On the Interplay Between Inductive Inference of Recursive Functions, Complexity Theory and Recursive Numberings

Thomas Zeugmann

Division of Computer Science,
Hokkaido University, Sapporo 060-0814, Japan
thomas@ist.hokudai.ac.jp

Abstract. The present paper surveys some results from the inductive inference of recursive functions, which are related to the characterization of inferrible function classes in terms of complexity theory, and in terms of recursive numberings. Some new results and open problems are also included.

1 Introduction

Inductive inference of recursive functions goes back to Gold [11], who considered *learning in the limit*, and has attracted a large amount of interest ever since. In learning in the limit an inference strategy S is successively fed the graph $f(0), f(1), \ldots$ of a recursive function in natural order and on every initial segment of it, the strategy has to output a hypothesis, which is a natural number. These numbers are interpreted as programs in a fixed Gödel numbering φ of all partial recursive functions over the natural numbers. The sequence of all hypotheses output on f has then to stabilize on a number i such that $\varphi_i = f$. A strategy infers a class $\mathcal{U}$ of recursive functions, if it infers every function from $\mathcal{U}$.

One of the most influential papers has been Blum and Blum [6], who introduced two types of characterizations of learnable function classes in terms of computational complexity. The gist behind such characterizations is that classes $\mathcal{U}$ of recursive functions are learnable with respect to a given learning criterion if and only if all functions in $\mathcal{U}$ possess a particular complexity theoretic property.

The learning criterion considered is reliable inference on the set $\mathcal{R}$, where $\mathcal{R}$ denotes the set of all recursive functions. We denote the family of all classes $\mathcal{U}$ which are reliably inferable by $\mathcal{R}$-REL. Here *reliability* means the learner converges on any function f from $\mathcal{R}$ iff it learns f in the limit. In the first version, operator honesty classes are used. If $\mathfrak{O}$ is a total effective operator then a function f is said to be $\mathfrak{O}$-honest if $\mathfrak{O}(f)$ is an upper bound for the complexity Φ_i for all but finitely many arguments of φ_i, where $\varphi_i = f$. Then Blum and Blum [6] showed that a class $\mathcal{U}$ is in $\mathcal{R}$-REL iff there is a total effective operator $\mathfrak{O}$ such that every function $f \in \mathcal{U}$ is $\mathfrak{O}$-honest. Operator honest characterizations are also called *a priori* characterizations.

In the second version, one considers functions which possess a fastest program modulo a general recursive operator $\mathfrak{O}$ (called $\mathfrak{O}$-compression index). Now, the *a posteriori* characterization is as follows: A class $\mathcal{U}$ is in $\mathcal{R}$-REL iff there is a general recursive operator $\mathfrak{O}$ such that every function from $\mathcal{U}$ has an $\mathfrak{O}$-compression index.

Combining these two characterizations yields that the family of operator honesty classes coincides with the family of the operator compressed classes.

While operator honesty characterizations have been obtained for many learning criterions (cf. [29] and the references therein), the situation concerning *a posteriori* characterizations is much less satisfactory. Some results were shown in [27], but many problems remain open. In particular, it would be quite interesting to have an *a posteriori* characterization for $\mathfrak{T}$-REL. The learning criterion $\mathfrak{T}$-REL is defined as $\mathcal{R}$-REL, but reliability is required on the set $\mathfrak{T}$ of all total functions. Note that $\mathfrak{T}\text{-REL} \subset \mathcal{R}\text{-REL}$. Also, we shall present an *a posteriori* characterization for the function classes which are $\mathcal{T}$-consistently learnable with δ-delay (cf. [1] for a formal definition). Intuitively speaking, a $\mathcal{T}$-consistent δ-delayed learning strategy correctly reflects all inputs seen so far except the last δ ones, where δ is a natural number.

Note that there are also prominent examples of learning criterions for which even *a priori* characterizations are missing. These include the *behaviorally correct* learnable functions classes (cf. [9, 8] for more information). So in both cases we also point to the open problem whether or not one can show the non-existence of such desired characterizations.

Moreover, in Blum and Blum [6] the *a posteriori* characterization of $\mathcal{R}$-REL has been used to show that some interesting function classes are in $\mathcal{R}$-REL, e.g., the class of approximations of the halting problem. Stephan and Zeugmann [22] extended these results to several classes based on approximations to non-recursive functions. Besides these results, our knowledge concerning the learnability of interesting function classes is severely limited, except the recursively enumerable functions classes (or subsets thereof), and with respect to function classes used to achieve separations.

Finally, the problem of suitable hypothesis spaces is considered. That is, instead of Gödel numberings one is interested in numberings having learner-friendly properties. Again, we survey some illustrative results, present some new ones, and outline open problems. Note that one can also combine the results obtained in this setting with the results mentioned above, i.e., one can derive some complexity theoretic properties of such numberings.

2 Preliminaries

Unspecified notations follow Rogers [21]. By $\mathbb{N} = \{0, 1, 2, \ldots\}$ we denote the set of all natural numbers. The set of all finite sequences of natural numbers is denoted by $\mathbb{N}^*$. For $a, b \in \mathbb{N}$ we define $a \dot{-} b$ to be $a - b$ if $a \geq b$ and 0, otherwise.

The cardinality of a set S is denoted by $|S|$. We write $\wp(S)$ for the power set of set S. Let $\emptyset$, $\in$, $\subset$, $\subseteq$, $\supset$, $\supseteq$, and $\#$ denote the empty set, element of, proper subset, subset, proper superset, superset, and incomparability of sets, respectively.

By $\mathfrak{P}$ and $\mathfrak{T}$ we denote the set of all partial and total functions of one variable over $\mathbb{N}$, respectively. The classes of all partial recursive and recursive functions of one, and two arguments over $\mathbb{N}$ are denoted by $\mathcal{P}$, $\mathcal{P}^2$, $\mathcal{R}$, and $\mathcal{R}^2$, respectively. Furthermore, for any $f \in \mathfrak{P}$ we use $\mathrm{dom}(f)$ to denote the *domain* of the function f, i.e., $\mathrm{dom}(f) =_{df} \{x \mid x \in \mathbb{N},\ f(x) \text{ is defined}\}$. Additionally, by $\mathrm{range}(f)$ we denote the *range* of f, i.e., $\mathrm{range}(f) =_{df} \{f(x) \mid x \in \mathrm{dom}(f)\}$. Let $f, g \in \mathfrak{P}$ be any partial functions. We write $f \subseteq g$ if for all $x \in \mathrm{dom}(f)$ the condition $f(x) = g(x)$ is satisfied. By $\mathcal{R}_{0,1}$ and $\mathcal{R}_{mon}$ we denote the set of all $\{0,1\}$-valued recursive functions (recursive predicates) and of all monotone recursive functions, respectively.

