On the Comparison of Inductive Inference Criteria for Uniform Learning of Finite Classes

Sandra Zilles

Fachbereich Informatik
Universität Kaiserslautern
Postfach 3049
D - 67653 Kaiserslautern
zilles@informatik.uni-kl.de

Abstract. We consider a learning model in which each element of a class of recursive functions is to be identified in the limit by a computable strategy. Given gradually growing initial segments of the graph of a function, the learner is supposed to generate a sequence of hypotheses converging to a correct hypothesis. The term correct means that the hypothesis is an index of the function to be learned in a given numbering. Restriction of the basic definition of learning in the limit yields several inference criteria, which have already been compared with respect to their learning power.
The scope of uniform learning is to synthesize appropriate identification strategies for infinitely many classes of recursive functions by a uniform method, i.e. a kind of meta-learning is considered. In this concept we can also compare the learning power of several inference criteria. If we fix a single numbering to be used as a hypothesis space for all classes of recursive functions, we obtain results similar to the non-uniform case. This hierarchy of inference criteria changes, if we admit different hypothesis spaces for different classes of functions. Interestingly, in uniform identification most of the inference criteria can be separated by collections of finite classes of recursive functions.

1 Introduction

Inductive Inference is concerned with theoretical models simulating learning processes. A model of quite simple mathematical description is for example identification of classes of recursive functions. This concept in general includes three main components:

- a partial-recursive function S – also called strategy – simulating the learner,
- a class U of total recursive functions which have to be identified by S,
- a partial-recursive numbering ψ – called hypothesis space – which enumerates at least all functions in U.

In each step of the identification process S is presented a finite subgraph of some unknown arbitrary function f contained in U; the strategy S then returns a hypothesis which is interpreted as an index of a function in the given numbering ψ. It is the learner's job to eventually return a single correct hypothesis, i.e. the sequence of outputs ought to converge to a ψ-number of f. This model – called identification in the limit – has first been analyzed by Gold in [5] and gave rise to the investigation and comparison of several new learning models ("inference criteria") basing on that principle. The common idea was to restrict the definition of identifiability by means of additional – and in some way natural – demands concerning the properties of the hypotheses. The corresponding models have been compared with respect to the resulting identification power; for some more background the reader is referred to [3], [4] and [7].

This paper studies Inductive Inference on a meta-level. Considering collections of infinitely many classes U of recursive functions we are looking for meta-learners synthesizing an appropriate strategy for each class U to be learned. For that purpose we agree on a method to describe a class U, because for the synthesis of a learner our meta-strategy should be given some description of U. That means we do not only try to solve a learning problem by an expert learner but to design a higher-level learner which constructs a method for solving a learning problem from a given description. Thus the meta-learner is able to simulate all the expert learners.

Uniform learning of classes of recursive functions has already been studied by Jantke in [6]. Unfortunately, his results are rather negative; he proves that there is no strategy which – given any description of an arbitrary class U consisting of just a single recursive function – synthesizes a learner which identifies U with respect to a fixed hypothesis space. Even if we allow different hypothesis spaces for the different classes of recursive functions, no meta-learner is successful for all descriptions of finite classes (cf. [12]). Since in the non-uniform case finite classes can be identified easily with respect to any common inference criterion, these results might suggest that the model of uniform learning yields a concept the investigation of which is not worthwile. As we will see, the results in this paper allow a more optimistic point of view. Of course it is quite natural to consider the same inference criteria known from the non-uniform model also in our meta-level. The aim of this paper is to investigate whether the comparison of these criteria concerning the resulting identification power yields hierarchies analogous to those approved in the classical context. In most cases we will see, that the classical separation results can be transferred to uniform learning. And we can prove even more. If we consider uniform learning with respect to fixed hypothesis spaces, all separations of inference criteria can be achieved by collections of *finite* classes of recursive functions. The resulting hierarchies correspond to the non-uniform case. If we drop the restrictions concerning the hypothesis spaces, we obtain slightly different results, although many of the criteria can still be separated by finite classes. So whereas finite classes are very simple regarding their identifiability in Gold's model, they are in most cases sufficient for the separation of inference criteria in uniform learning. Furthermore we conclude that the hierarchies obtained are very much influenced by the choice of the hypothesis spaces. Now, since the hierarchies of inference criteria do not collapse in our meta-level – even if we restrict ourselves to the choice of simple learning problems – we conclude that the concept of uniform learning is neither trivial nor fruitless. Furthermore this paper corroborates the interpretation that our different inference criteria possess some really substantial specific properties, which yield separations of such a strong nature that they still hold for uniform learning of finite classes.

In [12] the reader may also find positive results encouraging further research. It is shown that the choice of descriptions for the classes U has more influence on the uniform identifiability than the classes themselves, i.e. many meta-strategies fail rather because of a bad description of the learning problem than because of the complexity of the problem. So it might be interesting to find out what kinds of descriptions are suitable for uniform learnability and whether they can be characterized by any specific properties.

Further research on uniform identification has also been made in the context of language learning, see for example [8], [9] and [2]. Because of its numerous positive results, in particular the work of Baliga, Case and Jain [2] motivates the investigation of meta-strategies.

2 Preliminaries

Recursion theoretic terms used without explicit definition can be found in [10].

By $\mathbb{N}$ we denote the set of all nonnegative integers, $\mathbb{N}^*$ is the set of all finite tuples over $\mathbb{N}$; the variable n always ranges over $\mathbb{N}$. For fixed n, the notion $\mathbb{N}^n$ is used for the set of all n-tuples of integers. By implicit use of a bijective computable function $\mathrm{cod}: \mathbb{N}^* \mapsto \mathbb{N}$ we will identify any $\alpha \in \mathbb{N}^*$ with its coding $\mathrm{cod}(\alpha) \in \mathbb{N}$. A statement is quantified with $\forall^\infty n$ in order to indicate that the statement is fulfilled for all but finitely many n; quantifiers $\forall$ and $\exists$ are used in the common way.

For any set X the expression $\mathrm{card}\, X$ denotes the cardinality of X; $\wp X$ denotes the set of all subsets of X. As a symbol for set inclusion we use $\subseteq$, proper inclusion is indicated by $\subset$. Incomparability of sets is expressed by $\#$.

The set of all partial-recursive functions is denoted by $\mathcal{P}$, the set of total recursive functions by $\mathcal{R}$. In order to refer to functions of a fixed number n of input variables, we sometimes add the superscript n to these symbols. For any $f \in \mathcal{P}$, $x \in \mathbb{N}$ we write $f(x)\!\downarrow$, if f is defined on input x; $f(x)\!\uparrow$ otherwise. If $f \in \mathcal{P}$ and n fulfill $f(0)\!\downarrow, \ldots, f(n)\!\downarrow$ we set $f[n] := \mathrm{cod}(f(0), \ldots, f(n))$, i.e. $f[n]$ corresponds to the initial segment of length $n+1$ of f. Comparing $f, g \in \mathcal{P}$ we write $f =_n g$, if $\{(x, f(x)) \mid x \leq n,\ f(x)\!\downarrow\} = \{(x, g(x)) \mid x \leq n,\ g(x)\!\downarrow\}$; otherwise $f \neq_n g$. By the notion $f \subseteq g$ we indicate that $\{(x, f(x)) \mid x \in \mathbb{N},\ f(x)\!\downarrow\} \subseteq \{(x, g(x)) \mid x \in \mathbb{N},\ g(x)\!\downarrow\}$ and use proper inclusion by analogy. But $f \in \mathcal{P}$ may also be identified with the sequence $(f(n))_{n \in \mathbb{N}}$, so we sometimes write $f = 0^n 1 \uparrow^\infty$ and the like. We often identify a tuple $\alpha \in \mathbb{N}^*$ with the function $\alpha\uparrow^\infty$ implicitly. By $\mathrm{rng}(f)$ we refer to the range $\{f(x) \mid x \in \mathbb{N},\ f(x)\!\downarrow\}$ of a function $f \in \mathcal{P}$.

