$\mathbb{TCS}$-$\mathbb{TR}$-$\mathbb{A}$-**08-29**

$\mathbb{TCS}$ Technical Report

Consistent and Coherent Learning with δ-delay

by

YOHJI AKAMA AND THOMAS ZEUGMANN

Division of Computer Science

Report Series A

October 5, 2007

Hokkaido University
Graduate School of
Information Science and Technology

Email: thomas@ist.hokudai.ac.jp

Phone: +81-011-706-7684
Fax: +81-011-706-7684

Consistent and Coherent Learning with δ-delay

Yohji Akama
Mathematical Institute
Tohoku University
Sendai Miyagi Japan, 980-8578, Japan
akama@math.tohoku.ac.jp

Thomas Zeugmann
Division of Computer Science
Hokkaido University
N-14, W-9, Sapporo 060-0814, Japan
thomas@ist.hokudai.ac.jp

October 5, 2007

Abstract

A consistent learner is required to correctly and completely reflect in its actual hypothesis all data received so far. Though this demand sounds quite plausible, it may lead to the unsolvability of the learning problem.

Therefore, in the present paper several variations of consistent learning are introduced and studied. These variations allow a so-called δ–delay relaxing the consistency demand to all but the last δ data.

Additionally, we introduce the notion of *coherent* learning (again with δ–delay) requiring the learner to correctly reflect only the last datum (only the $n-\delta$ th datum) seen.

Our results are manyfold. First, it is shown that all models of coherent learning with δ–delay are exactly as powerful as their corresponding consistent learning models with δ–delay. Second, we provide characterizations for consistent learning with δ–delay in terms of complexity and computable numberings. Finally, we establish strict hierarchies for all consistent learning models with δ–delay in dependence on δ.

1. Introduction

Algorithmic learning has attracted much attention of researchers in various fields of computer science. Inductive inference addresses the question whether or not learning problems may be solved algorithmically at all. There has been huge progress since the pioneering paper of Gold [8] but several questions still deserve attention, in particular from the viewpoint of potential applications.

A main problem of algorithmic learning theory is to synthesize "global descriptions" for the objects to be learned from examples. Thus, one goal is the following. Let f be any computable function from $\mathbb{N}$ into $\mathbb{N}$. Given more and more examples $f(0), f(1), ..., f(n), \dots$ a learning strategy is required to compute a sequence of hypotheses $h_0, h_1, \dots, h_n, \dots$ the limit of which is a correct global description of the function f, i.e., a program that computes f. Since at any stage n of this learning process the strategy knows exclusively the examples $f(0), f(1), ..., f(n)$, one may be tempted to require the strategy to produce only hypotheses h_n such that for any $x \leq n$ the "hypothesis function" g described by h_n is defined and computes the value $f(x)$. Such a hypothesis is called *consistent.* If a hypothesis does not completely and correctly encode all information obtained so far about the unknown object it is called *inconsistent.* A learner exclusively outputting consistent hypotheses is called *consistent.* Requiring a consistent learner looks quite natural at first glance. Why a strategy should output a conjecture that is falsified by the data in hand?

But this is a misleading impression. One of the surprising phenomena discovered in inductive inference is the inconsistency phenomenon (cf., e.g., Barzdin [2], Blum and Blum [4], Wiehagen and Liepe [23], Jantke and Beick [12] as well as Osherson, Stob and Weinstein [17] and the references therein). That is, there are classes of recursive functions that can only be learned by inconsistent strategies.

Naturally, the inconsistency phenomenon has been studied subsequently by many researchers. The reader is encouraged to consult e.g., Jain *et al.* [11], Fulk [7], Freivalds, Kinber and Wiehagen [6] and Wiehagen and Zeugmann [24, 25] for further investigations concerning consistent and inconsistent learning.

In the present paper we introduce and study several variations of consistent learning that have not been considered in the literature. First, we introduce the notion of *coherent* learning. A learner is said to be coherent if it correctly reflects the last datum received (say $f(x_n)$), i.e., if every h_n output satisfies the requirement that the "hypothesis function" g described by h_n is defined on input x_n and $g(x_n) = f(x_n)$. Furthermore, we introduce the notion of δ–delay, where $\delta \in \mathbb{N}$. Then, *coherent* learning with δ–delay means that every h_n output satisfies that $g(x_n \dot{-} \delta)$ is defined and $g(x_n \dot{-} \delta) = f(x_n \dot{-} \delta)$ (cf. Definition 5).

Furthermore, we adopt the notion of δ–delay to the consistent learning types mainly studied so far, i.e., to $\mathcal{CONS}$ (defined by Barzdin [2]), $\mathcal{R}$-$\mathcal{CONS}$ (introduced by Jantke and Beick [12]) and $\mathcal{T}$-$\mathcal{CONS}$ (defined by Wiehagen and Liepe [23]) (cf. Definitions 2, 3 and 4, respectively).

Our results are threefold. First, it is shown that all models of coherent learning with δ–delay are exactly as powerful as their corresponding consistent learning models with δ–delay, see Theorem 1. Second, we provide characterizations for consistent learning with δ–delay in terms of complexity (cf. Theorems 2 and 3) and in terms of computable numberings (cf. Theorems 5, 6, and 7).

Finally, we establish strict hierarchies for all consistent learning models with δ–delay in dependence on δ, see Theorem 8 and Corollary 9.

The paper is structured as follows. Section 2 presents notation and definitions. Then we show the equivalence of coherent and consistent learning for all variants defined (cf. Section 3). The announced characterizations are shown in Section 4.

In Section 5 we prove three new infinite hierarchies for consistent learning with δ–delay. In Section 6 we discuss the results obtained and present open problems. The bibliography is provided in the References.

2. Preliminaries

Unspecified notations follow Rogers [18]. $\mathbb{N} = \{0, 1, 2, \ldots\}$ denotes 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. $\mathcal{R}_{0,1}$ denotes the set of all $0-1$ valued recursive functions (recursive predicates). Sometimes it will be suitable to identify a recursive function with the sequence of its values, e.g., let $\alpha = (a_0, \ldots, a_k) \in \mathbb{N}^*$, $j \in \mathbb{N}$, and $p \in \mathcal{R}_{0,1}$; then we write $\alpha j p$ to denote the function f for which $f(x) = a_x$, if $x \leq k$, $f(k+1) = j$, and $f(x) = p(x-k-2)$, if $x \geq k+2$. Let $g \in \mathcal{P}$, let $\delta \in \mathbb{N}$ and $\alpha = (a_0, \ldots, a_k) \in \mathbb{N}^*$; we write $\alpha \sqsubset_\delta g$ iff α is a δ–prefix of the sequence of values associated with g, i.e., for any x such that $x + \delta \leq k$, $g(x)$ is defined and $g(x) = a_x$. If $\delta = 0$, then we refer to a δ–prefix as to a prefix for short. If $\mathcal{U} \subseteq \mathcal{R}$, then we denote by $[\mathcal{U}]$ the set of all prefixes of functions from $\mathcal{U}$.

Every function $\psi \in \mathcal{P}^2$ is said to be a *numbering*. Furthermore, let $\psi \in \mathcal{P}^2$, then we write ψ_i instead of $\lambda x \psi(i, x)$ and set $\mathcal{P}_\psi = \{\psi_i \mid i \in \mathbb{N}\}$ as well as $\mathcal{R}_\psi = \mathcal{P}_\psi \cap \mathcal{R}$. 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. A numbering $\varphi \in \mathcal{P}^2$ is called a Gödel numbering (cf. Rogers [18]) iff $\mathcal{P}_\varphi = \mathcal{P}$, and for any 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 [5]. That is, $\varphi \in$ *Göd*, $\Phi \in \mathcal{P}^2$ and (1) $dom(\varphi_i) = dom(\Phi_i)$ for all $i \in \mathbb{N}$ and (2) the predicate $\Phi_i(x) = y$ is uniformly recursive for all $i, x, y \in \mathbb{N}$.

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

Moreover, using a fixed encoding $\langle \ldots \rangle$ of $\mathbb{N}^*$ onto $\mathbb{N}$ we write f^n instead of $\langle (f(0), \ldots, f(n)) \rangle$, 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 iff all but finitely many numbers of it are equal to j. Next we define some concepts of learning.

DEFINITION 1 (**Gold** [8]). *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 *iff 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 $\mathcal{U}$ is learnable in the limit with respect to ψ by a strategy S, then we write $\mathcal{U} \in \mathcal{LIM}_\psi(S)$. Let $\mathcal{LIM}_\psi = \{\mathcal{U} \mid \mathcal{U}$ is learnable in the limit w.r.t. $\psi\}$, and let $\mathcal{LIM} = \bigcup_{\psi\in\mathcal{P}^2} \mathcal{LIM}_\psi$.

As far as the semantics of the hypotheses produced 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 $\mathcal{LIM}_\varphi = \mathcal{LIM}$ for any Gödel numbering φ. In the above definition $\mathcal{LIM}$ stands for "limit." Note that in Definition 1 no 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 2. *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 *iff there is a strategy $S \in \mathcal{P}$ such that*

(1) *$\mathcal{U} \in \mathcal{LIM}_\psi(S)$,*

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

$\mathcal{CONS}^\delta_\psi(S)$, $\mathcal{CONS}^\delta_\psi$ and $\mathcal{CONS}^\delta$ are defined analogously to the above.