Every function $\psi \in \mathcal{P}^2$ is said to be a *numbering.* Let $\psi \in \mathcal{P}^2$, then we write ψ_i instead of $\lambda x.\psi(i,x)$, set $\mathcal{P}_\psi = \{\psi_i \mid i \in \mathbb{N}\}$ and $\mathcal{R}_\psi = \mathcal{P}_\psi \cap \mathcal{R}$. Consequently, if $f \in \mathcal{P}_\psi$, then there is a number i such that $f = \psi_i$. If $f \in \mathcal{P}$ and $i \in \mathbb{N}$ are such that $\psi_i = f$, then i is called a *ψ-program for f*. Let ψ be any numbering, and let $i \in \mathbb{N}$; if $\psi_i(x)$ is defined (abbr. $\psi_i(x)\downarrow$) then we also say that $\psi_i(x)$ *converges.* Otherwise, $\psi_i(x)$ is said to *diverge* (abbr. $\psi_i(x)\uparrow$).

A numbering $\varphi \in \mathcal{P}^2$ is called a Gödel numbering (cf. Rogers [21]) if $\mathcal{P}_\varphi = \mathcal{P}$, and for every numbering $\psi \in \mathcal{P}^2$, there is a $c \in \mathcal{R}$ such that $\psi_i = \varphi_{c(i)}$ for all $i \in \mathbb{N}$. *Göd* denotes the set of all Gödel numberings. Furthermore, we write (φ, Φ) to denote any *complexity measure* as defined in Blum [7]. That is, $\varphi \in \mathit{G\ddot{o}d}$, $\Phi \in \mathcal{P}^2$ and (1) $\mathrm{dom}(\varphi_i) = \mathrm{dom}(\Phi_i)$ for all $i \in \mathbb{N}$ and (2) the predicate "$\Phi_i(x) = y$" is uniformly recursive for all $i, x, y \in \mathbb{N}$.

Moreover, let $\mathrm{NUM} = \{\mathcal{U} \mid \exists\psi[\psi \in \mathcal{R}^2 \wedge \mathcal{U} \subseteq \mathcal{P}_\psi]\}$ denote the family of all subsets of all recursively enumerable classes of recursive functions.

Furthermore, using a fixed encoding $\langle \dots \rangle$ of $\mathbb{N}^*$ onto $\mathbb{N}$ we write f^n instead of $\langle (f(0), \dots, f(n)) \rangle$, for any $n \in \mathbb{N}$, $f \in \mathcal{R}$.

The quantifier $\forall^\infty$ stands for "almost everywhere" and means "all but finitely many." Finally, a sequence $(j_n)_{j \in \mathbb{N}}$ of natural numbers is said to *converge* to the number j if all but finitely many numbers of it are equal to j. Next we define some concepts of learning.

Definition 1 (Gold [11, 12]). *Let $\mathcal{U} \subseteq \mathcal{R}$ and let $\psi \in \mathcal{P}^2$. The class $\mathcal{U}$ is said to be* learnable in the limit with respect to ψ *if there is a strategy $S \in \mathcal{P}$ such that for each function $f \in \mathcal{U}$,*

(1) *for all $n \in \mathbb{N}$, $S(f^n)$ is defined,*
(2) *there is a $j \in \mathbb{N}$ such that $\psi_j = f$ and the sequence $(S(f^n))_{n \in \mathbb{N}}$ converges to j.*

If a class $\mathcal{U}$ is learnable in the limit with respect to ψ by a strategy S, then we write $\mathcal{U} \in \mathrm{LIM}_\psi(S)$. Let $\mathrm{LIM}_\psi = \{\mathcal{U} \mid \mathcal{U}$ is learnable in the limit w.r.t. $\psi\}$, and define $\mathrm{LIM} = \bigcup_{\psi \in \mathcal{P}^2} \mathrm{LIM}_\psi$.

As far as the semantics of the hypotheses output by a strategy S is concerned, whenever S is defined on input f^n, then we always interpret the number $S(f^n)$ as a ψ–number. This convention is adopted to all the definitions below. Furthermore, note that $\mathrm{LIM}_\varphi = \mathrm{LIM}$ for any $\varphi \in \mathit{Göd}$. In the above definition LIM stands for "limit."

Looking at Definition 1 one may be tempted to think that it is too general. Maybe we should add some requirements that seem very natural. Since it may be hard for a strategy to know which inputs may occur, it could be very convenient to require $S \in \mathcal{R}$. Furthermore, if the strategy outputs a program i such that $\varphi_i \notin \mathcal{R}$, then this output cannot be correct. Hence, it seems natural to require the strategy to output exclusively hypotheses describing recursive functions. These demands directly yield the following definition:

Definition 2 (Wiehagen [24]). *Let $\mathcal{U} \subseteq \mathcal{R}$ and let $\psi \in \mathcal{P}^2$. The class $\mathcal{U}$ is said to be* $\mathcal{R}$–totally learnable with respect to ψ *if there is a strategy $S \in \mathcal{R}$ such that*

(1) $\psi_{S(n)} \in \mathcal{R}$ *for all* $n \in \mathbb{N}$,
(2) *for each $f \in \mathcal{U}$ there is a $j \in \mathbb{N}$ such that $\psi_j = f$, and $(S(f^n))_{n\in\mathbb{N}}$ converges to j.*

$\mathcal{R}$-$\mathrm{TOTAL}_\psi(S)$, $\mathcal{R}$-TOTAL_ψ, *and* $\mathcal{R}$-TOTAL *are defined in analogy to the above.*

However, now it is not difficult to show that $\mathcal{R}$-TOTAL $=$ NUM (cf. Zeugmann and Zilles [29, Theorem 2]). This is the first characterization of a learning type in terms of recursive numberings. This characterization shows how $\mathcal{R}$-total learning can be achieved, i.e., by using the well-known *identification by enumeration* technique.

