D^* will be the value set of fp. As usual, τ denotes a fixed acceptable numbering to be used as a hypothesis space for uniform learning.

Definition of D^*. Let $(M^i)_{i\geq 0}$ be any fixed enumeration of all meta-IIMs. Now for each meta-IIM M^i we define numberings $\psi^{(i,d)}$ uniformly in such a manner, that none of the recursive cores of the numberings $\psi^{(i,d)}$, $d \geq 0$, is *Cons*-learned by the IIM M^i_d in τ. For any number i, the construction will then yield some description d^*, such that $\psi^{(i,d^*)} = \varphi^{d^*}$; in particular, the recursive core R_{d^*} of the numbering φ^{d^*} is not *Cons*-learned by the IIM $M^i_{d^*}$ in τ. Finally D^* will contain – for each number i – one such corresponding fixed point value d^*. Hence none of the meta-IIMs M^i, $i \geq 0$, will be suitable for *UniCons*-identification of D^*.

Given i, $d \geq 0$, the numbering $\eta = \psi^{(i,d)}$ is defined in stages according to the following instructions.

In stage 0, let $\eta_0(0) = i$ and go to stage 1. In each stage k, $k \geq 1$, proceed as follows.

Let $\eta_{2k-1}[k+1] = \eta_0[k-1]0(k+1)$ and $\eta_{2k}[k+1] = \eta_0[k-1]1(k+1)$ (the value $k+1$ can be used by a *Conf*-learner as an indicator for the functions η_{2k-1} and η_{2k}). Then extend η_{2k-1} with a sequence of the value $k+1$, until the computations of $M^i_d(\eta_0[k-1])$ and $M^i_d(\eta_0[k-1]0)$ terminate.

Remark 1. If $M^i_d(\eta_0[k-1])$ is undefined or $M^i_d(\eta_0[k-1]0)$ is undefined (i. e. neither case A nor case B below occurs), then stage k does not terminate. This yields $\eta_{2k-1} = \eta_0[k-1]0(k+1)^\infty$ as the only element of the recursive core of $\eta = \psi^{(i,d)}$, but M^i_d does not identify η_{2k-1}.

Case A. $M^i_d(\eta_0[k-1])$ and $M^i_d(\eta_0[k-1]0)$ are defined and $M^i_d(\eta_0[k-1]) \neq M^i_d(\eta_0[k-1]0)$.
In this case let $\eta_0(k) = 0$; η_{2k-1} and η_{2k} remain initial. Go to stage $k+1$.
(* Note that the IIM M^i_d changes its mind on the extension of η_0 defined in this case. *)

Case B. $M^i_d(\eta_0[k-1])$ and $M^i_d(\eta_0[k-1]0)$ are defined and equal.
Then extend η_{2k-1} with a sequence of the value $k+1$, until the computation of $\tau_{M^i_d(\eta_0[k-1])}(k)$ stops with the output 0.

> *Remark 2.* If $\tau_{M^i_d(\eta_0[k-1])}(k)$ is undefined or differs from 0 (i. e. case B.1 below does not occur), then stage k does not terminate. This yields $\eta_{2k-1} = \eta_0[k-1]0(k+1)^\infty$ as the only element of the recursive core of $\eta = \psi^{(i,d)}$, but the hypothesis $M^i_d(\eta_{2k-1}[k])$ $(= M^i_d(\eta_0[k-1]))$ is not consistent for $\eta_{2k-1}[k]$ with respect to τ. In particular, M^i_d does not *Cons*-learn η_{2k-1} in τ.

> *Case B.1.* $\tau_{M^i_d(\eta_0[k-1])}(k) = 0$.
> Then let η_{2k-1} remain initial and extend η_{2k} with a sequence of the value $k+1$, until the computation of $M^i_d(\eta_0[k-1]1)$ terminates.

Remark 3. If $M_d^i(\eta_0[k-1]1)$ is undefined (i. e. neither case B.1.1 nor case B.1.2 below occurs), then stage k does not terminate. Hence $\eta_{2k} = \eta_0[k-1]1(k+1)^\infty$ is the only element of the recursive core of $\eta = \psi^{(i,d)}$, but M_d^i does not identify η_{2k}.