Note that for $\delta = 0$ we get Barzdin's [2] original definition of $\mathcal{CONS}$. We therefore usually omit the upper index δ if $\delta = 0$. This is also done for all other versions of consistent learning defined below. Moreover, 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 2, then we refer to S as to a δ-delayed consistent strategy. If a strategy does not always works consistently with δ–delay we call it *δ–delayed inconsistent*. The latter two conventions are applied *mutatis mutandis* to Definitions 3 and 4 below.

Next, we modify $\mathcal{CONS}^\delta$ in the same way Jantke and Beick [12] changed $\mathcal{CONS}$, i.e., we add the requirement that the strategy is defined on every input.

DEFINITION 3. *Let $\mathcal{U} \subseteq \mathcal{R}$, let $\psi \in \mathcal{P}^2$ and let $\delta \in \mathbb{N}$. The class $\mathcal{U}$ is called* $\mathcal{R}$–consistently learnable in the limit with δ–delay with respect to ψ *iff there is a strategy $S \in \mathcal{R}$ such that $\mathcal{U} \in \mathcal{CONS}^\delta_\psi(S)$.*

$\mathcal{R}$-$\mathcal{CONS}^\delta_\psi(S)$, $\mathcal{R}$-$\mathcal{CONS}^\delta_\psi$ and $\mathcal{R}$-$\mathcal{CONS}^\delta$ are defined analogously to the above.

Note that in the latter definition consistency with δ–delay is only demanded for inputs that correspond to some function f from the target class $\mathcal{U}$. Therefore, in the following definition we incorporate Wiehagen and Liepe's [23] requirement on a strategy to work consistently on all inputs into our scenario of consistency with δ–delay.

Definition 4. *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 *iff there is a strategy $S \in \mathcal{R}$ such that*

(1) $\mathcal{U} \in \mathcal{CONS}^{\delta}_{\psi}(S)$,

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

$\mathcal{T}\text{-}\mathcal{CONS}^{\delta}_{\psi}(S)$, $\mathcal{T}\text{-}\mathcal{CONS}^{\delta}_{\psi}$ *and* $\mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ *are defined in the same way as above.*

Next, we introduce *coherent* learning (again with δ-delay). While our consistency with δ-delay demand requires a strategy to correctly reflect all but at most the last δ data seen so far, the coherence requirement only demands to correctly reflect the value $f(n \dot{-} \delta)$ on input f^n.

Definition 5. *Let $\mathcal{U} \subseteq \mathcal{R}$, let $\psi \in \mathcal{P}^2$ and let $\delta \in \mathbb{N}$. The class $\mathcal{U}$ is called* coherently learnable in the limit with δ–delay with respect to ψ *iff there is a strategy $S \in \mathcal{P}$ such that*

(1) $\mathcal{U} \in \mathcal{LIM}_{\psi}(S)$,

(2) $\psi_{S(f^n)}(n \dot{-} \delta) = f(n \dot{-} \delta)$ *for all* $f \in \mathcal{U}$ *and all* $n \in \mathbb{N}$ *such that* $n \geq \delta$.

$\mathcal{COH}^{\delta}_{\psi}(S)$, $\mathcal{COH}^{\delta}_{\psi}$ *and* $\mathcal{COH}^{\delta}$ *are defined analogously to the above.*

Now, performing the same modifications to coherent learning with δ–delay as we did in Definitions 3 and 4 to consistent learning with δ–delay results in the learning types $\mathcal{R}\text{-}\mathcal{COH}^{\delta}$ and $\mathcal{T}\text{-}\mathcal{COH}^{\delta}$, respectively. We therefore omit the formal definitions of these learning types here.

Using standard techniques one can show that for all $\delta \in \mathbb{N}$ and all learning types $LT \in \{\mathcal{CONS}^{\delta}, \mathcal{R}\text{-}\mathcal{CONS}^{\delta}, \mathcal{T}\text{-}\mathcal{CONS}^{\delta}, \mathcal{COH}^{\delta}, \mathcal{R}\text{-}\mathcal{COH}^{\delta}, \mathcal{T}\text{-}\mathcal{COH}^{\delta}\}$ we have $LT_{\varphi} = LT$ for every Gödel numbering φ.

3. Coherence and Consistency of Learning Strategies

In this section we study the problem whether or not the relaxation to learn coherently with δ–delay instead of demanding consistency with δ–delay does enhance the learning power of the corresponding learning types introduced in Section 2. The negative answer is provided by the following theorem.

Theorem 1. *Let $\delta \in \mathbb{N}$ be arbitrarily fixed. Then we have*

(1) $\mathcal{CONS}^{\delta} = \mathcal{COH}^{\delta}$,

(2) $\mathcal{R}\text{-}\mathcal{CONS}^{\delta} = \mathcal{R}\text{-}\mathcal{COH}^{\delta}$,

(3) $\mathcal{T}\text{-}\mathcal{CONS}^{\delta} = \mathcal{T}\text{-}\mathcal{COH}^{\delta}$.

Proof. By definition, we obviously have $\mathcal{CONS}^{\delta} \subseteq \mathcal{COH}^{\delta}$, $\mathcal{R}\text{-}\mathcal{CONS}^{\delta} \subseteq \mathcal{R}\text{-}\mathcal{COH}^{\delta}$ and $\mathcal{T}\text{-}\mathcal{CONS}^{\delta} \subseteq \mathcal{T}\text{-}\mathcal{COH}^{\delta}$.

For showing the opposite directions we can essentially use in all three cases the same idea. Let $\delta \in \mathbb{N}$, $\varphi \in Göd$, $\mathcal{U} \subseteq \mathcal{R}$ and any strategy $\hat{S}$ be arbitrarily fixed such that $\mathcal{U} \in LT_{\varphi}(\hat{S})$, where $LT \in \{\mathcal{COH}^{\delta},\ \mathcal{R}\text{-}\mathcal{COH}^{\delta},\ \mathcal{T}\text{-}\mathcal{COH}^{\delta}\}$. Next, we define a strategy S as follows. Let $f \in \mathcal{R}$ and let $n \in \mathbb{N}$. On input f^n do the following.

1. Compute $\hat{S}(f^0), \dots, \hat{S}(f^n)$ and determine the largest number $n^* \leq n$ such that $\hat{S}(f^{n^*-1}) \neq \hat{S}(f^{n^*})$.
2. Output the canonical φ–program i computing the following function g :

   $g(x) = f(x)$ for all $x \leq n^*$, and

   $g(x) = \varphi_{\hat{S}(f^{n^*})}(x)$ for all $x > n^*$.

First, we show that S learns $\mathcal{U}$ consistently with δ–delay.

By construction, we have $\varphi_{S(f^n)}(x) = f(x)$ for all $x \leq n^*$, and thus S is consistent on all data $f(0), \dots, f(n^*)$. If $n - n^* \leq \delta$, we are already done. Finally, if $n - n^* > \delta$, then we exploit the fact that $\hat{S}$ works coherently with δ–delay and that $\hat{S}(f^{n^*+k}) = \hat{S}(f^{n^*})$ for all $k = 1, \dots, n - n^*$. Thus, for all $k \in \{1, \dots, n - n^* - \delta\}$ we get

$$\varphi_{S(f^n)}(n^* + k) \;=\; \varphi_{\hat{S}(f^{n^*})}(n^* + k) = \varphi_{\hat{S}(f^{n^*+\delta+k})}(n^* + k) = f(n^* + k)\ . \tag{1}$$

Since in this case $\hat{S}(f^n)$ is defined for all $f \in \mathcal{U}$ and all $n \in \mathbb{N}$, we can directly conclude that $S(f^n)$ is defined for all $f \in \mathcal{U}$ and all $n \in \mathbb{N}$, too. This proves Assertion (1).

If $\hat{S} \in \mathcal{R}$, then so is S and thus Assertion (2) follows.

Finally, if $\hat{S} \in \mathcal{R}$ and $\hat{S}$ works $\mathcal{T}$–coherently, then we directly get $S \in \mathcal{R}$ and S is $\mathcal{T}$–consistent, since now (1) is true for all $f \in \mathcal{R}$. This completes the proof. ∎

4. Characterizations

Within this section, we characterize consistent learning with δ–delay in terms of complexity and in terms of computable numberings. 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. For having an example, suppose we want to learn a class $\mathcal{U}$ with respect to any fixed Gödel numbering φ . Then, a strategy may try to find a program i such that $\varphi_i^n = f^n$. Though this search will succeed, the strategy may face serious difficulties to converge. These difficulties are caused by the undecidability of the halting problem. When, on input f^n a strategy S has found a program i as described above and then sees $f(n+1)$ it may try to compute $\varphi_i(n+1)$ and, in parallel to find again an index, say j, such that $\varphi_j^{n+1} = f^{n+1}$. If it finds j and the computation of $\varphi_i(n+1)$ did not stop yet, then the strategy is in trouble. In order to converge, it may further try to compute $\varphi_i(n+1)$ thus risking that this try may fail to succeed or it may output j instead. But of course, switching to a new hypothesis can only be done finitely often, since otherwise S will not converge. Thus, additional information concerning the computational complexity of the functions to be learned can only help.