Next, we recall the definition of reliable learning introduced by Blum and Blum [6] and Minicozzi [19]. Intuitively, a learner M is reliable provided it converges if and only if it learns.

Definition 3 (Blum and Blum [6], Minicozzi [19]). *Let $\mathcal{U} \subseteq \mathcal{R}$, $\mathcal{M} \subseteq \mathfrak{T}$ and let $\varphi \in \mathit{Göd}$. The class $\mathcal{U}$ is said to be* reliably learnable on $\mathcal{M}$ *if there is a strategy $S \in \mathcal{R}$ such that*

(1) $\mathcal{U} \in \mathrm{LIM}_\varphi(S)$, *and*
(2) *for all functions $f \in \mathcal{M}$, if the sequence $(S(f^n))_{n\in\mathbb{N}}$ converges, say to j, then $\varphi_j = f$.*

Let $\mathcal{M}$-REL *denote the family of all classes $\mathcal{U}$ that are reliably learnable on $\mathcal{M}$.*

Note that neither in Definition 1 nor in Definition 3 a requirement is made concerning the intermediate hypotheses output by the strategy S. The following definition is obtained from Definition 1 by adding the requirement that S correctly reflects all but the last δ data seen so far.

Definition 4 (Akama and Zeugmann [1]). *Let* $\mathcal{U} \subseteq \mathcal{R}$*, let* $\psi \in \mathcal{P}^2$ *and let* $\delta \in \mathbb{N}$*. The class* $\mathcal{U}$ *is called* consistently learnable in the limit with δ-delay with respect to ψ *if there is a strategy* $S \in \mathcal{P}$ *such that*

(1) $\mathcal{U} \in \mathrm{LIM}_\psi(S)$,
(2) $\psi_{S(f^n)}(x) = f(x)$ *for all* $f \in \mathcal{U}$, $n \in \mathbb{N}$ *and all* x *such that* $x + \delta \leq n$.

We define $\mathrm{CONS}^\delta_\psi(S)$, $\mathrm{CONS}^\delta_\psi$, *and* CONS^δ *analogously to the above.*

We note that for $\delta = 0$ Barzdin's [4] original definition of CONS is obtained. We therefore usually omit the upper index δ if $\delta = 0$. This is also done for the other version of consistent learning defined below. We use the term δ*-delay*, since a consistent strategy with δ-delay correctly reflects all but at most the last δ data seen so far. If a strategy S learns a function class $\mathcal{U}$ in the sense of Definition 4, then we refer to S as a δ*-delayed consistent* strategy.

In Definition 4 consistency with δ-delay is only demanded for inputs that correspond to some function f from the target class $\mathcal{U}$. Note that for $\delta = 0$ the following definition incorporates Wiehagen and Liepe's [23] requirement on a strategy to work consistently on all inputs.

Definition 5 (Akama and Zeugmann [1]). *Let* $\mathcal{U} \subseteq \mathcal{R}$*, let* $\psi \in \mathcal{P}^2$ *and let* $\delta \in \mathbb{N}$*. The class* $\mathcal{U}$ *is called* $\mathcal{T}$–consistently learnable in the limit with δ-delay with respect to ψ *if there is a strategy* $S \in \mathcal{R}$ *such that*

(1) $\mathcal{U} \in \mathrm{CONS}^\delta_\psi(S)$,
(2) $\psi_{S(f^n)}(x) = f(x)$ *for all* $f \in \mathcal{R}$, $n \in \mathbb{N}$ *and all* x *such that* $x + \delta \leq n$.

We define $\mathcal{T}\text{-}\mathrm{CONS}^\delta_\psi(S)$, $\mathcal{T}\text{-}\mathrm{CONS}^\delta_\psi$, *and* $\mathcal{T}\text{-}\mathrm{CONS}^\delta$ *analogously to the above.*

We note that for all $\delta \in \mathbb{N}$ and all learning types $\mathrm{LT} \in \{\mathrm{CONS}^\delta, \mathcal{T}\text{-}\mathrm{CONS}^\delta\}$ we have $\mathrm{LT}_\varphi = \mathrm{LT}$ for every $\varphi \in G\ddot{o}d$.

Finally, we look at another mode of convergence which goes back to Feldman [9], who called it *matching in the limit* and considered it in the setting of learning languages. The difference to the mode of convergence used in Definition 1, which is actually *syntactic convergence*, is to relax the requirement that the sequence of hypotheses has to converge to a correct program, by *semantic convergence*. Here by semantic convergence we mean that after some point *all hypotheses are correct* but not necessarily identical. Nowadays, the resulting learning model is usually referred to as *behaviorally correct* learning. This term was coined by Case and Smith [8]. As far as learning of recursive functions is concerned, behaviorally correct learning was formalized by Barzdin [2, 3].

Definition 6 (Barzdin [2, 3]). *Let* $\mathcal{U} \subseteq \mathcal{R}$ *and let* $\psi \in \mathcal{P}^2$*. The class* $\mathcal{U}$ *is said to be* behaviorally correctly learnable with respect to ψ *if there is a strategy* $S \in \mathcal{P}$ *such that for each function* $f \in \mathcal{U}$,

(1) *for all* $n \in \mathbb{N}$, $S(f^n)$ *is defined,*
(2) $\psi_{S(f^n)} = f$ *for all but finitely many* $n \in \mathbb{N}$.

If $\mathcal{U}$ *is behaviorally correctly learnable with respect to* ψ *by a strategy* S, *we write* $\mathcal{U} \in \mathrm{BC}_\psi(S)$. BC_ψ *and* BC *are defined analogously to the above above.*

3 Characterizations in Terms of Complexity

We continue with characterizations in terms of computational complexity. Characterizations are a useful tool to get a better understanding of what different learning types have in common and where the differences are. They may also help to overcome difficulties that arise in the design of powerful learning algorithms.

Let us recall the needed definitions of several types of computable operators. Let $(F_x)_{x\in\mathbb{N}}$ be the canonical enumeration of all finite functions.