A function $\psi \in \mathcal{P}^{n+1}$ is used as a numbering for the set $\mathcal{P}_\psi := \{\psi_i \mid i \in \mathbb{N}\}$, where $\psi_i(x) := \psi(i, x)$ for all $i \in \mathbb{N}$, $x \in \mathbb{N}^n$ as usual. i is called ψ-number of the function ψ_i. In order to refer to the set of all total functions in $\mathcal{P}_\psi$, we use the notion $\mathcal{R}_\psi$, i.e. $\mathcal{R}_\psi := \mathcal{P}_\psi \cap \mathcal{R}$. $\mathcal{R}_\psi$ is called the recursive core or "$\mathcal{R}$-core" of $\mathcal{P}_\psi$. If $\psi \in \mathcal{P}^{n+2}$, every $b \in \mathbb{N}$ corresponds to a numbering $\psi^b \in \mathcal{P}^{n+1}$, if we define $\psi^b(i,x) := \psi(b,i,x)$ for all $i \in \mathbb{N},\ x \in \mathbb{N}^n$. Again i is a ψ^b-number for the function ψ_i^b defined in the common way.

Now we introduce our basic Inductive Inference criterion called identification in the limit, which was first defined in [5]. It may be regarded as a fundamental learning model from which we define further restrictive inference criteria (see Definitions 2 and 3). The notation EX in Definition 1 abbreviates the term "explanatory identification" which is also used to refer to learning in the limit.

Definition 1. *Let $U \subseteq \mathcal{R},\ \psi \in \mathcal{P}^2$. The class U belongs to EX_ψ and is called identifiable in the limit wrt the hypothesis space ψ iff there is a function $S \in \mathcal{P}$ (called strategy) such that for any $f \in U$:*

1. *$S(f[n])\downarrow$ for all $n \in \mathbb{N}$ ($S(f[n])$ is called hypothesis on $f[n]$),*
2. *there is some $j \in \mathbb{N}$ such that $\psi_j = f$ and $\forall^\infty n\ [S(f[n]) = j]$.*

If S is given, we also write $U \in EX_\psi(S)$. We set $EX := \bigcup_{\psi\in\mathcal{P}^2} EX_\psi$.

On any function $f \in U$ the strategy S must generate a sequence of hypotheses converging to a ψ-number of f. But a user reading the hypotheses generated by S up to a certain time will never know whether the actual hypothesis is correct or not, because he cannot decide whether the time of convergence is already reached. If there was a bound on the number of mind changes, he could at least rely on the actual hypothesis whenever the bound is reached. Learning with such bounds has first been studied in [3].

Definition 2. *Assume $U \subseteq \mathcal{R},\ \psi \in \mathcal{P}^2,\ m \in \mathbb{N}$. U belongs to $(EX_m)_\psi$ and is called identifiable (in the limit) with no more than m mind changes wrt ψ iff there exists an $S \in \mathcal{P}$ satisfying*

1. *$U \in EX_\psi(S)$ (where S is additionally permitted to return the sign "?"),*
2. *for all $f \in U$ there is an $n_f \in \mathbb{N}$ satisfying*
   - $\forall x < n_f\ [S(f[x]) = ?]$,
   - $\forall x \geq n_f\ [S(f[x]) \in \mathbb{N}]$,
3. *$card\{n \in \mathbb{N} \mid ? \neq S(f[n]) \neq S(f[n+1])\} \leq m$ for all $f \in U$.*

We use the notations $(EX_m)_\psi(S)$ and EX_m by analogy with Definition 1. A class $U \subseteq \mathcal{R}$ is identifiable with a bounded number of mind changes iff there is an $m \in \mathbb{N}$ such that $U \in EX_m$.

The output "?" allows our strategy to indicate that its hypothesis is left open for the actual time being, in order not to waste a mind change in the beginning of the learning process.

It is also a natural thought to strengthen the demands concerning the intermediate hypotheses themselves. A successful learning behaviour might be to generate intermediate hypotheses agreeing with the information received up to the actual time of the learning process ("consistent" hypotheses, cf. [1]). In order to be less demanding, one could also ask for hypotheses which do not disagree convergently (i.e. in their *defined* values) with the actual information ("conform" hypotheses, see [11]). Since any hypothesis representing a function not contained in $\mathcal{R}$ must be wrong, another natural demand would be to allow only ψ-numbers of total recursive functions ("total" hypotheses, cf. [7]) as outputs of S. Since in general the halting problem in ψ is not decidable, it might be hard for our strategy to detect the incorrectness of a hypothesis, if the corresponding function differs from the function to be learned only by being undefined for some arguments. For learning with "convergently incorrect" hypotheses (cf. [4]) such outputs are forbidden.

Definition 3. *Choose a pair $(I, \mathcal{C}_I)$ from those listed below. Let $U \subseteq \mathcal{R}$, $\psi \in \mathcal{P}^2$. U is called identifiable under the criterion I wrt ψ iff there exists a strategy $S \in \mathcal{P}$ such that $U \in EX_\psi(S)$ and for all $f \in U$, $n \in \mathbb{N}$ condition $\mathcal{C}_I$ is satisfied.*

I	$\mathcal{C}_I$
CONS	$\psi_{S(f[n])} =_n f$
CONF	$\forall x \leq n\ [\psi_{S(f[n])}(x)\downarrow\ \Rightarrow\ \psi_{S(f[n])}(x) = f(x)]$
TOTAL	$\psi_{S(f[n])} \in \mathcal{R}$
CEX	$\psi_{S(f[n])} \not\subset f$

We use the phrases "identification with consistent (conform, total, convergently incorrect) intermediate hypotheses" respectively. The notations I, I_ψ, $I_\psi(S)$ are used in the common way.

For the inference criteria introduced here the following comparison results have been proved:

Theorem 1. *1. $\forall m \in \mathbb{N}\ [EX_m \subset EX_{m+1} \subset EX]$ (see [3]).*
2. $TOTAL \subset CONS \subset CONF \subset EX \subset \wp\mathcal{R}$ (see [7], [11] and [5]).
3. $TOTAL \subset CEX \subset EX$ (see definitions and [4]).
4. $CEX \# CONS$ (see [4]).

For convenience $\mathcal{I} := \{\mathrm{EX}, \mathrm{CONS}, \mathrm{CONF}, \mathrm{TOTAL}, \mathrm{CEX}\} \cup \{\mathrm{EX}_m \mid m \in \mathbb{N}\}$ denotes the set of all inference criteria declared in this section. Furthermore let

- $J^* := \{U \subseteq \mathcal{R} \mid U \text{ is finite}\}$,
- $J^1 := \{U \subseteq \mathcal{R} \mid \text{card } U = 1\}\ (= \{\{f\} \mid f \in \mathcal{R}\})$.