Alternatively, particularly designed numberings possessing properties that facilitate learning may also help to overcome the difficulties described above.

We start with the complexity theoretic characterizations.

4.1. *Characterizations in Terms of Complexity*

First, we recall the definitions of recursive and general recursive operator. Let $(F_x)_{x\in\mathbb{N}}$ be the canonical enumeration of all finite functions.

Definition 6 (Rogers [18]). *A mapping $\mathfrak{O}\colon \mathfrak{P} \mapsto \mathfrak{P}$ from partial functions to partial functions is called a* partial recursive operator *iff there is a recursively enumerable set $W \subset \mathbb{N}^3$ such that for any $y, z \in \mathbb{N}$ it holds that $\mathfrak{O}(f)(y) = z$ iff 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 called a general recursive operator *iff $\mathfrak{T} \subseteq dom(\mathfrak{O})$, 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 *iff 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 $\mathcal{R} \subseteq dom(\mathfrak{O})$, and $\varphi_i \in \mathcal{R}$ implies $\mathfrak{O}(\varphi_i) \in \mathcal{R}$.*

For more information about general recursive operators and effective operators the reader is referred to [10, 15, 26]. 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. Any computable operator can be realized by a 3-tape Turing machine T which works as follows: If for an arbitrary function $f \in dom(\mathfrak{O})$, all pairs $(x, f(x))$, $x \in dom(f)$ 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 general recursive or total effective operator. Then, for $f \in dom(\mathfrak{O})$, $m \in \mathbb{N}$ we set: $\Delta\mathfrak{O}(f, m) =$ "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 $\Delta\mathfrak{O}(f, m) = \infty$."

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

Note that in the following we use mainly ideas and techniques from Wiehagen [22] who proved theses theorems for the case $\delta = 0$. Variants of these characterizations for $\delta = 0$ can also be found in Wiehagen and Liepe [23] as well as in Odifreddi [16].

Furthermore, in the following we always assume that learning is done with respect to any fixed $\varphi \in G\ddot{o}d$.

As in Blum and Blum [4] we define operator complexity classes as follows. Let $\mathfrak{O}$ be any computable operator; then we set

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

First, we characterize $\mathcal{T}\text{-}\mathcal{CONS}^\delta$.

THEOREM 2. *Let* $\mathcal{U} \subseteq \mathcal{R}$ *and let* $\delta \in \mathbb{N}$*; then we have:* $\mathcal{U} \in \mathcal{T}\text{-}\mathcal{CONS}^\delta$ *if and only if there exists a general recursive operator* $\mathfrak{O} \in \Omega_{id+\delta}$ *such that* $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ *and* $\mathcal{U} \subseteq \mathcal{R}_{\mathfrak{O}}$.

Proof. Necessity. Let $\mathcal{U} \in \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$, $S \in \mathcal{R}$. Then for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$ we define $\mathfrak{O}(f,n) = \Phi_{S(f^{n+\delta})}(n)$.

Since $\varphi_{S(f^{n+\delta})}(n)$ is defined for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$, by Condition (2) of Definition 4, we directly get from Condition (1) of the definition of a complexity measure that $\Phi_{S(f^{n+\delta})}(n)$ is defined for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$, too. Moreover, for every $t \in \mathfrak{T}$ and $n \in \mathbb{N}$ there is an $f \in \mathcal{R}$ such that $t^n = f^n$. Hence, we have $\mathfrak{O}(\mathfrak{T}) \subseteq \mathcal{R} \subseteq \mathfrak{T}$. Moreover, in order to compute $\mathfrak{O}(f,n)$ the operator $\mathfrak{O}$ reads only the values $f(0), \dots, f(n+\delta)$. Thus, we have $\mathfrak{O} \in \Omega_{id+\delta}$.

Now, let $f \in \mathcal{U}$. Then the sequence $(S(f^n))_{n\in\mathbb{N}}$ converges to a correct φ–program i for f. Consequently, $\mathfrak{O}(f,n) = \Phi_i(n)$ for almost all $n \in \mathbb{N}$. Therefore, we conclude $\mathcal{U} \subseteq \mathcal{R}_{\mathfrak{O}}$.

Sufficiency. Let $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{R}_{\mathfrak{O}}$. We have to define a strategy $S \in \mathcal{R}$ such that $\mathcal{U} \in \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$. By the definition of $\mathcal{R}_{\mathfrak{O}}$ we know that for every $f \in \mathcal{U}$ there are i and k such that $\varphi_i = f$ and $\Phi_i(x) \le \max\{k,\ \mathfrak{O}(f,x)\}$ for all x. Thus, the desired strategy S searches for the first pair (i,k) in the canonical enumeration c_2 of $\mathbb{N} \times \mathbb{N}$ and converges to i provided it has been found. Until this pair (i,k) is found, the strategy S outputs auxiliary consistent hypotheses. For doing this, we choose $g \in \mathcal{R}$ such that $\varphi_{g(\langle\alpha\rangle)}(x) = y_x$ for every tuple $\alpha \in \mathbb{N}^*$, $\alpha = (y_0, \dots, y_n)$ and all $x \le n$.

$S(f^n) =$ "Compute $\mathfrak{O}(f,x)$ for all $x \le n \dot{-} \delta$. Search for the least $z \le n$ such that for $c_2(z) = (i,k)$ the conditions

(A) $\Phi_i(x) \le \max\{k,\ \mathfrak{O}(f,x)\}$ for all $x \le n \dot{-} \delta$, and

(B) $\varphi_i(x) = f(x)$ for all $x \leq n \dot{-} \delta$

are fulfilled. If such a z is found, set $S(f^n) = i$.

Otherwise, set $S(f^n) = g(f^n)$."

Since $\mathfrak{O} \in \Omega_{id+\delta}$, the strategy can compute $\mathfrak{O}(f,x)$ for all $x \leq n \dot{-} \delta$ and since $c_2 \in \mathcal{R}$, it also can perform the desired search effectively. By Condition (2) of the definition of a complexity measure, the test in (A) can be performed effectively, too. If this test has succeeded, then Test (B) can also be effectively executed by Condition (1) of the definition of a complexity measure. Thus, we get $S \in \mathcal{R}$. Finally, by construction S is always consistent with δ-delay, and if $f \in \mathcal{U}$ it converges to a correct φ–program for f. ∎

THEOREM 3. *Let $\mathcal{U} \subseteq \mathcal{R}$ and let $\delta \in \mathbb{N}$; then we have: $\mathcal{U} \in \mathcal{CONS}^\delta$ if and only if there exists a partial recursive operator $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{R}_{\mathfrak{O}}$.*

Proof. The necessity is proved *mutatis mutandis* as in the proof of Theorem 2 with the only modification that $\mathfrak{O}(f,x)$ is now defined for all $f \in \mathcal{U}$ instead of $f \in \mathcal{R}$. This directly yields $\mathfrak{O} \in \Omega_{id+\delta}$, $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{R}_{\mathfrak{O}}$.

The only modification for the sufficiency part is to leave $S(f^n)$ undefined if $\mathfrak{O}(f,x)$ is not defined for $f \notin \mathcal{U}$. We omit the details. ∎

Unfortunately, we do not know how to characterize $\mathcal{R}$-$\mathcal{CONS}^\delta$ in terms of complexity.

We finish this subsection by using Theorem 2 to show that $\mathcal{T}$-$\mathcal{CONS}^\delta$ is closed under enumerable unions. Looking at applications this is a favorable property, since it provides a tool to build more powerful learners from simpler ones.

THEOREM 4. *Let $\delta \in \mathbb{N}$ and let $(S_i)_{i\in\mathbb{N}}$ be a recursive enumeration of strategies working $\mathcal{T}$-consistently with δ-delay. Then there exists a strategy $S \in \mathcal{R}$ such that $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$.*

Proof. The proof of the necessity of Theorem 2 shows that the construction of the operator $\mathfrak{O}$ is effective provided a program for the strategy is given. Thus, we effectively obtain a recursive enumeration $(\mathfrak{O}_i)_{i\in\mathbb{N}}$ of operators $\mathfrak{O}_i \in \Omega_{id+\delta}$ such that $\mathfrak{O}_i(\mathcal{R}) \subseteq \mathcal{R}$ and $\mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{R}_{\mathfrak{O}_i}$.

Now, we define an operator $\mathfrak{O}$ as follows. Let $f \in \mathcal{R}$ and $x \in \mathbb{N}$. We set $\mathfrak{O}(f,x) = \max\{\mathfrak{O}_i(f,x) \mid i \leq x\}$.