Case B.1.1. $M_d^i(\eta_0[k-1]1)$ is defined and differs from $M_d^i(\eta_0[k-1]0) = M_d^i(\eta_0[k-1])$.
Then define $\eta_0(k) = 1$; η_{2k} remains initial. Go to stage $k+1$.
(* Note that M_d^i changes its mind on the extension of η_0 constructed in this case. *)

Case B.1.2. $M_d^i(\eta_0[k-1]1)$ is defined and equal to $M_d^i(\eta_0[k-1])$.
Let $\eta_{2k} = \eta_0[k-1]1(k+1)^\infty$ be the only element of the recursive core of $\eta = \psi^{(i,d)}$.
(* Here $M_d^i(\eta_{2k}[k])$ $(= M_d^i(\eta_0[k-1]))$ is not consistent for $\eta_{2k}[k]$ with respect to τ, because $\tau_{M_d^i(\eta_0[k-1])}(k) = 0$. In particular, M_d^i does not *Cons*-learn the function η_{2k} in τ. *) *End stage k.*

As the construction of $\psi^{(i,d)}$ proceeds uniformly in i and d, there is a recursive function g satisfying $\varphi^{g(i,d)} = \psi^{(i,d)}$ for all $i, d \geq 0$. The recursion theorem then implies the existence of a total recursive function fp such that

$$\varphi^{fp(i)} = \varphi^{g(i,fp(i))} = \psi^{(i,fp(i))}$$

for all $i \geq 0$. Finally let $D^* := \{fp(i) \mid i \geq 0\}$. *End definition of D^*.*

It remains to verify the assertions (1) – (4).

ad (1). D^* is the value set of a recursive function and thus r. e. *qed (1).*

ad (2). Let $d \in D^*$, i. e. $d = fp(i)$ for some $i \geq 0$ and $\varphi^d = \psi^{(i,d)}$. If in the construction of $\psi^{(i,d)}$ one of the cases A or B.1.1 occurs infinitely often, then the recursive core R_d of $\psi^{(i,d)}$ equals $\{\psi_0^{(i,d)}\}$. If case B.1.2 is fulfilled in some stage k, then $R_d = \{\psi_{2k}^{(i,d)}\}$. Otherwise, by remarks 1, 2, and 3, R_d equals either $\{\psi_{2k-1}^{(i,d)}\}$ or $\{\psi_{2k}^{(i,d)}\}$. Consequently, the recursive core R_d is a singleton. *qed (2).*

ad (3) and (4). Assertions (3) and (4) are verified via 3 claims:

(i) $D^* \notin \mathit{UniCons}$,
(ii) $D^* \in \mathit{resUniConf}$,
(iii) if $D \in \mathit{UniCons}$ (or $\mathit{resUniCons}$), then there is some computable function e satisfying Property α for $D_1 = D^*$ and $D_2 = D$.

ad (i). Assume $D^* \in \mathit{UniCons}$, i. e. there is some meta-IIM M^i such that M_d^i learns the recursive core R_d consistently in τ, whenever $d \in D^*$. Now let

$d^* = fp(i)$, in particular, $d^* \in D^*$ and $\varphi^{d^*} = \psi^{(i,d^*)}$.

Firstly, if in the construction of $\psi^{(i,d^*)}$ one of the cases A or B.1.1 occurs infinitely often, then the recursive core R_{d^*} equals $\{\psi_0^{(i,d^*)}\}$, but $M^i_{d^*}$ changes its mind on $\psi_0^{(i,d^*)}$ infinitely often.

Secondly, if case B.1.2 occurs in some stage k, then – by the note in case B.1.2 of stage k – $M^i_{d^*}$ does not *Cons*-learn the recursive core R_{d^*} in τ.

Otherwise, by the remarks 1, 2, and 3 in stage k, the learner $M^i_{d^*}$ does not *Cons*-learn the recursive core R_{d^*} in τ either.

Since in no case $M^i_{d^*}$ learns R_{d^*} consistently in τ, we obtain a contradiction. Hence $D^* \notin \mathit{Uni\,Cons}$. *qed (i).*

ad (ii). Define a meta-IIM M by

$$M_d(f[n]) := \begin{cases} 2k-1\,, & \text{if } n>1 \text{ and } f[n] = f[n']0(k+1)^m \text{ for some} \\ & k,m \geq 1,\ n' < n-1\,, \\ 2k\,, & \text{if } n>1 \text{ and } f[n] = f[n']1(k+1)^m \text{ for some} \\ & k,m \geq 1,\ n' < n-1\,, \\ 0\,, & \text{otherwise}\,, \end{cases}$$

for all recursive functions f and all $d, n \geq 0$. Now let $d \in D^*$ and $f \in R_d$.

Firstly, if $f = \varphi^d_0$, then $f(n) \in \{0,1\}$ for all $n \geq 1$. Thus $M_d(f[n]) = 0$ for all $n \geq 0$, in particular, M_d *Conf*-learns the function f with respect to φ^d.

