Consistent and Coherent Learning with δ-Delay

Yohji Akama [a] and Thomas Zeugmann [b]

[a] *Mathematical Institute, Tohoku University, Sendai, Miyagi, 980-8578, Japan*

[b] *Division of Computer Science, Hokkaido University, N-14, W-9, Sapporo 060-0814, Japan*

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, we provide characterizations for consistent learning with δ–delay in terms of complexity and computable numberings. Second, we establish strict hierarchies for all consistent learning models with δ-delay in dependence on δ. Finally, 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.

Key words: Inductive inference, consistency, coherence, characterizations, recursion theory

1 Introduction

Algorithmic learning has attracted much attention of researchers in various fields of computer science. Inductive inference addresses the question whether

Email addresses: akama@math.tohoku.ac.jp (Yohji Akama), thomas@ist.hokudai.ac.jp (Thomas Zeugmann).

or not learning problems may be solved algorithmically at all. Since the pioneering paper of Gold [10], there has been huge progress in the area (cf., e.g., Jain *et al.* [13], Osherson *et al.* [22] and the references therein). On the other hand, several questions still deserve attention. One such question is consistency. A consistent learner is required to correctly and completely reflect all the data it has already seen during the learning process. It should be noted that consistency is also a common requirement in PAC learning, machine learning and statistical learning (cf., e.g., [1,19,26]).

For the sake of illustration, we may think of a typical classroom scenario where we have a teacher and students. In this scenario, a consistent student can always completely and correctly repeat what has been taught so far. However, as experience shows, this is not always the case. Consequently, it is only natural to ask whether or not this behavior of students constitutes an advantage or a disadvantage.

We study this question in the setting of inductive inference of recursive functions. In this setting is already known that, in general, inconsistent learners are more powerful than consistent ones. Therefore, we consider the following modification called *consistent learning with δ-delay.* In this framework, a student consistently learning with δ-delay can always completely and correctly repeat what has been taught so far except the last δ lectures.

Additionally, we introduce a new learning model called *coherent* learning. Intuitively, a coherent student has a very good short term memory, i.e., she can always completely and correctly repeat the content of the last lecture but not necessarily the content of previous lectures. Again, we consider the modification of *coherent learning with δ-delay.*

In order to be more precise, we proceed more formally. A main problem of inductive inference 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 $f(0), f(1), \ldots$, $f(n), \ldots$ a learning strategy is required to compute a sequence of hypotheses $h_0, h_1, \ldots, h_n, \ldots$ 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 $f(0), f(1), \ldots, f(n)$, 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 should a strategy 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 of recursive functions is the inconsistency phenomenon (cf., e.g., Barzdin [3], Blum and Blum [5], Wiehagen and Liepe [29], Jantke and Beick [14] as well as Osherson, Stob and Weinstein [22] 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.* [13], Freivalds, Kinber and Wiehagen [8], Fulk [9], Osherson *et al.* [22] and Wiehagen and Zeugmann [30,31] for further investigations concerning consistent and inconsistent learning.

As already mentioned above, in the present paper we introduce and study several variations of consistent learning that have not been considered in the literature. That is, we introduce the notion of δ-delay to the types of consistent learning mainly studied so far, i.e., to $\mathcal{CONS}$ (defined by Barzdin [3]), $\mathcal{R}$-$\mathcal{CONS}$ (introduced by Jantke and Beick [14]) and $\mathcal{T}$-$\mathcal{CONS}$ (defined by Wiehagen and Liepe [29]) (cf. Definitions 2, 3 and 4, respectively). These definitions differ with respect to the set of admissible strategies, i.e., partial recursive versus recursive and the consistency domain. In general, we are interested in a relaxation of the demand to learn consistently that yields additional learning power for *all* model of consistent learning. Further motivation is provided in Section 2.

Moreover, we define 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 $g(x_n) = f(x_n)$. Additionally, we consider coherent learning with δ-delay. 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 21); here $a \dot{-} b$ denotes the arithmetic difference.

While consistent learning with δ-delay requires correctness for all but the most recent δ data fed to the learner, coherence can be considered as the other extreme. That is, now for all intermediate hypotheses just the function value at argument $n \dot{-} \delta$ must be correctly reflected provided $n \geq \delta$.