Thus, we directly get $\mathfrak{O} \in \Omega_{id+\delta}$, $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ and $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{R}_{\mathfrak{O}}$. Hence, by Theorem 2 we can conclude $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$. ∎

On the other hand, $\mathcal{CONS}^\delta$ and $\mathcal{R}$-$\mathcal{CONS}^\delta$ are not even closed under finite union. This is a direct consequence of a more general result Barzdin [1] showed, i.e., there are classes $\mathcal{U} = \{f \mid f \in \mathcal{R}, \varphi_{f(0)} = f\}$ and $\mathcal{V} = \{\alpha 0^\infty \mid \alpha \in \mathbb{N}^*\}$ such that

$\mathcal{U}\cup\mathcal{V}\notin\mathcal{LIM}$. Now, it is easy to verify $\mathcal{U},\ \mathcal{V}\in\mathcal{R}\text{-}\mathcal{CONS}^{\delta}$ and thus $\mathcal{U},\ \mathcal{V}\in\mathcal{CONS}^{\delta}$ for every $\delta\in\mathbb{N}$, but since $\mathcal{U}\cup\mathcal{V}\notin\mathcal{LIM}$ we clearly have $\mathcal{U}\cup\mathcal{V}\notin\mathcal{R}\text{-}\mathcal{CONS}^{\delta}$ and $\mathcal{U}\cup\mathcal{V}\notin\mathcal{CONS}^{\delta}$ for all $\delta\in\mathbb{N}$.

Next, we continue with characterizations in terms of computable numberings.

4.2. *Characterizations in Terms of Computable Numberings*

As we shall see below, the differences and similarities between the different versions of consistent learning with δ–delay can be completely expressed by different versions of decidable consistency conditions. Therefore, adapting ideas from Wiehagen and Zeugmann [25], next we define these decidable consistency conditions.

DEFINITION 7. *Let $\mathcal{U}\subseteq\mathcal{R}$ and let $\psi\in\mathcal{P}^2$ be any numbering. We say that*

(1) δ-delayed $\mathcal{U}$–consistency with respect to ψ is decidable *iff there is a predicate $cons\in\mathcal{P}^2$ such that for every $\alpha\in[\mathcal{U}]$ and all $i\in\mathbb{N}$, $cons(\alpha,i)$ is defined and $cons(\alpha,i)=1$ if and only if $\alpha\sqsubset_\delta\psi_i$.*

(2) δ-delayed $\mathcal{U}$–consistency with respect to ψ is $\mathcal{R}$–decidable *iff there is a predicate $cons\in\mathcal{R}^2$ such that for every $\alpha\in[\mathcal{U}]$ and all $i\in\mathbb{N}$, $cons(\alpha,i)=1$ if and only if $\alpha\sqsubset_\delta\psi_i$.*

(3) δ-delayed consistency with respect to ψ is decidable *iff there is a predicate $cons\in\mathcal{R}^2$ such that for every $\alpha\in\mathbb{N}^*$ and all $i\in\mathbb{N}$, $cons(\alpha,i)=1$ if and only if $\alpha\sqsubset_\delta\psi_i$.*

Note that the proofs below use ideas from Wiehagen [21] and from Wiehagen and Zeugmann [25].

THEOREM 5. *Let $\mathcal{U}\subseteq\mathcal{R}$, then we have: $\mathcal{U}\in\mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ if and only if there exists a numbering $\psi\in\mathcal{P}^2$ such that*

(1) *$\mathcal{U}\subseteq\mathcal{P}_\psi$,*

(2) *δ-delayed consistency with respect to ψ is decidable.*

Proof. Necessity. Let $\mathcal{U}\in\mathcal{T}\text{-}\mathcal{CONS}^{\delta}_{\varphi}(S)$ where $\varphi\in\mathcal{P}^2$ is any Gödel numbering and S is a δ–delayed $\mathcal{T}$–consistent strategy. Let

$$M=\{(z,n)|\ \ z,n\in\mathbb{N},\ \varphi_z(x)\text{ is defined for all }x\le n,\ S(\varphi_z^n)=z\}$$

be recursively enumerated by a function e. Then define a numbering ψ as follows. Let $i,x\in\mathbb{N},\ e(i)=(z,n)$.

$$\psi_i(x)=\begin{cases}\varphi_z(x), & \text{if } x\le n\\ \varphi_z(x), & \text{if } x>n \text{ and, for any } y\in\mathbb{N} \text{ such that } n<y\le x,\\ & \varphi_z(y)\text{ is defined and } S(\varphi_z^y)=z\\ \text{undefined}, & \text{otherwise.}\end{cases}$$

For showing (1) let $f \in \mathcal{U}$ and $n, z \in \mathbb{N}$ be such that for all $m \geq n$, $S(f^m) = z$. Clearly, $\varphi_z = f$. Furthermore, $(z, n) \in M$. Let $i \in \mathbb{N}$ be such that $e(i) = (z, n)$. Then, by the definition of ψ, we have $\psi_i = \varphi_z = f$. Hence $\mathcal{U} \subseteq \mathcal{P}_\psi$.

In order to prove (2) we define $cons \in \mathcal{R}^2$ such that for all $\alpha \in \mathbb{N}^*$, $i \in \mathbb{N}$, $cons(\alpha, i) = 1$ iff $\alpha \sqsubset_\delta \psi_i$. Let $\alpha = (\alpha_0, \ldots, \alpha_x) \in \mathbb{N}^*$ and $i \in \mathbb{N}$. Let $e(i) = (z, n)$. Then define

$$cons(\alpha, i) = \begin{cases} 1, & \text{if } x < \delta \\ 1, & \text{if } \delta \leq x \leq n \text{ and, for every } y \text{ such that } y + \delta \leq x,\ \alpha_y = \psi_i(y) \\ 1, & \text{if } \delta \leq x,\ x > n \text{ and, for every } y \leq \min\{n,\ x - \delta\},\ \alpha_y = \psi_i(y), \\ & \text{and } S(\alpha_0, \ldots, \alpha_y) = z \text{ for every } y \in \mathbb{N} \text{ such that } n < y + \delta \leq x \\ 0, & \text{otherwise.} \end{cases}$$

Since $e(i) = (z, n) \in M$, by construction we know that $\varphi_z(m)$ is defined for all $m \leq n$ and $(\varphi_z^n) = z$. Thus, we have $\psi_i(m) = \varphi_z(m)$ for all $m \leq n$. Consequently, if $x \leq n$, then for all $y \leq x$ it can be effectively tested whether or not $\alpha_y = \psi_i(y)$. Furthermore, $S \in \mathcal{R}$ implies that $S(\alpha_0, \ldots, \alpha_y)$ can be computed for every $y \in \mathbb{N}$ such that $n < y + \delta \leq x$. Thus, if $x > n$ the condition $S(\alpha_0, \ldots, \alpha_y) = z$ can be effectively checked for every $y \in \mathbb{N}$ such that $n < y + \delta \leq x$. Consequently, $cons \in \mathcal{R}^2$.

It remains to show that for every $\alpha \in \mathbb{N}^*$, $i \in \mathbb{N}$, we have $cons(\alpha, i) = 1$ if and only if $\alpha \sqsubset_\delta \psi_i$.

First, assume $cons(\alpha, i) = 1$. As long as $x < \delta$, we have $\alpha \sqsubset_\delta \psi_i$ for every $\alpha = (\alpha_0, \ldots, \alpha_x)$. If $\delta \leq x \leq n$ then we have $\alpha \sqsubset_\delta \psi_i$, since this has been checked in the second case of the definition of *cons*.

Now, let $\delta \leq x$ and $x > n$. Then for every $y \leq \min\{n,\ x - \delta\}$ it has been checked that $\alpha_y = \psi_i(y)$. Thus, as long as $x - \delta \leq n$ we are done. If $x - \delta > n$ then we furthermore know that

$$S(\alpha_0, \ldots, \alpha_y) = z \text{ for every } y \in \mathbb{N} \text{ such that } n < y + \delta \leq x\ . \tag{2}$$

Since S is a δ–delayed $\mathcal{T}$–consistent strategy, we have

$$\varphi_z(m) = \alpha_m \text{ for all } m \text{ such that } m + \delta \leq x\ . \tag{3}$$

By construction $\psi_i(x) = \varphi_z(x)$ for all $x \leq n$. By (3) we can conclude that $\varphi_z(y)$ is defined for all y such that $n < y + \delta \leq x$. Finally, (2) implies that $S(\varphi_z^y) = z$ for all y such that $n < y + \delta \leq x$. Thus, $\psi_i(m) = \varphi_z(m)$ for all m such that $m + \delta \leq x$. Therefore, by (3) we get $\alpha \sqsubset_\delta \psi_i$.

Next, assume $\alpha \sqsubset_\delta \psi_i$. We have to show that $cons(\alpha, i) = 1$. This is obvious for all $x < \delta$. If $\delta \leq x$ and $x + \delta \leq n$ then the definition of *cons* directly yields $cons(\alpha, i) = 1$. Finally, if $n < x + \delta$ then, by construction, we know that

$\psi_i(m) = \varphi_z(m)$ for all m such that $m+\delta \leq x$, since otherwise $\psi_i(m)$ is not defined for $n < m+\delta \leq x$. Additionally, $S(\varphi_z^y) = z$ for all y such that $n < y+\delta \leq x$. Thus, $S(\alpha_0, \ldots, \alpha_y) = z$ for all y such $n < y + \delta \leq x$, and consequently $cons(\alpha, i) = 1$. This proves the necessity.