3 Uniform Learning – Definition and Basic Results

From now on let $\varphi \in \mathcal{P}^3$ be a fixed acceptable numbering of $\mathcal{P}^2$ and $\tau \in \mathcal{P}^2$ an acceptable numbering of $\mathcal{P}^1$. As φ is acceptable, it might be regarded as a numbering of all numberings $\psi \in \mathcal{P}^2$: every $b \in \mathbb{N}$ corresponds to the function φ^b which is defined by $\varphi^b(i, x) := \varphi(b, i, x)$ for any $i, x \in \mathbb{N}$. Thus b also describes a class $\mathcal{R}_b$ of recursive functions, where $\mathcal{R}_b := \mathcal{R}_{\varphi^b} = \mathcal{P}_{\varphi^b} \cap \mathcal{R}$; i.e. $\mathcal{R}_b$ is the recursive core of $\mathcal{P}_{\varphi^b}$. Therefore any set $B \subseteq \mathbb{N}$ will be called *description set* for the collection $\{\mathcal{R}_b \mid b \in B\}$ of recursive cores corresponding to the indices in B. Considering each recursive core as a set of functions to be identified, any description set $B \subseteq \mathbb{N}$ may be associated to a collection of learning problems. Now we are looking for a meta-learner which – given any description $b \in B$ – develops a special learner coping with the learning problem described by b, i.e. the special learner must identify each function in $\mathcal{R}_b$.

Definition 4. *Let $J \subseteq \wp\mathcal{R}$, $I \in \mathcal{I}$, $J \subseteq I$, $B \subseteq \mathbb{N}$. The set B is called suitable for uniform learning wrt J and I iff the following conditions are fulfilled:*

1. $\forall b \in B\ [\mathcal{R}_b \in J]$,
2. $\exists S \in \mathcal{P}^2\ \forall b \in B\ \exists \psi \in \mathcal{P}^2\ [\mathcal{R}_b \in I_\psi(\lambda x.S(b,x))]$.

We abbreviate this by $B \in suit(J,I)$ and write $B \in suit(J,I)(S)$, if S is given.

So $B \in \text{suit}(J,I)$ iff every recursive core described by some index $b \in B$ belongs to the class J and additionally there is a strategy $S \in \mathcal{P}^2$ which, given $b \in B$, synthesizes an I-learner successful for $\mathcal{R}_b$ with respect to some appropriate hypothesis space ψ. Note that the synthesis of these appropriate hypothesis spaces is *not* required. This means in particular, that in general the output of a meta-learner cannot be interpreted practically, because we might not know which numbering is actually used as a hypothesis space. Of course we might restrict our definition of suitable description sets by demanding uniform learnability with respect to the acceptable numbering τ for all classes $\mathcal{R}_b$. Another possibility is to use the numberings φ^b, $b \in B$, already given by the description set B as hypothesis spaces for I-identification of the classes $\mathcal{R}_b$.

Definition 5. *Let $I \in \mathcal{I}$, $J \subseteq I$, $B \subseteq \mathbb{N}$, $S \in \mathcal{P}^2$. Assume $B \in suit(J,I)(S)$. We write $B \in suit_\tau(J,I)(S)$ if $\mathcal{R}_b \in I_\tau(\lambda x.S(b,x))$ for all $b \in B$. The notation $B \in suit_\varphi(J,I)(S)$ shall indicate that $\mathcal{R}_b \in I_{\varphi^b}(\lambda x.S(b,x))$ for all $b \in B$. We also use the notations $suit_\tau(J,I)$ and $suit_\varphi(J,I)$ in the usual way.*

Of course it would be nice to find characterizations of the sets suitable for uniform learning with respect to J, I, where $I \in \mathcal{I}$ and $J \subseteq I$ are given. This paper compares the uniform identification power of several criteria $I \in \mathcal{I}$ and concentrates on the case $J = J^*$, i.e. all recursive cores to be identified are finite. Our first result follows obviously from our definitions.

Proposition 1. *Let $I \in \mathcal{I}$, $J \subseteq I$. Then $suit_\varphi(J,I) \subseteq suit_\tau(J,I) \subseteq suit(J,I)$.*

Whether these inclusions are proper inclusions or not depends on the choice of J and I. If they turned out to be equalities for all J and I, then Definition 5 would be superfluous. But in fact, as Theorem 5 will show, we have proper inclusions in the general case. That means that a restriction in the choice of the hypothesis spaces results in a restriction of the learning power of meta-strategies.

Any strategy identifying a class $U \subseteq \mathcal{R}$ with respect to some criterion $I \in \mathcal{I} \setminus \{\text{CONS}, \text{CONF}\}$ can be replaced by a *total* recursive strategy without loss of learning power. This new strategy is defined by computing the values of the old strategy for a bounded number of steps and a bounded number of input examples with increasing bounds. As long as no hypothesis is found, some temporary hypothesis agreeing with the restrictions in the definition of I is produced. Afterwards the hypotheses of the former strategy are put out "with delay".[1] Now we transfer these observations to the level of uniform learning and get the following result, which we will use in several proofs:

Proposition 2. *Let $I \in \mathcal{I} \setminus \{CONS, CONF\}$, $J \subseteq I$, $B \subseteq \mathbb{N}$. Assume $B \in suit(J, I)$ ($suit_\tau(J, I)$). Then there is a total recursive function S such that $B \in suit(J, I)(S)$ ($suit_\tau(J, I)(S)$ respectively).*

Let us now collect some simple examples of description sets suitable or not suitable for uniform learning. First we consider the identification of classes consisting of just one recursive function. Any set describing such classes turns out to be suitable for identification under any of our criteria:

Theorem 2. *Let $I \in \mathcal{I}$. Then $suit(J^1, I) = \{B \subseteq \mathbb{N} \mid \mathcal{R}_b \in J^1$ for all $b \in B\}$.*

Proof. Let $B \subseteq \mathbb{N}$ fulfill $\mathcal{R}_b \in J^1$ for all $b \in B$. Since for all $f \in \mathcal{R}$ there exists a numbering $\psi \in \mathcal{P}^2$ with $\psi_0 = f$, the strategy constantly zero yields $B \in \text{suit}(J^1, I)$. Thus $\{B \subseteq \mathbb{N} \mid \mathcal{R}_b \in J^1 \text{ for all } b \in B\} \subseteq \text{suit}(J^1, I)$. The other inclusion is obvious. □

Unfortunately, we would rather not regard the strategy defined in this proof as an "intelligent" learner, because its output does not depend on the input at all. Its success lies just in the choice of appropriate hypothesis spaces. If such a choice of hypothesis spaces is forbidden, we obtain an absolutely negative result:

Theorem 3. $\{b \in \mathbb{N} \mid \mathcal{R}_b \in J^1\} \notin suit_\tau(J^1, EX)$.
In particular even $\{b \in \mathbb{N} \mid card\ \{i \in \mathbb{N} \mid \varphi_i^b \in \mathcal{R}\} = 1\} \notin suit_\tau(J^1, EX)$.

For a proof see [6] or [12]. So, if we fix our hypothesis spaces in advance, not even the classes consisting of just one element can be identified in the limit uniformly. Regarding the identification of arbitrary finite classes (the learnability of which is trivial in the non-uniform case), the situation gets worse still. Even by free choice of the hypothesis spaces we cannot achieve uniform EX-identifiability.