Definition 7 (Rogers [21]). *A mapping $\mathfrak{O}\colon \mathfrak{P} \mapsto \mathfrak{P}$ from partial functions to partial functions is called a* partial recursive operator *if there is a recursively enumerable set $W \subset \mathbb{N}^3$ such that for any $y, z \in \mathbb{N}$ it holds that $\mathfrak{O}(f)(y) = z$ if there is $x \in \mathbb{N}$ such that $(x, y, z) \in W$ and f extends the finite function F_x.*

Furthermore, $\mathfrak{O}$ is said to be a general recursive operator *if $\mathfrak{T} \subseteq \mathrm{dom}(\mathfrak{O})$, and $f \in \mathfrak{T}$ implies $\mathfrak{O}(f) \in \mathfrak{T}$.*

A mapping $\mathfrak{O}\colon \mathcal{P} \mapsto \mathcal{P}$ is called an effective operator *if there is a function $g \in \mathcal{R}$ such that $\mathfrak{O}(\varphi_i) = \varphi_{g(i)}$ for all $i \in \mathbb{N}$. An effective operator $\mathfrak{O}$ is said to be* total effective *provided that $\mathcal{R} \subseteq \mathrm{dom}(\mathfrak{O})$, and $\varphi_i \in \mathcal{R}$ implies $\mathfrak{O}(\varphi_i) \in \mathcal{R}$.*

For more information about general recursive operators and effective operators we refer the reader to [14, 20, 28]. If $\mathfrak{O}$ is an operator which maps functions to functions, we write $\mathfrak{O}(f, x)$ to denote the value of the function $\mathfrak{O}(f)$ at the argument x.

Definition 8. *A partial recursive operator $\mathfrak{O}\colon \mathfrak{P} \mapsto \mathfrak{P}$ is said to be* monotone *if for all functions $f, g \in \mathrm{dom}(\mathfrak{O})$ the following condition is satisfied:*
If $\forall^\infty x[f(x) \le g(x)]$ then $\forall^\infty x[\mathfrak{O}(f, x) \le \mathfrak{O}(g, x)]$.

Let $\mathfrak{O}$ be any arbitrarily fixed operator and let $M \subseteq \mathfrak{P}$. Then the abbreviation "$\mathfrak{O}(M) \subseteq M$" stands for "$M \subseteq \mathrm{dom}(\mathfrak{O})$ and $f \in \mathcal{M}$ implies that $\mathfrak{O}(f) \in M$."

Any computable operator can be realized by a 3-tape Turing machine T which works as follows: If for an arbitrary function $f \in \mathrm{dom}(\mathfrak{O})$, all pairs $(x, f(x))$, $x \in \mathrm{dom}(f)$ are written down on the input tape of T (repetitions are allowed), then T will write exactly all pairs $(x, \mathfrak{O}(f, x))$ on the output tape of T (under unlimited working time).

Let $\mathfrak{O}$ be a partial recursive operator, a general recursive operator or a total effective operator. Then, for $f \in \mathrm{dom}(\mathfrak{O})$, $m \in \mathbb{N}$ we set: $\varDelta\mathfrak{O}(f, m) =$ "the least n such that, for all $x \le n$, $f(x)$ is defined and, for the computation of $\mathfrak{O}(f, m)$, the Turing machine T only uses the pairs $(x, f(x))$ with $x \le n$; if such an n does not exist, we set $\varDelta\mathfrak{O}(f, m) = \infty$."

For any function $u \in \mathcal{R}$ we define $\varOmega_u$ to be the set of all partial recursive operators $\mathfrak{O}$ satisfying $\varDelta\mathfrak{O}(f, m) \le u(m)$ for all $f \in \mathrm{dom}(\mathfrak{O})$. For the sake of notation, below we shall use $\mathrm{id} + \delta$, $\delta \in \mathbb{N}$, to denote the function $u(x) = x + \delta$ for all $x \in \mathbb{N}$.

Blum and Blum [6] initiated the characterization of learning types in terms of computational complexity. Here they distinguished between *a priori characterizations* and *a posteriori characterizations.* In order to obtain an *a priori* *characterization* one starts from classes of *operator honesty complexity classes*, which are defined as follows: Let $\mathfrak{O}$ be a computable operator. Then we define

$$\mathcal{C}_{\mathfrak{O}} =_{df} \{f \mid \exists i[\varphi_i = f \wedge \forall^\infty x[\Phi_i(x) \leq \mathfrak{O}(f,x)]]\} \cap \mathcal{R}\ . \tag{1}$$

That is, every function in $\mathcal{C}_{\mathfrak{O}}$ possesses a program i such that the complexity of program i is in a computable way bounded by its function values, namely by $\mathfrak{O}(f,x)$ almost everywhere. So let LT be any learning type, e.g., learning in the limit. Then the general form of an *a priori characterization* of LT looks as follows:

Theorem 1. *Let $\mathcal{U} \subseteq \mathcal{R}$ be any class. Then we have $\mathcal{U} \in$ LT if and only if there is a computable operator $\mathfrak{O}$ such that $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$, where the operator $\mathfrak{O}$ has to fulfill some additional properties.*

Example 1. Consider the set of all operators which can be defined as follows: For any $t \in \mathcal{R}$ we define $\mathfrak{O}(f,x) =_{df} t(x)$ for every $f \in \mathcal{R}$ and $x \in \mathbb{N}$. Then the complexity classes defined in (1) have the form

$$\mathcal{C}_t = \{f \mid \exists i[\varphi_i = f \wedge \forall^\infty x[\Phi_i(x) \leq t(x)]]\} \cap \mathcal{R}\ . \tag{2}$$

and Theorem 1 yields the following *a priori* characterization of $\mathcal{R}$-TOTAL:

Let $\mathcal{U} \subseteq \mathcal{R}$ be any class. Then we have $\mathcal{U} \in \mathcal{R}$-TOTAL if and only if there is a recursive function $t \in \mathcal{R}$ such that $\mathcal{U} \subseteq \mathcal{C}_t$.

Since this theorem holds obviously also in case that $\mathcal{U} = \mathcal{C}_t$, we can directly use the fact that $\mathcal{R}$-TOTAL $=$ NUM and conclude that $\mathcal{C}_t \in$ NUM for every $t \in \mathcal{R}$. Thus, using the *a priori* characterization of $\mathcal{R}$-TOTAL we could easily reprove $\mathcal{C}_t \in$ NUM, which was originally shown by McCreight and Meyer [17].