Sufficiency. Let $\psi \in \mathcal{P}^2$ be any numbering. Let $cons \in \mathcal{R}^2$ be such that for all $\alpha \in \mathbb{N}^*$, $i \in \mathbb{N}$, $cons(\alpha, i) = 1$ iff $\alpha \sqsubset_\delta \psi_i$. Let $\mathcal{U} \subset \mathcal{P}_\psi$. In order to consistently learn any function $f \in \mathcal{U}$ it suffices to define $S(f^n) = \min\{i \mid \ cons(f^n, i) = 1\}$. However, S would be undefined if, for $f \notin \mathcal{U}$, $n \in \mathbb{N}$, there is *no* $i \in \mathbb{N}$ such that $f^n \sqsubset_\delta \psi_i$. This difficulty is circumvented by the following definition. Let $\varphi \in \textit{Göd}$. Let $aux \in \mathcal{R}$ be such that for any $\alpha \in \mathbb{N}^*$, $\varphi_{aux(\alpha)} = \alpha 0^\infty$. Finally, let $c \in \mathcal{R}$ be such that for all $i \in \mathbb{N}$, $\psi_i = \varphi_{c(i)}$. Then, for any $f \in \mathcal{R}$, $n \in \mathbb{N}$, define a strategy S as follows.

$$S(f^n) = \begin{cases} c(j), & \text{if } I = \{i \mid \ i \leq n,\ cons(f^n, i) = 1\} \neq \emptyset \text{ and } j = \min\ I \\ aux(f^n), & I = \emptyset\ . \end{cases}$$

Clearly, $S \in \mathcal{R}$ and S outputs only δ–delayed consistent hypotheses. Now let $f \in \mathcal{U}$. Then, obviously, $(S(f^n))_{n\in\mathbb{N}}$ converges to $c(min\{i \mid \ \psi_i = f\})$. Hence, S witnesses $U \in \mathcal{T}\text{-}\mathcal{CONS}_\varphi^\delta$. ∎

Next, we characterize $\mathcal{R}\text{-}\mathcal{CONS}^\delta$ in terms of computable numberings.

THEOREM 6. *Let $\mathcal{U} \subseteq \mathcal{R}$, then we have: $\mathcal{U} \in \mathcal{R}\text{-}\mathcal{CONS}^\delta$ if and only if there exists a numbering $\psi \in \mathcal{P}^2$ such that*

(1) $\mathcal{U} \subseteq \mathcal{P}_\psi$,

(2) *δ-delayed $\mathcal{U}$–consistency with respect to ψ is $\mathcal{R}$–decidable.*

Proof. The proof is similar to that of Theorem 5. The only difference affects the predicate $cons$. Though its formal definition remains unchanged, the properties of $cons$ change. That is, now we get $cons \in \mathcal{R}^2$ such that for all $\alpha \in [\mathcal{U}]$, $i \in \mathbb{N}$, $cons(\alpha, i) = 1$ iff $\alpha \sqsubset_\delta \psi_i$. ∎

Finally, we characterize $\mathcal{CONS}^\delta$. Again the proof is analogous to the one of Theorem 5 and therefore omitted.

THEOREM 7. *Let $\mathcal{U} \subseteq \mathcal{R}$, then we have: $\mathcal{U} \in \mathcal{CONS}^\delta$ if and only if there exists a numbering $\psi \in \mathcal{P}^2$ such that*

(1) $\mathcal{U} \subseteq \mathcal{P}_\psi$,

(2) *δ-delayed $\mathcal{U}$–consistency with respect to ψ is decidable.*

5. Hierarchy Results

Within this section we study the problem whether or not the introduction of δ–delay to consistent learning yields an advantage with respect to the learning power of the defined learning types.

For answering this problem it is advantageous to recall the definition of reliable learning introduced by Blum and Blum [4] and Minicozzi [14]. Intuitively, a learner M is reliable provided it converges if and only if it learns.

Definition 8 **(Blum and Blum [4], Minicozzi [14]).** *Let* $\mathcal{U} \subseteq \mathcal{R}$ *, let* $\mathcal{M} \subseteq \mathfrak{T}$ *and let* $\varphi \in$ *Göd ; then* $\mathcal{U}$ *is said to be* reliably learnable on $\mathcal{M}$ *if there is a strategy* $S \in \mathcal{R}$ *such that*

(1) $\mathcal{U} \in \mathcal{LIM}_\varphi(S)$ *, and*

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

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

In particular, we shall consider the cases where $\mathcal{M} = \mathfrak{T}$ and $\mathcal{M} = \mathcal{R}$, i.e., reliable learnability on the set of all total functions and all recursive functions, respectively.

Our first theorem shows that incrementing δ yields more learning power for δ–delayed $\mathcal{T}$-consistent strategies in general. On the other hand, when restricted to learning predicates, the learning capabilities of $\mathcal{T}$-$\mathcal{CONS}^\delta$ are *not* enlarged. Only classes of predicates contained in $\mathcal{NUM}$ can be identified by δ–delayed $\mathcal{T}$-consistent strategies.

Theorem 8. *The following statements hold for all* $\delta \in \mathbb{N}$ *:*

(1) $\mathcal{T}$-$\mathcal{CONS}^\delta \subset \mathcal{T}$-$\mathcal{CONS}^{\delta+1} \subset \mathfrak{T}$-$\mathcal{REL}$,

(2) $\mathcal{NUM} \cap \wp(\mathcal{R}_{0,1}) = \mathcal{T}$-$\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{0,1}) = \mathcal{T}$-$\mathcal{CONS}^{\delta+1} \cap \wp(\mathcal{R}_{0,1}) = \mathfrak{T}$-$\mathcal{REL} \cap \wp(\mathcal{R}_{0,1})$,

(3) $\mathcal{T}$-$\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{0,1}) \subset \mathcal{R}$-$\mathcal{REL} \cap \wp(\mathcal{R}_{0,1})$.

Proof. We first prove Assertion (1). Let $\delta \in \mathbb{N}$ be arbitrarily fixed. Then by Definition 4 we obviously have $\mathcal{T}$-$\mathcal{CONS}^\delta \subseteq \mathcal{T}$-$\mathcal{CONS}^{\delta+1}$. For showing $\mathcal{T}$-$\mathcal{CONS}^{\delta+1} \setminus \mathcal{T}$-$\mathcal{CONS}^\delta \neq \emptyset$ we use the following class. Let (φ, Φ) be any complexity measure; we set

$$\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} = \{f \mid f \in \mathcal{R},\ \varphi_{f(0)} = f,\ \forall x[\Phi_{f(0)}(x) \leq f(x+\delta+1)]\} .$$

Claim 1. $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \in \mathcal{T}$-$\mathcal{CONS}^{\delta+1}$.

The desired strategy S is defined as follows. Let $g \in \mathcal{R}$ be the function defined in the sufficiency proof of Theorem 2. For all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$ we set

$$S(f^n) = \begin{cases} f(0), & \text{if } n \le \delta \text{ or } n > \delta \text{ and } \Phi_{f(0)}(y) \le f(y+\delta+1) \\ & \text{and } \varphi_{f(0)}(y) = f(y) \text{ for all } y \le n \dot{-} \delta \dot{-} 1 \\ g(f^n), & \text{otherwise.} \end{cases}$$

Now, by construction one easily verifies $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1}(S)$. This proves Claim 1.

Claim 2. $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \notin \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$.

Suppose the converse. Then there must be a strategy $S \in \mathcal{R}$ such that $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \in \mathcal{T}\text{-}\mathcal{CONS}_{\varphi}^{\delta}(S)$. We continue by constructing a function φ_{i^*} belonging to $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$ but on which S fails.

Furthermore, let $r \in \mathcal{R}$ be such that $\Phi_i = \varphi_{r(i)}$ for all $i \in \mathbb{N}$ and r is strongly monotone growing, i.e., $r(i) < r(i+1)$ for all $i \in \mathbb{N}$. Then $range(r)$ is recursive (cf. Rogers [18]). Choose $s \in \mathcal{R}$ such that for all $j \in \mathbb{N}$ we have for all $x \le \delta$

$$\varphi_{s(j)}(x) = \begin{cases} i, & \text{if there is an } i \text{ with } r(i) = j\ , \\ 0, & \text{otherwise.} \end{cases}$$

For the further definition of $\varphi_{s(j)}$ we also use in every step $\delta+1$ arguments. For $x = 0,\ \delta+1,\ 2\delta+2,\ 3\delta+3, \ldots$ we set

$$\begin{aligned} \varphi_{s(j)}(x+\delta+1) &= \varphi_j(x)+1 \\ &\vdots \\ \varphi_{s(j)}(x+2\delta+1) &= \varphi_j(x+\delta)+1 \end{aligned}$$

provided $\varphi_j(x)$, $\varphi_j(x+1), \ldots, \varphi_j(x+\delta)$ are all defined, $\varphi_{s(j)}^{x+\delta}$ is defined and

$$S\left(\varphi_{s(j)}^{x+\delta}\right) = S\left(\langle(\varphi_{s(j)}(0), \ldots, \varphi_{s(j)}(x+\delta), \varphi_j(x), \ldots, \varphi_j(x+\delta))\rangle\right)$$

and we set

$$\begin{aligned} \varphi_{s(j)}(x+\delta+1) &= \varphi_j(x) \\ &\vdots \\ \varphi_{s(j)}(x+2\delta+1) &= \varphi_j(x+\delta) \end{aligned}$$

provided $\varphi_j(x)$, $\varphi_j(x+1), \ldots, \varphi_j(x+\delta)$ are all defined, $\varphi_{s(j)}^{x+\delta}$ is defined and

$$S\left(\varphi_{s(j)}^{x+\delta}\right) \neq S\left(\langle(\varphi_{s(j)}(0), \ldots, \varphi_{s(j)}(x+\delta), \varphi_j(x), \ldots, \varphi_j(x+\delta))\rangle\right) .$$

Otherwise, $\varphi_{s(j)}(x+\delta+1), \ldots, \varphi_{s(j)}(x+2\delta+1)$ remain undefined.