Theorem 4. $\{b \in \mathbb{N} \mid \mathcal{R}_b \in J^*\} \notin suit(J^*, EX)$.

A proof can be found in [12]. How can we interpret these results? Is the concept of uniform learning fruitless and further research on this area not worthwile? Fortunately, many results in [2] and [12] allow a more optimistic point of view.

[1] This does not work for CONS and CONF, since in general after the delay the hypotheses are no longer consistent or conform with the information in the actual time of the learning process.

For example, [12] shows that some constraints on the descriptions $b \in B$ – especially concerning the topological structure of the numberings φ^b – yield uniform learnability of huge classes of functions, even with consistent *and* total intermediate hypotheses and also with respect to our acceptable numbering τ. The sticking point seems to be that uniform identifiability is not so much influenced by the classes to be learned, but by the numberings φ^b chosen as representations for these classes. So the numerous negative results should be interpreted carefully. For example the reason that there is no uniform EX-learner for $\{b \in \mathbb{N} \mid \mathcal{R}_b \in J^*\}$ is not so much the complexity of finite classes but rather the need to cope with *any* numbering possessing a finite $\mathcal{R}$-core. Based on these aspects we should not tend to a pessimistic view concerning the fruitfulness of the concept of uniform learning. Our results in the following sections will substantiate this opinion.

Theorems 2 and 3 now enable the proof of the following example of a strict version of Proposition 1.

Theorem 5. *$suit_\varphi(J^1, I) \subset suit_\tau(J^1, I) \subset suit(J^1, I)$ for all $I \in \mathcal{I}$.*

Proof. $\mathrm{suit}_\tau(J^1, I) \subset \mathrm{suit}(J^1, I)$ is obtained as follows: by Theorem 3 we know that $B_1 := \{b \in \mathbb{N} \mid \mathcal{R}_b \in J^1\} \notin \mathrm{suit}_\tau(J^1, I)$ (otherwise B_1 was also an element of $\mathrm{suit}_\tau(J^1, \mathrm{EX})$). Thus by Theorem 2 we obtain $B_1 \in \mathrm{suit}(J^1, I) \backslash \mathrm{suit}_\tau(J^1, I)$.

It remains to prove $\mathrm{suit}_\varphi(J^1, I) \subset \mathrm{suit}_\tau(J^1, I)$. Again by Theorem 3 we know that there exists a set $B \subseteq \mathbb{N}$ such that card $\{i \in \mathbb{N} \mid \varphi_i^b \in \mathcal{R}\} = 1$ for all $b \in B$ and $B \notin \mathrm{suit}_\varphi(J^1, \mathrm{EX})$. Now let $g \in \mathcal{R}$ be a computable function satisfying

$$\varphi_i^{g(b)}(x) = \begin{cases} 0 & \text{if } \varphi_i^b(y)\!\downarrow \text{ for all } y \leq x \\ \uparrow & \text{otherwise} \end{cases} \quad \text{for any } b, i, x \in \mathbb{N} \,.$$

Let $B' := \{g(b) \mid b \in B\}$. Since $\mathcal{R}_{g(b)} = \{0^\infty\}$ for $b \in B$, we get $B' \in \mathrm{suit}_\tau(J^1, I)$ (via a strategy which constantly returns a τ-index of the function 0^∞).

Obviously $\{i \in \mathbb{N} \mid \varphi_i^{g(b)} \in \mathcal{R}\} = \{i \in \mathbb{N} \mid \varphi_i^b \in \mathcal{R}\}$ for all $b \in \mathbb{N}$. If there was a strategy $S \in \mathcal{P}^2$ satisfying $B' \in \mathrm{suit}_\varphi(J^1, I)(S)$, we would achieve $B \in \mathrm{suit}_\varphi(J^1, \mathrm{EX})(T)$ by defining $T(b, f[n]) := S(g(b), 0^n)$ for $f \in \mathcal{R},\ b, n \in \mathbb{N}$. This contradicts the choice of B, so $B' \in \mathrm{suit}_\tau(J^1, I) \backslash \mathrm{suit}_\varphi(J^1, I)(S)$. Hence $\mathrm{suit}_\varphi(J^1, I) \subset \mathrm{suit}_\tau(J^1, I)$. □

4 Separation of Inference Criteria – Special Hypothesis Spaces

From now on we will compare the learning power of our inference criteria for uniform learning of finite classes of recursive functions, i.e. we try to find results in the style of Theorem 1, where the criteria $I \in \mathcal{I}$ are replaced by the sets $\mathrm{suit}(J^*, I)$, $\mathrm{suit}_\tau(J^*, I)$ or $\mathrm{suit}_\varphi(J^*, I)$. Please note that a separation like for example $\mathrm{suit}_\tau(\mathrm{CONS}, \mathrm{CONS}) \subset \mathrm{suit}_\tau(\mathrm{CONS}, \mathrm{EX})$ is not a very astonishing result. The remarkable point is that even collections of *finite* classes of recursive functions suffice for a separation (note that in the non-uniform case finite classes can be identified under *any* criterion $I \in \mathcal{I}$ easily).

Since all proofs for the theorems stated in Section 4 proceed in a similar manner and include rather long constructions, we will omit most of them and just give sketches of the proofs for Theorem 7 and Theorem 9.

In this section we concentrate on uniform learning with respect to fixed hypothesis spaces, i.e. according to Definition 5. Our aim is to show that all the comparison results in Theorem 1 hold analogously for these concepts, even if all classes to be learned are finite. Lemma 1 summarizes some simple observations.

Lemma 1. *1. $suit_\varphi(J^*, EX_m) \subseteq suit_\varphi(J^*, EX_{m+1}) \subseteq suit_\varphi(J^*, EX)$ for arbitrary $m \in \mathbb{N}$,*
2. $suit_\varphi(J^*, TOTAL) \subseteq suit_\varphi(J^*, CONS) \subseteq suit_\varphi(J^*, CONF) \subseteq suit_\varphi(J^*, EX)$,
3. $suit_\varphi(J^*, TOTAL) \subseteq suit_\varphi(J^*, CEX) \subseteq suit_\varphi(J^*, EX)$.
These results hold analogously if we substitute $suit_\varphi$ by $suit_\tau$.

Proof. All these inclusions except for $\mathrm{suit}_\varphi(J^*, \mathrm{TOTAL}) \subseteq \mathrm{suit}_\varphi(J^*, \mathrm{CONS})$ (or analogously with τ instead of φ) follow immediately from the definitions. If a set $B \subseteq \mathbb{N}$ fulfills $B \in \mathrm{suit}_\varphi(J^*, \mathrm{TOTAL})(S)$ for some strategy $S \in \mathcal{P}^2$, we can easily define $T \in \mathcal{P}^2$ such that $B \in \mathrm{suit}_\varphi(J^*, \mathrm{CONS})(T)$. On input $(b, f[n])$ the strategy T just has to check the hypothesis $S(b, f[n])$ for consistency wrt φ^b. For $b \in B$, $f \in \mathcal{R}_b$ this is possible, because the function $\varphi^b_{S(b,f[n])}$ is total. If consistency is verified, T returns the same index as S, otherwise it returns some consistent hypothesis (which can be found, if $f \in \mathcal{R}_b$). Convergence to a correct hypothesis follows from the choice of S. The τ-case is proved by analogy. □

Now we want to prove that all these inclusions are in fact proper inclusions. For that purpose consider Theorem 6 first.