Example 2. Note that for every general recursive operator $\mathfrak{O}$ there is a monotone general recursive operator $\mathfrak{M}$ such that $\mathfrak{O}(f,x) \leq \mathfrak{M}(f,x)$ for every function $f \in \mathfrak{T}$ and almost all $x \in \mathbb{N}$ (cf. Meyer and Fischer [18]). Furthermore, Grabowski [13] proved the following *a priori* characterization of $\mathfrak{T}$-REL:

Let $\mathcal{U} \subseteq \mathcal{R}$ be any class. Then we have $\mathcal{U} \in \mathfrak{T}$-REL if and only if there exists a general recursive operator $\mathfrak{O}$ such that $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.

Using that every function $f \in \mathcal{R}_{0,1}$ satisfies $f(x) \leq 1$ for all $x \in \mathbb{N}$ we directly see by an easy application of Meyer and Fischer's [18] result that

$$\mathfrak{T}\text{-REL} \cap \wp(\mathcal{R}_{0,1}) = \mathcal{R}\text{-TOTAL} \cap \wp(\mathcal{R}_{0,1}) \ = \ \text{NUM} \cap \wp(\mathcal{R}_{0,1})\ ; \tag{3}$$

i.e., reliable learning on the total functions restricted to classes of recursive predicates is exactly as powerful as $\mathcal{R}$-total learning restricted to classes of recursive predicates. On the other hand, $\mathcal{R}$-TOTAL $\subset \mathfrak{T}$-REL (cf. Grabowski [13]).

Moreover, Stephan and Zeugmann [22] showed that

$$\text{NUM} \cap \wp(\mathcal{R}_{0,1}) \subset \mathcal{R}\text{-REL} \cap \wp(\mathcal{R}_{0,1})\ . \tag{4}$$

The latter result was already published in Grabowski [13], but the new proof is much easier. It uses the class of approximations to the halting problem that has been considered in [6]. This class is defined as follows: Let (φ, Φ) be any complexity measure, and let $\tau \in \mathcal{R}$ be such that for all $i \in \mathbb{N}$

$$\varphi_{\tau(i)}(x) =_{df} \begin{cases} 1, & \text{if } \Phi_i(x)\downarrow \text{ and } \Phi_x(x) \leq \Phi_i(x) \ ; \\ 0, & \text{if } \Phi_i(x)\downarrow \text{ and } \neg[\Phi_x(x) \leq \Phi_i(x)] \ ; \\ \uparrow, & \text{otherwise} \ . \end{cases}$$

Now, we set $\mathcal{B} = \{\varphi_{\tau(i)} \mid i \in \mathbb{N} \text{ and } \Phi_i \in \mathcal{R}_{mon}\}$. Then in [22, Theorems 2 and 3] both results were proved $\mathcal{B} \notin \text{NUM}$ and $\mathcal{B} \in \mathcal{R}\text{-REL}$ yielding (4). We shall come back to this class.

Now, one can combine this with the *a priori* characterization of $\mathcal{R}$-REL obtained by Blum and Blum [6], which is as follows:

Let $\mathcal{U} \subseteq \mathcal{R}$ be any class. Then we have $\mathcal{U} \in \mathcal{R}\text{-REL}$ if and only if there exists a total effective operator $\mathfrak{O}$ such that $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.

Note that the difference between the *a priori* characterization of $\mathfrak{T}$-REL and $\mathcal{R}$-REL is that the operator $\mathfrak{O}$ is general recursive and total effective, respectively.

Putting this all together, we directly see that Meyer and Fischer's [18] boundability theorem cannot be strengthened by replacing "general recursive operator" by "total effective operator." And it also allows to show that there is an operator honesty complexity class $\mathcal{C}_{\mathfrak{O}}$ generated by a total effective operator $\mathfrak{O}$ such that $\mathcal{C}_{\mathfrak{O}} \nsubseteq \mathcal{C}_{\widetilde{\mathfrak{O}}}$ for every general recursive operator $\widetilde{\mathfrak{O}}$. For an explicit construction of such an operator $\mathfrak{O}$ we refer the reader to [14, 28].

Furthermore, Theorem 1 can be precisely stated for LIM, CONS^{δ}, and $\mathcal{T}\text{-CONS}^{\delta}$ by using techniques from Blum and Blum [6], Wiehagen [24] and Akama and Zeugmann [1]. The proofs can be found in [29, Theorems 37,35,34].

Let $\mathcal{U} \subseteq \mathcal{R}$, then we have Let $\mathcal{U} \subseteq \mathcal{R}$, then we have $\mathcal{U} \in \text{LIM}$ if and only if there exists an effective operator $\mathfrak{O}$ such that $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.

Let $\mathcal{U} \subseteq \mathcal{R}$ and let $\delta \in \mathbb{N}$; then we have

(1) *$\mathcal{U} \in \text{CONS}^{\delta}$ if and only if there exists an effective operator $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.*
(2) *$\mathcal{U} \in \mathcal{T}\text{-CONS}^{\delta}$ if and only if there is a general recursive operator $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.*

These *a priori* characterizations shed also additional light to the fact that the learning types $\mathfrak{T}$-REL, $\mathcal{R}$-REL, and $\mathcal{T}\text{-CONS}^{\delta}$ are closed under union, while LIM and CONS^{δ} are not. In the former the operator $\mathfrak{O}$ maps $\mathcal{R}$ to $\mathcal{R}$, and in the latter we only have $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$.

As we have seen, operator honesty characterizations have been found for many learning types, but some important ones are missing. These include BC, TOTAL, and conform learning. The learning criterion TOTAL is obtained from Definition 2 by replacing $S \in \mathcal{R}$ by $S \in \mathcal{P}$ and adding $S(f^n) \in \mathcal{R}$ for all $f \in \mathcal{U}$ and all $n \in \mathbb{N}$. Conform learning is a modification of consistent learning, where the requirement to correctly reflect all the functions values seen so far is replaced by the demand that the hypothesis output does never convergently contradict inputs already seen (cf. [29, Definition 22]).