By the Fixed Point Theorem (cf. Rogers [18]) there exists a number i^* such that $\varphi_{s(r(i^*))} = \varphi_{i^*}$.

Next, we show that $\varphi_{i^*} \in \mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$. This is done inductively. For the induction base, by construction we have $\varphi_{i^*}(0) = \cdots = \varphi_{i^*}(\delta) = i^*$. Hence, $\Phi_{i^*}(0), \ldots, \Phi_{i^*}(\delta)$ are all defined, too. Therefore, we know that $\varphi^{\delta}_{s(r(i^*))}$ is defined and so either $\varphi_{s(r(i^*))}(\delta+1) = \Phi_{i^*}(0) + 1$, ..., $\varphi_{s(r(i^*))}(2\delta+1) = \Phi_{i^*}(\delta) + 1$ provided

$$S\left(\varphi^{\delta}_{s(r(i^*))}\right) = S\left(\langle(\varphi_{s((r(i^*))}(0), \ldots, \varphi_{s((r(i^*))}(\delta), \Phi_{i^*}(0), \ldots, \Phi_{i^*}(\delta))\rangle\right)$$

or $\varphi_{s(r(i^*))}(\delta+1) = \Phi_{i^*}(0)$, ..., $\varphi_{s(r(i^*))}(2\delta+1) = \Phi_{i^*}(\delta)$ if

$$S\left(\varphi^{\delta}_{s(r(i^*))}\right) \neq S\left(\langle(\varphi_{s((r(i^*))}(0), \ldots, \varphi_{s((r(i^*))}(\delta), \Phi_{i^*}(0), \ldots, \Phi_{i^*}(\delta))\rangle\right) .$$

One of these cases must happen, since otherwise S would not be $\mathcal{T}$-consistent with δ–delay.

Hence, $\Phi_{i^*}(0) \leq \varphi_{i^*}(\delta+1), \ldots, \Phi_{i^*}(\delta) \leq \varphi_{i^*}(2\delta+1)$, since $\varphi_{s(r(i^*))} = \varphi_{i^*}$. So we know that $\varphi_{i^*}(\delta+1), \ldots, \varphi_{i^*}(2\delta+1)$ as well as $\Phi_{i^*}(\delta+1), \ldots, \Phi_{i^*}(2\delta+1)$ are all defined. This completes the induction base.

Consequently, we have the induction hypothesis that for some $x = 0, \delta+1,\ 2\delta+2,\ 3\delta+3, \ldots$ the values $\varphi_{i^*}(z)$ are defined and $\Phi_{i^*}(z) \leq \varphi_{i^*}(z+\delta+1)$ for all $z \leq x+\delta$. This of course implies $\varphi^{x+\delta}_{s(r(i^*))}$ is defined, too. The induction step is done from x to $x+\delta+1$. First, we either have $\varphi_{s(r(i^*))}(x+\delta+1) = \Phi_{i^*}(x) + 1$, ..., $\varphi_{s(r(i^*))}(x+2\delta+1) = \Phi_{i^*}(x+\delta) + 1$ provided

$$S\left(\varphi^{x+\delta}_{s(r(i^*))}\right) = S\left(\langle(\varphi_{s(r(i^*))}(0), \ldots, \varphi_{s(r(i^*))}(x+\delta), \Phi_{i^*}(x), \ldots, \Phi_{i^*}(x+\delta))\rangle\right)$$

or $\varphi_{s(r(i^*))}(x+\delta+1) = \Phi_{i^*}(x)$, ..., $\varphi_{s(r(i^*))}(x+2\delta+1) = \Phi_{i^*}(x+\delta)$ if

$$S\left(\varphi^{x+\delta}_{s(r(i^*))}\right) \neq S\left(\langle(\varphi_{s(r(i^*))}(0), \ldots, \varphi_{s(r(i^*))}(x+\delta), \Phi_{i^*}(x), \ldots, \Phi_{i^*}(x+\delta))\rangle\right) .$$

Again, one of these cases must happen, since otherwise S would not be $\mathcal{T}$-consistent with δ–delay.

Therefore, $\varphi_{i^*}(x+\delta+1), \ldots, \varphi_{i^*}(x+2\delta+1)$ are all defined and, additionally, $\Phi_{i^*}(x) \leq \varphi_{i^*}(x+\delta+1), \ldots, \Phi_{i^*}(x+\delta) \leq \varphi_{i^*}(x+2\delta+1)$.

Now, we also know that $\Phi_{i^*}(x+\delta+1)$, ..., $\Phi_{i^*}(x+2\delta+1)$ are all defined. Therefore, we have shown that $\varphi_{i^*} \in \mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$. Finally, by construction we directly obtain that S performs infinitely mind changes when successively fed φ_{i^*}, a contradiction to $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta}(S)$. This proves Claim 2.

Taking into account that a strategy working $\mathcal{T}$-consistently with δ–delay converges when successively fed any function f if and only if it learns f, we directly get $\mathcal{T}\text{-}\mathcal{CONS}^{\delta} \subseteq \mathfrak{T}\text{-}\mathcal{REL}$ for every $\delta \in \mathbb{N}$. Furthermore, as shown in Minicozzi [14], $\mathfrak{T}\text{-}\mathcal{REL}$ is closed under recursively enumerable union. Therefore, setting $\mathcal{U} = \bigcup_{\delta\in\mathbb{N}} \mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$ we can conclude $\mathcal{U} \in \mathfrak{T}\text{-}\mathcal{REL}$. But obviously $\mathcal{U} \notin \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ for any δ. This proves Assertion (1).

For showing Assertion (2), we prove that for every operator $\mathfrak{O} \in \Omega_{id+\delta}$ there is a monotone operator $\hat{\mathfrak{O}} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(f,x) \leq \hat{\mathfrak{O}}(f,x)$ for all $f \in \mathcal{R}$ and all $x \in \mathbb{N}$. Here, we call an operator monotone if, for all $f,\, g \in \mathcal{R}$ and $\forall x[f(x) \leq g(x)]$ implies $\forall x[\mathfrak{O}(f,x) \leq \mathfrak{O}(g,x)]$. This can be seen as follows. Let $\alpha = (\alpha_0, \ldots, \alpha_m)$ and $\beta = (\beta_0, \ldots, \beta_m)$ be any tuples of length $m+1$ from $\mathbb{N}^*$. We write $\alpha \leq \beta$ if $\alpha_i \leq \beta_i$ for all $i = 1, \ldots, m$. Now, let $\mathfrak{O} \in \Omega_{id+\delta}$. We define

$$\hat{\mathfrak{O}}(\beta, x) = \max\{\mathfrak{O}(\alpha, x) \big|\ |\alpha| = |\beta| = x + \delta,\ \alpha \leq \beta\}\ .$$

Since the operator $\mathfrak{O}$ for computing $\mathfrak{O}(\alpha, x)$ just needs $id + \delta$ values, $\hat{\mathfrak{O}}$ is properly defined, and, by its definition, $\hat{\mathfrak{O}} \in \Omega_{id+\delta}$, $\hat{\mathfrak{O}}$ is monotone, and $\mathfrak{O}(f,x) \leq \hat{\mathfrak{O}}(f,x)$ for all $f \in \mathcal{R}$ and all $x \in \mathbb{N}$.

By Theorem 2, for every class $\mathcal{U} \in \mathcal{T}\text{-}\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{0,1})$ there is an operator $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{R}_{\mathfrak{O}}$. Let $\hat{\mathfrak{O}}$ be the monotone operator constructed for $\mathfrak{O}$. Consequently, for every function $f \in \mathcal{U}$ there is a φ–program i such that $\varphi_i = f$ and $\forall^\infty x[\Phi_i(x) \leq \hat{\mathfrak{O}}(1^\infty, x)]$. Thus, by the Extrapolation Theorem we can conclude $\mathcal{U} \in \mathcal{NUM}$ (cf. Barzdin and Freivalds [3]).

The same ideas can be used to show[1] the remaining part for $\mathfrak{T}\text{-}\mathcal{REL}$ (cf. Grabowski [9]). Hence, Assertion (2) is shown.

Finally, Assertion (3) is an immediate consequence of Assertion (2) and Theorems 2 and 3 from Stephan and Zeugmann [20] which together show that $\mathcal{NUM} \cap \wp(\mathcal{R}_{0,1}) \subset \mathcal{R}\text{-}\mathcal{REL} \cap \wp(\mathcal{R}_{0,1})$. This completes the proof. ∎

Together with Theorem 4 the latter proof allows a nice corollary.