Theorem 6. *$suit_\varphi(J^*, EX_{m+1}) \setminus suit(J^*, EX_m) \neq \emptyset$ for any $m \in \mathbb{N}$.*[2]

Note that this result is even stronger than required. We just needed to prove $\mathrm{suit}_\varphi(J^*, \mathrm{EX}_{m+1}) \setminus \mathrm{suit}_\varphi(J^*, \mathrm{EX}_m) \neq \emptyset$ and the corresponding statement for the τ-case. Besides we have not only verified $\mathrm{suit}(J^*, \mathrm{EX}_{m+1}) \setminus \mathrm{suit}(J^*, \mathrm{EX}_m) \neq \emptyset$, but we observe a further fact: though we know uniform learning with respect to the hypothesis spaces given by φ to be much more restrictive than uniform learning without special demands concerning the hypothesis spaces, we still can find collections of class-descriptions which are

- restrictive enough to describe finite classes of recursive functions only,
- suitable for uniform EX_{m+1}-identification with respect to the hypothesis spaces corresponding to their descriptions,
- but *not* suitable for uniform EX_m-identification even if the hypothesis spaces can be chosen without restrictions.

Similar strict separations are obtained by the following theorems.

Theorem 7. *$suit_\varphi(J^*, EX) \setminus suit(J^*, CONF) \neq \emptyset$.*

[2] The proof is omitted but proceeds similar to the proof of Theorem 7.

Proof. We will just give a sketch of the relevant parts of the proof; details and formal constructions are not needed to explain the general idea common to most of the proofs of our results. We use a strategy $T \in \mathcal{R}$ to define a description set $B \subseteq \mathbb{N}$ suitable for uniform identification in the limit by T. The set B shall describe only finite recursive cores and will not be suitable for uniform conform identification. The choice of the strategy T may seem rather arbitrary, but it will enable an indirect proof.

Define $T \in \mathcal{R}$ by

$$T(f[n]) := \begin{cases} 0 & \text{if } f[n] \in \{0,1\}^* \\ \max\{f(0), \ldots, f(n)\} - 1 & \text{otherwise} \end{cases}$$

for arbitrary $f \in \mathcal{R}$ and $n \in \mathbb{N}$.
Then set $B := \{b \in \mathbb{N} \mid \mathcal{R}_b \text{ is finite and } \mathcal{R}_b \in \mathrm{EX}_{\varphi^b}(T)\}$.

We will prove $B \in \mathrm{suit}_\varphi(J^*, \mathrm{EX}) \setminus \mathrm{suit}(J^*, \mathrm{CONF})$. By definition of B we obviously have $B \in \mathrm{suit}_\varphi(J^*, \mathrm{EX})$. Now $B \notin \mathrm{suit}(J^*, \mathrm{CONF})$ is verified by way of contradiction.

Assumption. $B \in \mathrm{suit}(J^*, \mathrm{CONF})$.
Then there is some $S \in \mathcal{P}^2$ such that $\mathcal{R}_b \in \mathrm{CONF}(\lambda x.S(b,x))$ for all $b \in B$.

Aim. Construct an integer b_0, such that $b_0 \in B$, but $\mathcal{R}_{b_0} \notin \mathrm{CONF}(\lambda x.S(b_0,x))$, in contradiction to our assumption. The strategy $\lambda x.S(b_0,x)$ will fail for at least one function $f \in \mathcal{R}_{b_0}$ by either

- changing its hypothesis for f infinitely often or
- not terminating its computation on input of some initial segment of f or
- violating the conformity demand on input of some initial segment of f.

Construction of b_0. We define $\eta^b \in \mathcal{P}^2$ uniformly in $b \in \mathbb{N}$. First we define $\eta_0^b(0) := 0$. If we set $y_0 := 0$, the segment $\eta_0^b[y_0]$ is already defined. We start in stage 0.

In general, in stage k we proceed as follows:
For the definition of further values of η_0^b one computes $S(b, \eta_0^b[y_k])$, $S(b, \eta_0^b[y_k]0)$ and $S(b, \eta_0^b[y_k]1)$. If one of these values is undefined, then $\eta_0^b = 0{\uparrow}^\infty$. Else, if these values are all equal, we append zeros until we observe that the strategy $\lambda x.S(b,x)$ changes its mind on the initial segment constructed so far. Otherwise we just append one value $t \in \{0,1\}$, such that $S(b, \eta_0^b[y_k]) \neq S(b, \eta_0^b[y_k]t)$.

The functions η_{2k+1}^b and η_{2k+2}^b are defined as follows:
$\eta_{2k+1}^b[y_k + 2] := \eta_0^b[y_k]0(2k+2)$, $\eta_{2k+2}^b[y_k + 2] := \eta_0^b[y_k]1(2k+3)$. Both functions will be extended by zeros until the values $S(b, \eta_0^b[y_k]0)$ and $S(b, \eta_0^b[y_k]1)$ are computed and the definition of η_0^b is stopped temporarily because of a mind change of $\lambda x.S(b,x)$ on the initial segment of η_0^b constructed so far (if

these conditions are never satisfied, we obtain $\eta^b_{2k+1} = \eta^b_0[y_k]0(2k+2)0^\infty$ and $\eta^b_{2k+2} = \eta^b_0[y_k]1(2k+3)0^\infty$). If the definition of η^b_0 is stopped temporarily, let y_{k+1} be the maximal argument for which η^b_0 is defined. If y_{k+1} exists, go to stage $k+1$.

The Recursion Theorem then yields an integer $b_0 \in \mathbb{N}$ satisfying $\varphi^{b_0} = \eta^{b_0}$.

Claim. 1. We have $\mathrm{rng}(\varphi^{b_0}_0) \subseteq \{0,1\}$; if $x \in \mathbb{N}$, then $\max(\mathrm{rng}(\varphi^{b_0}_{x+1})) = x+2$ or $\mathrm{rng}(\varphi^{b_0}_{x+1}) = \emptyset$.

2. If in the construction of φ^{b_0} all stages are reached, then $\mathcal{R}_{b_0} = \{\varphi^{b_0}_0\}$. If stage k ($k \in \mathbb{N}$) is the last stage to be reached, then $\mathcal{R}_{b_0} = \{\varphi^{b_0}_0, \varphi^{b_0}_{2k+1}, \varphi^{b_0}_{2k+2}\}$ or $\mathcal{R}_{b_0} = \{\varphi^{b_0}_{2k+1}, \varphi^{b_0}_{2k+2}\}$.

This claim implies $b_0 \in B$. For the proof of $\mathcal{R}_{b_0} \notin \mathrm{CONF}(\lambda x.S(b_0,x))$ we assume by way of contradiction that $\mathcal{R}_{b_0} \in \mathrm{CONF}_\psi(\lambda x.S(b_0,x))$ for some numbering $\psi \in \mathcal{P}^2$. By Claim 2 it suffices to consider the following three cases:

Case 1. $\mathcal{R}_{b_0} = \{\varphi^{b_0}_0\}$.
Then all stages are reached in our construction. We observe that in the identification process for $\varphi^{b_0}_0$ the strategy $\lambda x.S(b_0,x)$ changes its hypothesis infinitely often.