Blum and Blum [6] also initiated the study of *a posteriori* characterizations of learning types. In particular they showed that any class $\mathcal{U} \in \mathcal{R}$-REL can be characterized in a way such that there exists a general recursive operator $\mathfrak{O}$ for which every function from $\mathcal{U}$ is *everywhere* $\mathfrak{O}$*-compressed.*

For the sake of completeness we include here the definition of everywhere $\mathfrak{O}$-compressed.

Definition 9 (Blum and Blum [6]). *Let* (φ, Φ) *be a complexity measure, let* $f \in \mathcal{R}$, *and let* $\mathfrak{O}$ *be a general recursive operator. Then a program* $i \in \mathbb{N}$ *is said to be an* $\mathfrak{O}$-compression index of f (*relative to* (φ, Φ)) *if*

(1) $\varphi_i = f$,
(2) $\forall j[\varphi_j = f \rightarrow \forall x\, \Phi_i(x) \leq \mathfrak{O}(\Phi_j, \max\{i, j, x\})]$.

In this case we also say that the function f *is* everywhere $\mathfrak{O}$-compressed.

We note that Definition 9 formalizes the concept of a fastest program (modulo an operator $\mathfrak{O}$) in a useful way. The $\mathfrak{O}$-compression index i satisfies the condition $\Phi_i(x) \leq \mathfrak{O}(\Phi_j, x)$ for all but finitely many $x \in \mathbb{N}$ and all programs j computing the same function as program i does. Additionally, it also provides an upper bound for the least argument n such that $\Phi_i \leq \mathfrak{O}(\Phi_j, x)$ for all $x > n$, i.e., $\max\{j, i\}$, and a computable majorante for those values $m \leq n$ for which possibly $\Phi_i(m) > \mathfrak{O}(\Phi_j, m)$; i.e., the value $\mathfrak{O}(\Phi_j, \max\{i, j\}$.

Of course, one can also consider the notion of everywhere $\mathfrak{O}$-compressed functions for total effective operators or any other type of computable operator $\mathfrak{O}$ provided that all considered complexity functions Φ_j are in $\mathrm{dom}(\mathfrak{O})$.

Theorem 2 (Blum and Blum [6]). *For every class* $\mathcal{U} \subseteq \mathcal{R}$ *we have the following:* $\mathcal{U} \in \mathcal{R}$-REL *if and only if there is a general recursive operator* $\mathfrak{O}$ *such that every function from* $\mathcal{U}$ *is everywhere* $\mathfrak{O}$*-compressed.*

However, in [6] it remained open whether or not one can also reliably learn on $\mathcal{R}$ an $\mathfrak{O}$-compression index for every function f in the target class $\mathcal{U}$. We were able to show (cf. [26]) that this is not always the case, when using the algorithm described in [6]. Furthermore, in [27] we provided a suitable modification of Definition 9 resulting in a *reliable* $\mathfrak{O}$*-compression index*, and then showed that such reliable $\mathfrak{O}$-compression indices are reliable learnable on $\mathcal{R}$.

On the other hand, Blum and Blum [6, Section 8] used Theorem 2 to show that several interesting functions classes are contained in $\mathcal{R}$-REL including the class $\mathcal{B}$ of approximations to the halting problem. Using different techniques, this result was extend in [22]. Conversely, one can also consider any particular general recursive operator *op* and ask for the resulting funcion class of everywhere *op*-compressed functions, which are, via Theorem 2, known to be in $\mathcal{R}$-REL. Unfortunately, almost nothing is known in this area. Therefore, we would like to encourage research along these two lines, i.e., considering interesting function classes and figuring out to which learning type they belong, or to study special general recursive operators with respect to the learning power they generate.

In order to characterize the learning type $\mathcal{T}$-CONS$^{\delta}$, the following modification of Definition 9 turned out to be suitable:

Definition 10. *Let* (φ, Φ) *be a complexity measure, let* $f \in \mathcal{R}$, *and let* $\mathfrak{O}$ *be a general recursive operator. Then a program* $i \in \mathbb{N}$ *is said to be an* absolute $\mathfrak{O}$-compression index of f (*relative to* (φ, Φ)) *if*

(1) $\varphi_i = f$,
(2) $\forall j \forall x[\varphi_j(y) = f(y)$ *for all* $y \leq \Delta\mathfrak{O}(\Phi_j, \max\{i, x\})$
$\rightarrow \Phi_i(x) \leq \mathfrak{O}(\Phi_j, \max\{i, x\})]$.

In this case we also say that the function f *is* absolutely $\mathfrak{O}$-compressed.

To show the following lemma we have to restrict the class of complexity measures a bit. We shall say that a complexity measure (φ, Φ) satisfies Property $(+)$ if for all $i, x \in \mathbb{N}$ such that $\Phi_i(x)$ is defined the condition $\Phi_i(x) \geq \varphi_i(x)$ is satisfied.

Note that Property $(+)$ is not very restrictive, since various "natural" complexity measures satisfy it.

Lemma 1. *Let* (φ, Φ) *be a complexity measure satisfying Property* $(+)$, *and let* $\delta \in \mathbb{N}$ *be arbitrarily fixed. Furthermore, let* $\mathcal{U} \in \mathcal{T}\text{-}\mathrm{CONS}^{\delta}$. *Then there is a general recursive operator* $\mathfrak{O} \in \Omega_{\mathrm{id}+\delta}$ *such that every function from* $\mathcal{U}$ *is absolutely* $\mathfrak{O}$*-compressed.*

The following lemma shows that the condition presented in Lemma 1 is also sufficient. Furthermore, this lemma holds for all complexity measures.

Lemma 2. *Let* (φ, Φ) *be any complexity measure, let* $\delta \in \mathbb{N}$ *be arbitrarily fixed, let* $\mathfrak{O} \in \Omega_{\mathrm{id}+\delta}$, *and let* $\mathcal{U} \subseteq \mathcal{R}$ *such that every function from* $\mathcal{U}$ *is absolutely* $\mathfrak{O}$*-compressed. Then there is a strategy* $S \in \mathcal{R}$ *such that*

(1) $\mathcal{U} \in \mathcal{T}\text{-}\mathrm{CONS}^{\delta}_{\varphi}(S)$,
(2) *for every* $f \in \mathcal{U}$ *the sequence* $(S(f^n))_{n \in \mathbb{N}}$ *converges to an absolute* $\mathfrak{O}$*-compression index of* f.