Corollary 9. *For all* $\delta \in \mathbb{N}$ *we have:*

(1) $\mathcal{CONS}^\delta \subset \mathcal{CONS}^{\delta+1}$,

(2) $\mathcal{R}\text{-}\mathcal{CONS}^\delta \subset \mathcal{R}\text{-}\mathcal{CONS}^{\delta+1}$.

Proof. We use $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$ from the proof of Theorem 8 and $\mathcal{V} = \{\alpha 0^\infty |\ \ \alpha \in \mathbb{N}^*\}$. Clearly, $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$, $\mathcal{V} \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1}$ and hence, by Theorem 4 we also have $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \cup \mathcal{V} \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1}$. Consequently, $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \cup \mathcal{V} \in \mathcal{R}\text{-}\mathcal{CONS}^{\delta+1}$ and $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \cup \mathcal{V} \in \mathcal{CONS}^{\delta+1}$. It remains to argue that $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \cup \mathcal{V} \notin \mathcal{CONS}^\delta$. This will suffice, since $\mathcal{R}\text{-}\mathcal{CONS}^\delta \subseteq \mathcal{CONS}^\delta$.

Suppose the converse, i.e., there is a strategy $S \in \mathcal{P}$ such that $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \cup \mathcal{V} \in \mathcal{CONS}^\delta(S)$. By the choice of $\mathcal{V}$ we can directly conclude that $S \in \mathcal{R}$ and that S has to work consistently with δ–delay on every f^n, $f \in \mathcal{R}$ and $n \in \mathbb{N}$. But this would imply $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \cup \mathcal{V} \in \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$, a contradiction to $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \notin \mathcal{T}\text{-}\mathcal{CONS}^\delta$. ∎

[1]Of course, Grabowski's [9] result that $\mathfrak{T}\text{-}\mathcal{REL} \cap \wp(\mathcal{R}_{0,1}) = \mathcal{NUM} \cap \wp(\mathcal{R}_{0,1})$ directly implies Assertion (2) by using Assertion (1). We included the part $\mathcal{T}\text{-}\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{0,1}) = \mathcal{NUM} \cap \wp(\mathcal{R}_{0,1})$ here to make the paper more self-contained and for explaining the basic proof ideas.

A closer look at the latter proof shows that we have even proved the following corollary shedding some light on the power of our notion of δ–delay.

COROLLARY 10. *$\mathcal{T}$-$\mathcal{CONS}^{\delta+1} \setminus \mathcal{CONS}^{\delta} \neq \emptyset$ for all $\delta \in \mathbb{N}$.*

The situation is comparable to Lange and Zeugmann's [13] bounded example memory learnability BEM_k of languages from positive data, where BEM_k yields an infinite hierarchy such that $\bigcup_{k\in\mathbb{N}} BEM_k$ is a proper subclass of the class of all indexed families that can be conservatively learned.

On the one hand, the latter corollary shows the strength of δ–delay. On the other hand, the δ–delay cannot compensate all the learning power that is provided by the different consistency demands on the domain of the strategies.

THEOREM 11. *$\mathcal{R}$-$\mathcal{CONS} \setminus \mathcal{T}$-$\mathcal{CONS}^{\delta} \neq \emptyset$ for all $\delta \in \mathbb{N}$.*

Proof. The proof can be done by using the class $\mathcal{U} = \{f \mid f \in \mathcal{R},\ \varphi_{f(0)} = f\}$ of self-describing functions. Obviously, $\mathcal{U} \in \mathcal{R}$-$\mathcal{CONS}(S)$ as witnessed by the strategy $S(f^n) = f(0)$ for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$. Now, assuming $\mathcal{U} \in \mathcal{T}$-$\mathcal{CONS}^{\delta}$ for some $\delta \in \mathbb{N}$ would directly imply that $\mathcal{U} \cup \mathcal{V} \in \mathcal{T}$-$\mathcal{CONS}^{\delta}$ for the same δ (here $\mathcal{V}$ is the class defined in the proof of Corollary 9) by Theorem 4. But this is a contradiction to $\mathcal{U} \cup \mathcal{V} \notin \mathcal{LIM}$ as shown in Barzdin [1]. ∎

Corollary 10 and Theorem 11 together imply the following incomparabilities.

COROLLARY 12. *$\mathcal{T}$-$\mathcal{CONS}^{\delta} \# \mathcal{CONS}^{\mu}$ and $\mathcal{T}$-$\mathcal{CONS}^{\delta} \# \mathcal{R}$-$\mathcal{CONS}^{\mu}$ for all $\delta, \mu \in \mathbb{N}$ provided $\delta > \mu$.*

Next we show the analogue to Theorem 11 for $\mathcal{R}$-$\mathcal{CONS}^{\delta}$ and $\mathcal{CONS}$.

THEOREM 13. $\mathcal{CONS} \setminus \mathcal{R}$-$\mathcal{CONS}^{\delta} \neq \emptyset$ for all $\delta \in \mathbb{N}$.

Proof. The proof uses the class $\mathcal{U} = \{f \mid f \in \mathcal{R},$ either $\varphi_{f(0)} = f$ or $\varphi_{f(1)} = f\}$ and ideas from Wiehagen and Zeugmann [25]. As shown in [25], $\mathcal{U} \in \mathcal{CONS}$.

It remains to show that $\mathcal{U} \notin \mathcal{R}$-$\mathcal{CONS}^{\delta}$ for all $\delta \in \mathbb{N}$. Let $\delta \in \mathbb{N}$ be arbitrarily fixed. Suppose there is a strategy $S \in \mathcal{R}$ such that $\mathcal{U} \in \mathcal{R}$-$\mathcal{CONS}^{\delta}_{\varphi}(S)$.

Applying Smullyan's Recursion Theorem, cf. Smullyan [19], we construct a function $f \in \mathcal{U}$ such that either S on successive input f^n the strategy S changes its mind infinitely often or there is an $x \in \mathbb{N}$ such that $\varphi_{S(f^x)}$ violates the δ-delay consistency condition. Since both cases yield a contradiction to the definition of $\mathcal{R}$-$\mathcal{CONS}^{\delta}$, we are done. The wanted function f is defined as follows. Let h and s be two recursive functions such that for all $i, j \in \mathbb{N}, \varphi_{h(i,j)}(0) = \varphi_{s(i,j)}(0) = i$ and $\varphi_{h(i,j)}(1) = \varphi_{s(i,j)}(1) = j$. For any $i, j \in \mathbb{N}$, $x \geq 2$ we proceed inductively.

Suspend the definition of $\varphi_{s(i,j)}$. Define $\varphi_{h(i,j)}$ for more and more arguments via the following procedure. Note that (A) and (B) can be effectively checked, since $S \in \mathcal{R}$.

(T) Test whether or not (A) or (B) happens:

(A) $S(\varphi^x_{h(i,j)}) \neq S(\varphi^x_{h(i,j)}0^{\delta+1})$,

(B) $S(\varphi^x_{h(i,j)}) \neq S(\varphi^x_{h(i,j)}1^{\delta+1})$.

If (A) happens, then let $\varphi_{h(i,j)}(x+1) = \cdots = \varphi_{h(i,j)}(x+\delta+1) = 0$, let $x := x+\delta+2$, and goto (T).
In case (B) happens, set $\varphi_{h(i,j)}(x+1) = \cdots = \varphi_{h(i,j)}(x+\delta+1) = 1$, let $x := x+\delta+2$, and goto (T).
If neither (A) nor (B) happens, then define $\varphi_{h(i,j)}(x') = 0$ for all $x' > x$, and goto ($*$).

($*$) Set $\varphi_{s(i,j)}(n) = \varphi_{h(i,j)}(n)$ for all $n \leq x$, and $\varphi_{s(i,j)}(x') = 1$ for all $x' > x$.

By Smullyan's Recursion Theorem [19], there are numbers i and j such that $\varphi_i = \varphi_{h(i,j)}$ and $\varphi_j = \varphi_{s(i,j)}$. Now we distinguish the following cases.

Case 1. The loop in (T) is never left.

Then $\varphi_i \in \mathcal{R}$ and $\varphi_i(0) = i$. Since $\varphi_j = ij$, and hence a finite function, we obtain $\varphi_i \in \mathcal{U}$. Moreover, in accordance with the definition of the loop (T), on input φ_i^n the strategy S changes its mind infinitely often and thus does not learn φ_i.

Case 2. The loop in (T) is left.

Then there exists an x such that $S(\varphi^x_{h(i,j)}) = S(\varphi^x_{h(i,j)}0^{\delta+1}) = S(\varphi^x_{h(i,j)}1^{\delta+1})$. Moreover, we have $\varphi_{h(i,j)} = \varphi_i$, $\varphi_{s(i,j)} = \varphi_j$ and, by ($*$), $\varphi_i(n) = \varphi_j(n)$ for all $n \leq x$. Additionally, by construction we know that $\varphi_i(0) = i$ and $\varphi_j(1) = j$ as well as $\varphi_i,\ \varphi_j \in \mathcal{R}$. Since in particular $\varphi_i(x+1) \neq \varphi_j(x+1)$, we get $\varphi_i \neq \varphi_j$. Consequently, both functions φ_i and φ_j belong to $\mathcal{U}$.