Case 2. $\mathcal{R}_{b_0} = \{\varphi^{b_0}_{2k+1}, \varphi^{b_0}_{2k+2}\}$ for some $k \in \mathbb{N}$.
In this case we have $S(b_0, \varphi^{b_0}_{2k+1}[y_k+1])\uparrow$ or $S(b_0, \varphi^{b_0}_{2k+2}[y_k+1])\uparrow$ (with y_k as in our construction), so $\lambda x.S(b_0,x)$ cannot be successful for both $\varphi^{b_0}_{2k+1}$ and $\varphi^{b_0}_{2k+2}$.

Case 3. $\mathcal{R}_{b_0} = \{\varphi^{b_0}_0, \varphi^{b_0}_{2k+1}, \varphi^{b_0}_{2k+2}\}$ for $k \in \mathbb{N}$.
Then stage k is reached; stage $k+1$ is not reached. Furthermore

$$S(b_0, \varphi^{b_0}_{2k+1}[y_k+1]) = S(b_0, \eta^{b_0}_0[y_k]0) = S(b_0, \eta^{b_0}_0[y_k]1) = S(b_0, \varphi^{b_0}_{2k+2}[y_k+1]) ,$$

although $\varphi^{b_0}_{2k+1}[y_k+1] \neq \varphi^{b_0}_{2k+2}[y_k+1]$. Thus $i := S(b_0, \varphi^{b_0}_{2k+1}[y_k+1])$ cannot be a ψ-number for both $\varphi^{b_0}_{2k+1}$ and $\varphi^{b_0}_{2k+2}$. There are two possibilities:

Case 3.1. $\psi_i(y_k+1)\uparrow$.
Then the sequence of hypotheses produced by $\lambda x.S(b_0,x)$ on the function $\varphi^{b_0}_0$ converges to an index incorrect for $\varphi^{b_0}_0$ with respect to ψ.

Case 3.2. $\psi_i(y_k+1)\downarrow$.
Then i is not conform for both $\varphi^{b_0}_{2k+1}[y_k+1]$ and $\varphi^{b_0}_{2k+2}[y_k+1]$ wrt ψ.

We conclude $\mathcal{R}_{b_0} \notin \mathrm{CONF}_\psi(\lambda x.S(b_0,x))$; thus $\mathcal{R}_{b_0} \notin \mathrm{CONF}(\lambda x.S(b_0,x))$. The properties of b_0 now contradict our assumption, so $B \notin \mathrm{suit}(J^*, \mathrm{CONF})$. This completes the proof. □

A separation of the criteria CONF and CONS in the uniform learning model can be verified with similar methods; the proof is omitted.

Theorem 8. $suit_\varphi(J^*, CONF) \setminus suit(J^*, CONS) \neq \emptyset$.

So, the results CONS $\subset$ CONF $\subset$ EX can also be transferred to uniform learning with respect to τ and the numberings given a priori by φ. Again, finite classes are sufficient for the separations.

In order to prove $\text{suit}_\varphi(J^*, \text{TOTAL}) \subset \text{suit}_\varphi(J^*, \text{CONS})$ (and the same result for suit_τ) we use Theorem 9. Since $\text{suit}_\tau(J^*, \text{TOTAL}) \subseteq \text{suit}_\tau(J^*, \text{CEX})$, we even obtain $\text{suit}_\varphi(J^*, \text{CONS}) \setminus \text{suit}_\tau(J^*, \text{TOTAL}) \neq \emptyset$.

Theorem 9. $suit_\varphi(J^*, CONS) \setminus suit_\tau(J^*, CEX) \neq \emptyset$.

Proof. We will omit some formal details and concentrate on the main ideas. Again we use a strategy $T \in \mathcal{P}^2$ to define a description set $B \subseteq \mathbb{N}$ suitable for uniform consistent identification by T. Though B describes only finite recursive cores, it will not be suitable for uniform CEX-identification with respect to τ.

Define $T \in \mathcal{P}^2$ by

$$T(b, f[n]) := \begin{cases} 0 & \text{if } 0 \notin \{f(0), \dots, f(n)\} \\ \min\{i \geq 1 \mid \exists \alpha \in (\mathbb{N} \setminus \{0\})^* & \text{if } 0 \in \{f(0), \dots, f(n)\} \text{ and} \\ \qquad [\alpha 0 \subseteq \varphi_i^b \text{ and } \alpha 0 \subseteq f]\} & \text{such a minimum is found} \\ \uparrow & \text{otherwise} \end{cases}$$

for $f \in \mathcal{R}$ and $b, n \in \mathbb{N}$.
Then set $B := \{b \in \mathbb{N} \mid \mathcal{R}_b \text{ is finite and } \mathcal{R}_b \in \text{CONS}_{\varphi^b}(\lambda x.T(b, x))\}$.

We will prove $B \in \text{suit}_\varphi(J^*, \text{CONS}) \setminus \text{suit}_\tau(J^*, \text{CEX})$. The definitions imply $B \in \text{suit}_\varphi(J^*, \text{CONS})$. The claim $B \notin \text{suit}_\tau(J^*, \text{CEX})$ is verified by way of contradiction.

Assumption. $B \in \text{suit}_\tau(J^*, \text{CEX})$,
i.e. there is some $S \in \mathcal{R}^2$ such that $\mathcal{R}_b \in \text{CEX}_\tau(\lambda x.S(b, x))$ for any $b \in B$.

Aim. Construction of an integer $b_0 \in B$ with $\mathcal{R}_{b_0} \notin \text{CEX}_\tau(\lambda x.S(b_0, x))$, in contradiction to our assumption. The strategy $\lambda x.S(b_0, x)$ will fail for at least one $f \in \mathcal{R}_{b_0}$ by either

- changing its hypothesis for f infinitely often or
- generating a hypothesis incorrect for f with respect to τ for infinitely many initial segments of f or
- guessing a τ-number of a proper subfunction of f on input of some initial segment of f.

Construction of b_0.
Define a function $\psi \in \mathcal{P}^3$ with the help of initial segments α_k^b ($b, k \in \mathbb{N}$) as follows: for arbitrary $b \in \mathbb{N}$ set $\alpha_0^b := 1$ and begin in stage 0.

In general, in stage k we proceed as follows:
$e := S(b, \alpha_k^b)$. Start a parallel check until *(i)* or *(ii)* turns out to be true.

(i). There is some $y < |\alpha_k^b|$ such that $\tau_e(y)$ is defined and $\tau_e(y) \neq \alpha_k^b(y)$.
(ii). There is some $y \geq |\alpha_k^b|$ such that $\tau_e(y)$ is defined.

The function ψ_{k+1}^b shall have the initial segment $\alpha_k^b 0$ which will be extended by a sequence of 0's, until *(i)* or *(ii)* turns out to be true. If condition *(i)* turns out to be true first, then ψ_0^b shall have the initial segment α_k^b which will be extended by a sequence of 1's, until $\lambda x.S(b, x)$ is forced to change its mind on ψ_0^b; then α_{k+1}^b shall be the initial segment of ψ_0^b constructed so far. If condition *(ii)* turns out to be true first – with $\tau_e(y_k)\downarrow$, $y_k > n$ – then $\alpha_{k+1}^b := \alpha_k^b 1 \ldots 1(\tau_e(y_k) + 1)$, where the last argument in the domain of α_{k+1}^b is y_k. In case α_{k+1}^b is defined go to stage $k + 1$. If neither *(i)* nor *(ii)* is fulfilled, ψ_0^b remains initial.