Furthermore, Lemmata 1 and 2 directly allow for the following theorem:

Theorem 3. *Let* (φ, Φ) *be a complexity measure satisfying Property* $(+)$, *let* $\delta \in \mathbb{N}$ *be arbitrarily fixed, and let* $\mathcal{U} \subseteq \mathcal{R}$. *Then we have*

$\mathcal{U} \in \mathcal{T}\text{-}\mathrm{CONS}^{\delta}_{\varphi}(S)$ *if and only if there is an operator* $\mathfrak{O} \in \Omega_{\mathrm{id}+\delta}$ *such that every function* f *from* $\mathcal{U}$ *is absolutely* $\mathfrak{O}$*-compressed. Furthermore, for every* $f \in \mathcal{U}$ *the sequence* $(S(f^n))_{n \in \mathbb{N}}$ *converges to an absolute* $\mathfrak{O}$*-compression index of* f.

Though we succeeded to show an *a posteriori* characterization for the learning type $\mathcal{T}\text{-}\mathrm{CONS}^{\delta}$, it is not completely satisfactory, since it restricts the class of admissible complexity measures. Can this restriction be removed?

Nevertheless, combining the *a priori* characterization of $\mathcal{T}\text{-}\mathrm{CONS}^{\delta}$ with the *a posteriori* characterization provided in Theorem 3 shows that the family of operator honesty classes coincides for every $\delta \in \mathbb{N}$ with the family of absolutely operator compressed classes.

In this regard, it would be very nice to have also an *a posteriori* characterization of $\mathfrak{T}$-REL.

4 Characterizations in Terms of Computable Numberings

The reader may be curious why our definitions of learning types include a numbering ψ with respect to which we aim to learn. After all, if one can learn a class $\mathcal{U}$ with respect to some numbering ψ, then one can also infer $\mathcal{U}$ with respect to any Gödel numbering φ. However, ψ may possess properties which facilitate learning. For example, since $\mathcal{R}$-TOTAL $=$ NUM, for every class $\mathcal{U} \in \mathcal{R}$-TOTAL there is numbering $\psi \in \mathcal{R}^2$ such that $\mathcal{U} \subseteq \mathcal{R}_\psi$, and so the identification by enumeration technique over ψ always succeeds.

Next, one may consider *measurable* numberings, which are defined as follows: A numbering $\psi \in \mathcal{P}^2$ is said to be *measurable* if the predicate "$\psi_i(x) = y$" is uniformly recursive in i, x, y (cf. Blum [7]). So, if $\psi \in \mathcal{P}^2$ is a measurable numbering and $\mathcal{U} \subseteq \mathcal{R}$ is such that $\mathcal{U} \subseteq \mathcal{P}_\psi$, then the identification by enumeration technique is still applicable. A prominent example is the class $\mathcal{U} = \{\Phi_i \mid i \in \mathbb{N}\} \cap \mathcal{R}$, where (φ, Φ) is a complexity measure. Note that the halting problem for the numbering $\Phi \in \mathcal{P}^2$ is undecidable (cf. [25, Lemma 3]).

Blum and Blum [6] also considered $\mathcal{P}$-REL and $\mathfrak{P}$-REL (cf. Definition 3 for $\mathcal{M} = \mathcal{P}$ and $\mathcal{M} = \mathfrak{P}$, respectively) and showed that $\mathcal{P}$-REL $=$ $\mathfrak{P}$-REL. Furthermore, they proved that a class $\mathcal{U} \subseteq \mathcal{R}$ is in $\mathcal{P}$-REL if and only if there is measurable numbering ψ such that $\mathcal{U} \subseteq \mathcal{P}_\psi$. Furthermore, the reliably on $\mathcal{P}$ learnable function classes are characterized as the h-honesty function classes, i.e., $\mathcal{U} \subseteq \mathcal{C}_h$, where the operator $\mathfrak{O}$ is defined as $\mathfrak{O}(f, n) = h(n, f(n))$ (for a more detailed proof see [29, Theorems 12, 27]).

Note that these results also allow for a first answer of how inductive inference strategies discover their errors. This problem was studied in detail in Freivalds, Kinber, and Wiehagen [10]. The results obtained clearly show the importance of characterizations in terms of computable numberings and related techniques.

One such technique is the *amalgamation* technique, which is given implicitly in Barzdin and Podnieks [5] and then formalized in Wiehagen [24]. It was also independently discovered by Case and Smith [8], who gave it its name. Let amal be a recursive function mapping any finite set I of ψ-programs to a φ-program such that for any $x \in \mathbb{N}$, $\varphi_{\mathrm{amal}(I)}(x)$ is defined by running $\varphi_i(x)$ for every $i \in I$ in parallel and taking the first value obtained, if any.

In order to have a further example, let us take a closer look at $\mathcal{R}$-REL. Here we have the additional problem that the strategy S has to diverge on input initially growing finite segments of any function f it cannot learn. We are interested in learning of how this can be achieved. We need the following notation: For every $f \in \mathcal{R}$ and $n \in \mathbb{N}$ we write $f[n]$ to denote the tuple $(f(0), \cdots, f(n))$. Moreover, for any $f \in \mathcal{R}$, $d \in \mathcal{R}$, and $\psi \in \mathcal{P}^2$ we define $H_f = \{i \mid i \in \mathbb{N},\ f[d(i)] \subseteq \psi_i\}$. In [15, Theorem 44] the following was shown:

Let $\mathcal{U} \subseteq \mathcal{R}$ be any function class. Then $\mathcal{U} \in \mathcal{R}$-REL if and only if there is a numbering $\psi \in \mathcal{P}^2$ and a function $d \in \mathcal{R}$ such that

(1) *for every $f \in \mathcal{R}$, if H_f is finite, then H_f contains a ψ-program of the function f, and*
(2) *for every $f \in \mathcal{U}$, the set H_f is finite.*

The proof of this theorem instructively answers where the ability to infer a class $\mathcal{U}$ reliably on $\mathcal{R}$ may be come from. On the one hand, it comes from a well chosen hypothesis space ψ. For any function $f \in \mathcal{U}$ there are only finitely many "candidates" in the set H_f including a ψ-program of f. So, in this case, the amalgamation technique succeeds. On the other hand, the infinity of this set H_f for every function which is not learned, then ensures that the strategy provided in the proof has to diverge. This is also guaranteed by the amalgamation technique, since the sets of "candidates" forms a proper chain of finite sets and so arbitrary large hypotheses are output on every function $f \in \mathcal{R}$ with H_f being infinite.