Now, we see that $S(\varphi_i^x) = S(\varphi_j^x) = S(\varphi_i^x 0^{\delta+1}) = S(\varphi_i^x 1^{\delta+1})$ and additionally

$$\begin{aligned}\varphi_i(x+1) &= \cdots = \varphi_i(x+\delta+1) = 0\\ \varphi_j(x+1) &= \cdots = \varphi_j(x+\delta+1) = 1\ .\end{aligned}$$

Let $k = S(\varphi_j^x)$ and distinguish the following subcases.

Subcase 2.1. $\varphi_k(x+1)$ is not defined or $\varphi_k(x+1)$ is defined and $0 \neq \varphi_k(x+1) \neq 1$.

Then, since we also have $k = S(\varphi_i^x 0^{\delta+1}) = S(\varphi_i^x 1^{\delta+1})$ the δ-delay consistency condition is violated on both inputs $\varphi_i^x 0^{\delta+1}$ and $\varphi_i^x 1^{\delta+1}$ to S, a contradiction to $\mathcal{U} \in \mathcal{R}\text{-}\mathcal{CONS}^{\delta}_{\varphi}(S)$.

Subcase 2.2. $\varphi_k(x+1)$ is defined $\varphi_k(x+1) = 0$ or $\varphi_k(x+1) = 1$.

First, let $\varphi_k(x+1) = 0$. Then we know that $\varphi_k(x+1) \neq \varphi_j(x+1) = 1$. Thus, the hypothesis k which is also output on input $\varphi_j^x 1^{\delta+1}$ (recall that $\varphi_i^x = \varphi_j^x$) is violating the δ-delay consistency condition.

The case $\varphi_k(x+1) = 1$ is handled analogously. Therefore, we get again a contradiction to $\mathcal{U} \in \mathcal{R}\text{-}\mathcal{CONS}^{\delta}_{\varphi}(S)$, and thus there is no strategy $S \in \mathcal{R}$ such that $\mathcal{U} \in \mathcal{R}\text{-}\mathcal{CONS}^{\delta}_{\varphi}(S)$. ∎

Finally, putting Theorem 11 and 13 together we directly arrive at the following corollary.

COROLLARY 14. $\mathcal{T}\text{-}\mathcal{CONS}^\delta \subset \mathcal{R}\text{-}\mathcal{CONS}^\delta \subset \mathcal{CONS}^\delta$ *for all* $\delta \in \mathbb{N}$.

6. Conclusions and Future Work

Looking for possible relaxations for the demand to learn consistently we have introduced the notions of coherent learning and of δ-delay. As our results show, coherent learning with δ-delay has the same learning power as consistent learning with δ-delay for all versions considered. Thus, coherence is in fact no weakening of the consistency demand.

On the other hand, we could establish three new infinite hierarchies of consistent learning in dependence on the delay δ.

The figure below summarizes the achieved separations and coincidences of the various coherent and consistent learning models investigated in this paper.

$$\begin{array}{ccccccccccc}
\mathcal{T}\text{-}\mathcal{COH} & \subset & \mathcal{T}\text{-}\mathcal{COH}^1 & \subset \cdots \subset & \mathcal{T}\text{-}\mathcal{COH}^\delta & \subset & \mathcal{T}\text{-}\mathcal{COH}^{\delta+1} & \subset \cdots \subset & \mathfrak{T}\text{-}\mathcal{REL} \\
\| & & \| & & \| & & \| & & \| \\
\mathcal{T}\text{-}\mathcal{CONS} & \subset & \mathcal{T}\text{-}\mathcal{CONS}^1 & \subset \cdots \subset & \mathcal{T}\text{-}\mathcal{CONS}^\delta & \subset & \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1} & \subset \cdots \subset & \mathfrak{T}\text{-}\mathcal{REL} \\
\cap & & \cap & & \cap & & \cap & & \cap \\
\mathcal{R}\text{-}\mathcal{COH} & \subset & \mathcal{R}\text{-}\mathcal{COH}^1 & \subset \cdots \subset & \mathcal{R}\text{-}\mathcal{COH}^\delta & \subset & \mathcal{R}\text{-}\mathcal{COH}^{\delta+1} & \subset \cdots \# & \mathcal{R}\text{-}\mathcal{REL} \\
\| & & \| & & \| & & \| & & \| \\
\mathcal{R}\text{-}\mathcal{CONS} & \subset & \mathcal{R}\text{-}\mathcal{CONS}^1 & \subset \cdots \subset & \mathcal{R}\text{-}\mathcal{CONS}^\delta & \subset & \mathcal{R}\text{-}\mathcal{CONS}^{\delta+1} & \subset \cdots \# & \mathcal{R}\text{-}\mathcal{REL} \\
\cap & & \cap & & \cap & & \cap & & \cap \\
\mathcal{COH} & \subset & \mathcal{COH}^1 & \subset \cdots \subset & \mathcal{COH}^\delta & \subset & \mathcal{COH}^{\delta+1} & \subset \cdots \subset & \mathcal{LIM} \\
\| & & \| & & \| & & \| & & \| \\
\mathcal{CONS} & \subset & \mathcal{CONS}^1 & \subset \cdots \subset & \mathcal{CONS}^\delta & \subset & \mathcal{CONS}^{\delta+1} & \subset \cdots \subset & \mathcal{LIM}
\end{array}$$

Figure 1: Hierarchies of consistent learning with δ-delay

Moreover, we showed characterization theorems for $\mathcal{CONS}^\delta$ and $\mathcal{T}\text{-}\mathcal{CONS}^\delta$ in terms of complexity. These theorems provide a first explanation for the increase in learning power caused by the δ-delay. On the other hand, the characterization for $\mathcal{T}\text{-}\mathcal{CONS}^\delta$ proved to be very useful for showing the closure of $\mathcal{T}\text{-}\mathcal{CONS}^\delta$ under recursively enumerable unions. Thus, it would be nice to find also a characterization for $\mathcal{R}\text{-}\mathcal{CONS}^\delta$ in terms of complexity. This seems to be a challenging problem.

References

[1] J. M. Barzdin. Две теоремы о предельном синтезе функций. In J. M. Barzdin, editor, Теория Алгоритмов и Программ, volume I, pages 82 – 88. Latvian State University, 1974.

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

[3] J. M. Barzdin and R. V. Freivalds. On the prediction of general recursive functions. *Soviet Math. Dokl.*, 13:1224–1228, 1972.

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

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

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

[7] M. A. Fulk. Saving the phenomenon: Requirements that inductive inference machines not contradict known data. *Inform. Comput.*, 79(3):193–209, 1988.

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

[9] J. Grabowski. Starke Erkennung. In R. Linder and H. Thiele, editors, *Strukturerkennung diskreter kybernetischer Systeme*, volume 82, pages 168–184. Seminarberichte der Sektion Mathematik der Humboldt-Universität zu Berlin, 1986.

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

[11] S. Jain, D. Osherson, J. S. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory, second edition.* MIT Press, Cambridge, Massachusetts, 1999.

[12] K. P. Jantke and H.-R. Beick. Combining postulates of naturalness in inductive inference. *Elektronische Informationsverarbeitung und Kybernetik*, 17(8/9):465–484, 1981.

[13] S. Lange and T. Zeugmann. Incremental learning from positive data. *J. of Comput. Syst. Sci.*, 53(1):88–103, 1996.

[14] E. Minicozzi. Some natural properties of strong identification in inductive inference. *Theoret. Comput. Sci.*, 2:345–360, 1976.

[15] P. Odifreddi. *Classical Recursion Theory.* North Holland, Amsterdam, 1989.

[16] P. Odifreddi. *Classical Recursion Theory, Vol. II.* North Holland, Amsterdam, 1999.

[17] D. N. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists.* MIT Press, Cambridge, Massachusetts, 1986.

[18] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967. Reprinted, MIT Press 1987.

[19] R. M. Smullyan. *Theory of formal systems*, volume 47 of *Annals of Mathematics Studies.* Princton University Press, Princeton, New Jersey, USA, 1961.

[20] F. Stephan and T. Zeugmann. Learning classes of approximations to non-recursive functions. *Theoret. Comput. Sci.*, 288(2):309–341, 2002.

[21] R. Wiehagen. Characterization problems in the theory of inductive inference. In *Automata, Languages and Programming Fifth Colloquium, Udine, Italy, July 17-21, 1978*, volume 62 of *Lecture Notes in Computer Science*, pages 494 – 508, Berlin, 1978. Springer-Verlag.

[22] R. Wiehagen. *Zur Theorie der Algorithmischen Erkennung.* Dissertation B, Humboldt-Universität zu Berlin, 1978.

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

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

[25] R. Wiehagen and T. Zeugmann. Learning and consistency. In *Algorithmic Learning for Knowledge-Based Systems*, volume 961 of *Lecture Notes in Artificial Intelligence*, pages 1–24. Springer, 1995.

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