The Recursion Theorem then yields an integer $b_0 \in \mathbb{N}$ satisfying $\varphi^{b_0} = \psi^{b_0}$.

Claim. The construction in stage k implies

1. $\varphi_{k+1}^{b_0} \in \mathcal{R}$ iff $[\alpha_k^{b_0}$ is defined and $\tau_{S(b_0, \alpha_k^{b_0})} \subseteq \alpha_k^{b_0}\uparrow^\infty \ (\subset \varphi_{k+1}^{b_0})]$,
2. if $\varphi_{k+1}^{b_0} \notin \mathcal{R}$ and $\alpha_{k+1}^{b_0}\uparrow$, then $\varphi_0^{b_0} = \alpha_k^{b_0} 1^\infty \in \mathcal{R}$ and the sequence of hypotheses produced by $\lambda x.S(b_0, x)$ on $\varphi_0^{b_0}$ converges to an index incorrect for $\varphi_0^{b_0}$ with respect to τ,
3. if $\varphi_{k+1}^{b_0} \notin \mathcal{R}$ and $\alpha_{k+1}^{b_0}\downarrow$, then $\alpha_k^{b_0} \subseteq \alpha_{k+1}^{b_0} \subseteq \varphi_0^{b_0}$; furthermore
   (a) $S(b_0, \alpha_{k+1}^{b_0}) \neq S(b_0, \alpha_k^{b_0})$ or
   (b) $S(b_0, f[|\alpha_k^{b_0}| - 1])$ is incorrect wrt τ for any $f \in \mathcal{R}$ satisfying $\alpha_{k+1}^{b_0} \subset f$,
4. if $\varphi_0^{b_0} \in \mathcal{R}$, then $\varphi_{k+1}^{b_0} \notin \mathcal{R}$ for all $k \in \mathbb{N}$. Furthermore $0 \notin \mathrm{rng}(\varphi_0^{b_0})$.
5. There is exactly one index i such that $\varphi_i^{b_0} \in \mathcal{R}$.

With this claim and our construction we can verify $b_0 \in B$. Now we assume by way of contradiction that $\mathcal{R}_{b_0} \in \mathrm{CEX}_\tau(\lambda x.S(b_0, x))$. It suffices to regard two cases.

Case 1. $\mathcal{R}_{b_0} = \{\varphi_0^{b_0}\}$.
Then on $\varphi_0^{b_0}$ the strategy $\lambda x.S(b_0, x)$ changes its hypothesis infinitely often or returns a hypothesis incorrect with respect to τ infinitely often. We obtain $\mathcal{R}_{b_0} \notin \mathrm{CEX}_\tau(\lambda x.S(b_0, x))$.

Case 2. $\mathcal{R}_{b_0} = \{\varphi_i^{b_0}\}$ with $i \geq 1$.
With Claim 1 we have $\tau_{S(b_0, \varphi_i^{b_0}[n])} \subseteq \alpha_{i-1}^{b_0}\uparrow^\infty \subset \alpha_{i-1}^{b_0} 0^\infty = \varphi_i^{b_0}$ for some $n \in \mathbb{N}$. Hence $S(b_0, \varphi_i^{b_0}[n])$ is a τ-number of a proper subfunction of $\varphi_i^{b_0}$. We conclude $\mathcal{R}_{b_0} \notin \mathrm{CEX}_\tau(\lambda x.S(b_0, x))$.

In each case we have $\mathcal{R}_{b_0} \notin \mathrm{CEX}_\tau(\lambda x.S(b_0, x))$. As $b_0 \in B$, this contradicts our initial assumption; so $B \notin \mathrm{suit}_\tau(J^*, \mathrm{CEX})$. This completes the proof. □

Thus it only remains to show that the separations in Lemma 1.3 are proper inclusions. From Theorem 9 and $\mathrm{suit}_\varphi(J^*, \mathrm{CONS}) \subseteq \mathrm{suit}_\varphi(J^*, \mathrm{EX})$ (analogously for suit_τ) we obtain that the second inclusion $\mathrm{suit}_\varphi(J^*, \mathrm{CEX}) \subseteq \mathrm{suit}_\varphi(J^*, \mathrm{EX})$ and its τ-version are indeed proper. For the first inclusion regard Theorem 10.

Theorem 10. *$suit_\varphi(J^*, CEX) \setminus suit(J^*, CONS) \neq \emptyset$.*[3]

Together with $\mathrm{suit}_\tau(J^*, \mathrm{TOTAL}) \subseteq \mathrm{suit}_\tau(J^*, \mathrm{CONS})$ this theorem yields $\mathrm{suit}_\varphi(J^*, \mathrm{CEX}) \setminus \mathrm{suit}_\tau(J^*, \mathrm{TOTAL}) \neq \emptyset$ and in particular $\mathrm{suit}_\varphi(J^*, \mathrm{TOTAL}) \subset \mathrm{suit}_\varphi(J^*, \mathrm{CEX})$, where again suit_φ may be replaced by suit_τ.

With Theorems 9 and 10 we have also verified the following corollary.

Corollary 1. *1. $suit_\varphi(J^*, CEX) \# suit_\varphi(J^*, CONS)$,*
2. $suit_\tau(J^*, CEX) \# suit_\tau(J^*, CONS)$.

Now we can summarize our separation results for uniform learning of finite classes with respect to fixed hypothesis spaces.

Theorem 11. *1. $suit_\varphi(J^*, EX_m) \subset suit_\varphi(J^*, EX_{m+1}) \subset suit_\varphi(J^*, EX)$ for arbitrary $m \in \mathbb{N}$,*
2. $suit_\varphi(J^*, TOTAL) \subset suit_\varphi(J^*, CONS) \subset suit_\varphi(J^*, CONF) \subset suit_\varphi(J^*, EX)$,
3. $suit_\varphi(J^*, TOTAL) \subset suit_\varphi(J^*, CEX) \subset suit_\varphi(J^*, EX)$.
These results hold analogously if we substitute $suit_\varphi$ by $suit_\tau$.

Thus we have transferred the comparison results of Theorem 1 to the concept of meta-learning in fixed hypothesis spaces. Each separation is achieved already by restricting ourselves to the synthesis of strategies for finite classes of recursive functions.

5 Separation of Inference Criteria – General Hypothesis Spaces

In this section we investigate the hierarchies of inference criteria for uniform learning without restrictions in the choice of the hypothesis spaces. Again we will concentrate on description sets corresponding to collections of finite classes of recursive functions. Some of the comparison results in Section 4 hold analogously for this concept, but there are differences, too. Our first simple observations in Lemma 2 follow immediately from the definitions.

Lemma 2. *1. $suit(J^*, EX_m) \subseteq suit(J^*, EX_{m+1}) \subseteq suit(J^*, EX)$ for all $m \in \mathbb{N}$,*
2. $suit(J^*, CONS) \subseteq suit(J^*, CONF) \subseteq suit(J^*, EX)$,
3. $suit(J^*, TOTAL) \subseteq suit(J^*, CEX) \subseteq suit(J^*, EX)$.

[3] The proof is omitted but proceeds similar to the proof of Theorem 7.