Secondly, if there is some $k \geq 1$, such that $f = \varphi^d_{2k-1}$, then $f = f[n']0(k+1)^\infty$ for some $n' \geq 0$ such that $f(1), \ldots, f(n') \in \{0,1\}$, if $n' \geq 1$. By definition, M_d returns the hypothesis 0 for f in the first $n'+2$ steps, the hypothesis $2k-1$ afterwards. As φ^d_0 is a proper subfunction of φ^d_{2k-1}, M_d learns f conformly with respect to φ^d.

Thirdly, if $f = \varphi^d_{2k}$ for some $k \geq 1$, a similar argumentation shows that M_d learns f conformly in φ^d.

So M witnesses $D^* \in \mathit{res\,Uni\,Conf}$. *qed (ii).*

ad (iii). Let D be any description set; moreover define a computable function e for any $d, i \geq 0$ by

$$e(d,i) := \begin{cases} 0\,, & \text{if } fp(i) = d\,, \\ 1\,, & \text{if } fp(i) \neq d\,. \end{cases}$$

Now if $d \in D^*$ and $f \in R_d$, then $d = fp(i)$ for some $i \geq 0$ such that $f(0) = i$. This implies $e(d, f(0)) = e(d, i) = e(fp(i), i) = 0$. If $d \in D \setminus D^*$, then $d \neq fp(i)$ for all $i \geq 0$. So $e(d, f(0)) = 1$ for all $f \in R_d$. Consequently, e satisfies Property α for D^* and D. *qed (iii).*

Finally, to verify assertions (3) and (4), assume that D belongs to $\mathit{UniCons}$ (or $\mathit{resUniCons}$). As some computable function e satisfies Property α for D^*, D and $D^* \in \mathit{resUniConf}$, Theorem 23 implies $D \cup D^* \in \mathit{UniConf}$ (or $\mathit{resUniConf}$, respectively). Moreover, since $D^* \notin \mathit{UniCons}$, we obtain $D \cup D^* \notin \mathit{UniCons}$ and all in all $D \cup D^* \in \mathit{UniConf} \setminus \mathit{UniCons}$ (or $\mathit{resUniConf} \setminus \mathit{UniCons}$, respectively). *qed (3) and (4).*

This completes the proof of Theorem 24 for $I = \mathit{Conf}$ and $J = \mathit{Cons}$. □

The idea for the construction of the numberings η in the proof above is taken from a corresponding proof in [26]. There the existence of some set $D \in \mathit{resUniConf} \setminus \mathit{UniCons}$, describing only singletons, is verified. Here this proof is combined with a few new ideas. Similarly, all other statements of Theorem 24, Theorem 25, and Theorem 27 can be witnessed by such constructions using the corresponding ideas in [26]. Details are omitted.

6 Conclusions

This paper investigates inductive inference of recursive functions on the meta-level of three versions of uniform learning. Inference types resulting from different learning criteria have been analysed and the identification capacities of the corresponding meta-learners have been compared to each other. The desired strong separations have been successfully verified, thereby observing two additional properties:

- any strong separation verified above can even be witnessed by a fixed description set D^*;
- there are differences in the results concerning the three models of uniform learning (which stem from distinct requirements involving the choice of hypothesis spaces).

The strong separation results themselves show that it might in some cases be reasonable to give up certain constraints concerning the inference type J, because thus an increase of learning potential of the corresponding meta-learners can be achieved, even if it is required to learn at least some given description set $D \in \mathit{UniJ}$ (or $D \in \mathit{resUniJ}$, $D \in \mathit{extUniJ}$).

Given suitable inference types I and J, the proofs moreover indicate *how* to modify a uniform J-learner into a uniform I-learner of higher capacity. That means methods for designing more powerful learners are provided.

The existence of a fixed description set D^*, witnessing to the strong separations for any description set $D \in \mathit{Uni}J$ (analogously for $D \in \mathit{resUni}J$, $D \in \mathit{extUni}J$) suggests some structure for a somehow characteristic description set unsuitable for uniform J-learning. This structure is on the one hand complex enough to disallow for uniform learning according to J, but on the other hand simple enough to enable uniform I-learning, even in composition with any description set suitable for uniform I-learning and uniform J-learning.

The differences in the results concerning the three investigated models of uniform learning are evidence to the influence of the hypothesis spaces chosen for uniform learning.

Acknowledgements

The author thanks Dr. Steffen Lange, Prof. Dr. Rolf Wiehagen, Prof. Dr. Thomas Zeugmann, as well as the anonymous referees for many useful hints contributing to the quality of the presentation.