There are many more characterization theorems in terms of computable numberings including some for LIM and BC (cf., e.g., [29, Section 8] and the references therein), and consistent learning with δ-delay (cf. [1, Section 3.2]).

However, there are also many open problems. For example, Kinber and Zeugmann [16] generalized reliable learning in the limit as defined in this paper to reliable behaviorally correct learning and reliable frequency inference. All these learning types share the useful properties of reliable learning such as closure under recursively enumerable unions and finite invariance (cf. Minicozzi [19]). But we are not aware of any characterization of reliable behaviorally correct learning and reliable frequency inference in terms of computable numberings or in terms of computational complexity.

References

[1] Akama, Y., Zeugmann, T.: Consistent and coherent learning with δ-delay. Information and Computation **206**(11) (2008) 1362–1374

[2] Barzdin, J.M.: Prognostication of automata and functions. In Freiman, C.V., Griffith, J.E., Rosenfeld, J.L., eds.: Information Processing 71, Proceedings of IFIP Congress 71, Volume 1 - Foundations and Systems, Ljubljana, Yugoslavia, August 23-28, 1971, North-Holland (1972) 81–84

[3] Barzdin, J.M.: Две теоремы о предельном синтезе функций. In Barzdin, J.M., ed.: Теория Алгоритмов и Программ. Volume I. Latvian State University (1974) 82 – 88

[4] Barzdin, J.M.: Inductive inference of automata, functions and programs. In: Proc. of the 20-th International Congress of Mathematicians, Vancouver, Canada. (1974) 455–460 (republished in *Amer. Math. Soc. Transl.* (2) 109, 1977, pp.107-112).

[5] Барздинь, Я. М., Подниекс, К. М.: К теорию индуктивного вывода. In: Mathematical Foundations of Computer Science: Proceedings of Symposium and Summer School, Strbské Pleso, High Tatras, Czechoslovakia, September 3-8, 1973., Mathematical Institute of the Slovak Academy of Sciences (1973) 9–15

[6] Blum, L., Blum, M.: Toward a mathematical theory of inductive inference. Information and Control **28**(2) (1975) 125–155

[7] Blum, M.: A machine-independent theory of the complexity of recursive functions. Journal of the ACM **14**(2) (1967) 322–336

[8] Case, J., Smith, C.: Comparison of identification criteria for machine inductive inference. Theoretical Computer Science **25**(2) (1983) 193–220

[9] Feldman, J.: Some decidability results on grammatical inference and complexity. Information and Control **20**(3) (1972) 244–262

[10] Freivalds, R., Kinber, E.B., Wiehagen, R.: How inductive inference strategies discover their errors. Information and Computation **118**(2) (1995) 208–226

[11] Gold, E.M.: Limiting recursion. Journal of Symbolic Logic **30** (1965) 28–48

[12] Gold, E.M.: Language identification in the limit. Information and Control **10**(5) (1967) 447–474

[13] Grabowski, J.: Starke Erkennung. In Linder, R., Thiele, H., eds.: Strukturerkennung diskreter kybernetischer Systeme. Volume 82. Seminarberichte der Sektion Mathematik der Humboldt-Universität zu Berlin (1986) 168–184

[14] Helm, J.: On effectively computable operators. Zeitschrift für mathematische Logik und Grundlagen der Mathematik (ZML) **17** (1971) 231–244

[15] Jain, S., Kinber, E., Wiehagen, R., Zeugmann, T.: On learning of functions refutably. Theoretical Computer Science **298**(1) (2003) 111–143

[16] Kinber, E., Zeugmann, T.: One-sided error probabilistic inductive inference and reliable frequency identification. Inform. Comput. **92**(2) (1991) 253–284

[17] McCreight, E.M., Meyer, A.R.: Classes of computable functions defined by bounds on computation: Preliminary report. In: Proceedings of the 1st Annual ACM Symposium on Theory of Computing, Marina del Rey, California, United States, ACM Press (1969) 79–88

[18] Meyer, A., Fischer, P.C.: On computational speed-up. In: Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory, Papers Presented at the Ninth Annual Symposium Schenectady, New York. October 15-18, 1968, IEEE Computer Society (1968) 351–355

[19] Minicozzi, E.: Some natural properties of strong identification in inductive inference. Theoretical Computer Science **2** (1976) 345–360

[20] Odifreddi, P.: Classical Recursion Theory. North Holland, Amsterdam (1989)

[21] Rogers, H.: Theory of Recursive Functions and Effective Computability. McGraw-Hill (1967) Reprinted, MIT Press 1987.

[22] Stephan, F., Zeugmann, T.: Learning classes of approximations to non-recursive functions. Theoretical Computer Science **288**(2) (2002) 309–341

[23] Wiehagen, R., Liepe, W.: Charakteristische Eigenschaften von erkennbaren Klassen rekursiver Funktionen. Elektronische Informationsverarbeitung und Kybernetik **12**(8/9) (1976) 421–438

[24] Wiehagen, R.: Zur Theorie der Algorithmischen Erkennung. Dissertation B, Humboldt-Universität zu Berlin (1978)

[25] Wiehagen, R., Zeugmann, T.: Ignoring data may be the only way to learn efficiently. J. of Experimental and Theoret. Artif. Intell. **6**(1) (1994) 131–144

[26] Zeugmann, T.: Optimale Erkennung. Diplomarbeit, Sektion Mathematik, Humboldt-Universität zu Berlin (1981)

[27] Zeugmann, T.: A-posteriori characterizations in inductive inference of recursive functions. Elektronische Informationsverarbeitung und Kybernetik **19**(10/11) (1983) 559–594

[28] Zeugmann, T.: On the nonboundability of total effective operators. Zeitschrift für mathematische Logik und Grundlagen der Mathematik (ZML) **30** (1984) 169–172

[29] Zeugmann, T., Zilles, S.: Learning recursive functions: A survey. Theoretical Computer Science **397**(1-3) (2008) 4–56