Note that we dropped the inclusion for TOTAL-identification in the second line. Since in general a uniform strategy S satisfying $B \in \text{suit}(J^*, \text{TOTAL})(S)$ for some $B \subseteq \mathbb{N}$ can *not* synthesize an appropriate hypothesis space for $\mathcal{R}_b$ from $b \in B$, the hypotheses returned by S cannot be checked for consistency. Therefore the proof of Lemma 1 cannot be transferred. By Theorems 6, 7, 8 all inclusions in Lemma 2.1 and 2.2 are proper inclusions. But for the other separations we observe a different connection, as Theorem 12 states.

Theorem 12. *suit*$(J^*, \mathit{TOTAL}) = $ *suit*$(J^*, \mathit{CEX}) = $ *suit*(J^*, EX).

Proof. $\text{suit}(J^*, \text{TOTAL}) \subseteq \text{suit}(J^*, \text{CEX}) \subseteq \text{suit}(J^*, \text{EX})$ follows by definition. It remains to prove $\text{suit}(J^*, \text{EX}) \subseteq \text{suit}(J^*, \text{TOTAL})$. For that purpose fix a description set $B \in \text{suit}(J^*, \text{EX})$. Then we know

1. $\mathcal{R}_b$ is finite for all $b \in B$,
2. there is a strategy $S \in \mathcal{P}^2$ such that for any $b \in B$ there is a hypothesis space $\psi^{[b]} \in \mathcal{P}^2$ satisfying $\mathcal{R}_b \in \text{EX}_{\psi^{[b]}}(\lambda x.S(b, x))$.

Note that the hypothesis spaces $\psi^{[b]}$ do not have to be computable uniformly in b. Now we want to prove that $B \in \text{suit}(J^*, \text{TOTAL})$. We even will see that our given strategy S is already an appropriate strategy for uniform TOTAL-identification from B. This requires a change of the hypothesis spaces $\psi^{[b]}$ for $b \in B$.

Idea. Assume $b \in B$ was fixed. Since $\lambda x.S(b, x)$ identifies the finite class $\mathcal{R}_b$ in the limit, there are only finitely many initial segments of functions in $\mathcal{R}_b$ which force the strategy $\lambda x.S(b, x)$ into a "non-total" guess. If we replace the functions in $\psi^{[b]}$ associated with these non-total guesses by an element of $\mathcal{R}$ (for example 0^∞), we obtain a hypothesis space appropriate for TOTAL-identification of $\mathcal{R}_b$ by $\lambda x.S(b, x)$.

More formally: Fix $b \in B$. From 2 we obtain $\text{card}\,\{n \in \mathbb{N} \mid \psi^{[b]}_{S(b,f[n])} \notin \mathcal{R}\} < \infty$ for all $f \in \mathcal{R}_b$. Defining the set of "forbidden" hypotheses on "relevant" initial segments by

$$H^{[b]} := \{i \in \mathbb{N} \mid \psi^{[b]}_i \notin \mathcal{R} \wedge \exists f \in \mathcal{R}_b \; \exists n \in \mathbb{N} \; [S(b, f[n]) = i]\} \ ,$$

we conclude with statement 1, that $H^{[b]}$ is finite. Now we define a new hypothesis space $\eta^{[b]}$ by

$$\eta^{[b]}_i := \begin{cases} \psi^{[b]}_i & \text{if } i \notin H^{[b]} \\ 0^\infty & \text{if } i \in H^{[b]} \end{cases} \quad \text{for all } i \in \mathbb{N} \ .$$

Since $\psi^{[b]} \in \mathcal{P}^2$ and $H^{[b]}$ is finite, $\eta^{[b]}$ is computable. The definition of $\eta^{[b]}$ then implies $\mathcal{R}_b \in \text{TOTAL}_{\eta^{[b]}}(\lambda x.S(b, x))$. As $b \in B$ was chosen arbitrarily, we conclude $B \in \text{suit}(J^*, \text{TOTAL})$. □

With this result we also observe a difference to our separation of TOTAL and CONS in the classical learning model.

Corollary 2. *suit*$(J^*, CONS) \subset$ *suit*$(J^*, CONF) \subset$ *suit*$(J^*, TOTAL)$.

Proof. This fact follows immediately from Theorem 7 and Theorem 8 and by the result $\text{suit}(J^*, \text{EX}) = \text{suit}(J^*, \text{TOTAL})$ in Theorem 12. □

Obviously, a further change in the hierarchies of inference criteria is witnessed by the fact $\text{suit}(J^*, \text{CONS}) \subset \text{suit}(J^*, \text{CEX})$, which follows by the same argumentation as in the proof of Corollary 2. We summarize:

Theorem 13. *1.* *suit*$(J^*, EX_m) \subset$ *suit*$(J^*, EX_{m+1}) \subset$ *suit*(J^*, EX) *for arbitrary* $m \in \mathbb{N}$*,*
2. *suit*$(J^*, CONS) \subset$ *suit*$(J^*, CONF) \subset$ *suit*$(J^*, TOTAL) =$ *suit*$(J^*, CEX) =$ *suit*(J^*, EX).

So in contrast to uniform identification of finite classes with respect to fixed hypothesis spaces the separations in Theorem 1 cannot be transferred to the unrestricted concept of uniform learning. Still it is remarkable, how many inference criteria for uniform identification can be separated by collections of finite classes of functions – even with very strong results (cf. the remarks below Theorem 6).

References

1. Barzdins, J. (1974); *Inductive Inference of Automata, Functions and Programs*, In: Proceedings International Congress of Mathematicians, 455-460.
2. Baliga, G.; Case, J.; Jain, S. (1996); *Synthesizing Enumeration Techniques for Language Learning*, In: Proceedings of the Ninth Annual Conference on Computational Learning Theory, ACM Press, 169-180
3. Case, J.; Smith, C. (1983); *Comparison of Identification Criteria for Machine Inductive Inference*, Theoretical Computer Science 25, 193-220.
4. Freivalds, R.; Kinber, E.B.; Wiehagen, R. (1995); *How Inductive Inference Strategies Discover Their Errors*, Information and Computation 118, 208-226.
5. Gold, E.M. (1967); *Language Identification in the Limit*, Information and Control 10, 447-474.
6. Jantke, K.P. (1979); *Natural Properties of Strategies Identifying Recursive Functions*, Elektronische Informationsverarbeitung und Kybernetik 15, 487-496.
7. Jantke, K.P.; Beick, H. (1981); *Combining Postulates of Naturalness in Inductive Inference*, Elektronische Informationsverarbeitung und Kybernetik 17, 465-484.
8. Kapur, S.; Bilardi, G. (1992); *On Uniform Learnability of Language Families*, Information Processing Letters 44, 35-38.
9. Osherson, D.N.; Stob, M.; Weinstein, S. (1988); *Synthesizing Inductive Expertise*, Information and Computation 77, 138-161.
10. Rogers, H. (1987); *Theory of Recursive Functions and Effective Computability*, MIT Press, Cambridge, Massachusetts.
11. Wiehagen, R. (1978); *Zur Theorie der algorithmischen Erkennung*, Dissertation B, Humboldt-University, Berlin (in German).
12. Zilles, S. (2000); *On Uniform Learning of Classes of Recursive Functions*, Technical Report LSA-2000-05E, Centre for Learning Systems and Applications, University of Kaiserslautern.