References

[1] Angluin, D.; Smith, C. (1983); Inductive Inference: Theory and Methods, *Computing Surveys* 15:237–269.

[2] Baliga, G.; Case, J.; Jain, S. (1999); The Synthesis of Language Learners, *Information and Computation* 152:16–43.

[3] Barzdin, J. (1974); Two Theorems on the Limiting Synthesis of Functions, *Theory of Algorithms and Programs*, Latvian State University, Riga 210:82–88 (in Russian).

[4] Barzdin, J. (1974); Inductive Inference of Automata, Functions and Programs, *Proceedings of the International Congress of Mathematicians*, Vancouver, pp. 455–460.

[5] Blum, M. (1967); A Machine-Independent Theory of the Complexity of Recursive Functions, *Journal of the ACM* 14:322–336.

[6] Blum, L.; Blum, M. (1975); Toward a Mathematical Theory of Inductive Inference, *Information and Control* 28:125-155.

[7] Case, J.; Chen, K.; Jain, S. (1992); Strong Separation of Learning Classes, *Journal of Experimental and Theoretical Artificial Intelligence* 4:281–293.

[8] Case, J.; Jain, S.; Ngo Manguelle, S. (1994); Refinements of Inductive Inference by Popperian and Reliable Machines, *Kybernetika* 30:23–52.

[9] Case, J.; Ngo Manguelle, S. (1979); Refinements of Inductive Inference by Popperian Machines, Technical Report 152, SUNY Buffalo.

[10] Case, J.; Smith, C. (1983); Comparison of Identification Criteria for Machine Inductive Inference, *Theoretical Computer Science* 25:193–220.

[11] Freivalds, R.; Kinber, E.; Wiehagen, R. (1995); How Inductive Inference Strategies Discover Their Errors, *Information and Computation* 118:208–226.

[12] Fulk, M. (1988); Saving the Phenomena: Requirements that Inductive Inference Machines Not Contradict Known Data, *Information and Computation* 79:193–209.

[13] Gold, E. M. (1967); Language Identification in the Limit, *Information and Control* 10:447–474.

[14] Jantke, K. P. (1979); Natural Properties of Strategies Identifying Recursive Functions, *Elektronische Informationsverarbeitung und Kybernetik* 15:487–496.

[15] Jantke, K. P.; Beick, H.-R. (1981); Combining Postulates of Naturalness in Inductive Inference, *Elektronische Informationsverarbeitung und Kybernetik* 17:465–484.

[16] Kapur, S.; Bilardi, G. (1992); On Uniform Learnability of Language Families, *Information Processing Letters* 44:35–38.

[17] Langley, P. (1994); *Elements of Machine Learning*, Morgan Kaufmann, San Francisco, California.

[18] Mitchell, T. (1997); *Machine Learning*, McGraw-Hill, Boston, Massachusetts.

[19] Osherson, D.; Stob, M.; Weinstein, S. (1988); Synthesizing Inductive Expertise, *Information and Computation* 77:138–161.

[20] Rogers, H. (1987); *Theory of Recursive Functions and Effective Computability*, MIT Press, Cambridge, Massachusetts.

[21] Wiehagen, R. (1978); *Zur Theorie der algorithmischen Erkennung*, Diss. B, Humboldt University, Berlin (in German).

[22] Wiehagen, R.; Zeugmann, T. (1995); Learning and Consistency, *Algorithmic Learning for Knowledge-Based Systems*, Lecture Notes in Artificial Intelligence 961, Springer-Verlag, Berlin, pp. 1–24.

[23] Zilles, S. (2001); On the Synthesis of Strategies Identifying Recursive Functions, *Proceedings of the 14th Annual Conference on Computational Learning Theory*, Lecture Notes in Artificial Intelligence 2111, Springer-Verlag, Berlin, pp. 160–176.

[24] Zilles, S. (2001); On the Comparison of Inductive Inference Criteria for Uniform Learning of Finite Classes, *Proceedings of the 12th International Conference on Algorithmic Learning Theory*, Lecture Notes in Artificial Intelligence 2225, Springer-Verlag, Berlin, pp. 251–266.

[25] Zilles, S. (2002); Merging Uniform Inductive Learners, *Proceedings of the 15th Annual Conference on Computational Learning Theory*, Lecture Notes in Artificial Intelligence 2375, Springer-Verlag, Berlin, pp. 201–215.

[26] Zilles, S.; Separation of Uniform Learning Classes, an extended and improved version of [24] submitted to *Theoretical Computer Science*.