Our results are manyfold. First, we provide characterizations for consistent learning with δ-delay in terms of complexity (cf. Theorems 6 and 7) and in terms of computable numberings (cf. Theorems 10, 11, and 12). Second, we establish strict hierarchies for all consistent learning models with δ-delay in dependence on δ, see Theorem 14 and Corollary 15.

Finally, 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 22.

The paper is structured as follows. Section 2 presents notation and definitions. The announced characterizations are shown in Section 3. In Section 4 we prove three new infinite hierarchies for consistent learning with δ-delay. Then we show the equivalence of coherent and consistent learning for all variants defined (cf. Section 5). In Section 6 we discuss the results obtained and present open problems.

2 Preliminaries

Unspecified notations follow Rogers [23]. $\mathbb{N} = \{0, 1, 2, \ldots\}$ 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 the *arithmetic difference* $a \dot{-} b$ 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 $\mathbb{N}$ is denoted by $\mathcal{P}$, $\mathcal{P}^2$, $\mathcal{R}$, and $\mathcal{R}^2$, respectively. Let $f \in \mathcal{P}$, then we use $dom(f)$ to denote the *domain* of the function f, i.e., $dom(f) = \{x \mid x \in \mathbb{N},\ f(x) \text{ is defined}\}$. By $\mathcal{R}_{\{0,1\}}$ we denote 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 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 [23]) 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 [6]. 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 $\mathcal{NUM} = \{\mathcal{U} \mid \exists\psi[\psi \in \mathcal{R}^2 \wedge \mathcal{U} \subseteq \mathcal{P}_\psi]\}$ denote the family of all subsets of all recursively enumerable classes of recursive functions.

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 [10]) *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 $\varphi \in Göd$. 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 [3] 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 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.

In machine learning it is often assumed that learning algorithms are defined on all inputs (cf., e.g., Kodratoff and Michalski [15] as well as Michalski *et al.* [16,17]). On the one hand, this requirement is partially justified by a theorem of Gold [10]. He showed the following. Let $\mathcal{U} \subseteq \mathcal{R}$; if $\mathcal{U} \in \mathcal{LIM}(S)$ then there exists a strategy $S' \in \mathcal{R}$ such that $\mathcal{U} \in \mathcal{LIM}(S')$. That is, learning in the limit is *insensitive* with respect to the requirement to learn exclusively with recursive strategies. Therefore, it is natural to consider the two basic variants of consistent learning where only recursive strategies are admissible.

Jantke and Beick [14] required recursive strategies but restricted the demand to output exclusively consistent hypotheses to functions from the target class $\mathcal{U}$. The second basic variant was introduced by Wiehagen and Liepe [29]. Their definition requires a strategy to be defined and to be consistent on every input. Therefore, we adopt both definitions to our scenario to learn consistently with δ-delay.

Next, we modify $\mathcal{CONS}^{\delta}$ in the same way Jantke and Beick [14] 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.

As mentioned above, in Definition 3 consistency with δ-delay is only demanded for inputs that correspond to some function f from the target class $\mathcal{U}$. In the following definition we incorporate Wiehagen and Liepe's [29] 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* $x + \delta \leq n$.

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

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

Note that another relaxation of the demand to learn consistently has been proposed by Wiehagen [28]. Using the terminology of Daley [7] we say that an error of *commission* occurs at an argument x if $\psi_{S(f^n)}(x)$ is defined and $\psi_{S(f^n)}(x) \neq f(y)$. Furthermore, if $\psi_{S(f^n)}(x)$ is undefined then we have an error of *ommission* at argument x. Now, Wiehagen [28] relaxed $\mathcal{CONS}$ by requiring that $\psi_{S(f^n)}(x)$ does not make any error of commission for $0 \leq x \leq n$, denoted the resulting learning type by $\mathcal{CONF}$, and called it conformity. Moreover, Wiehagen [28] showed $\mathcal{CONS} \subset \mathcal{CONF} \subset \mathcal{LIM}$. Fulk [9] then considered the variation of $\mathcal{CONF}$ that corresponds to Wiehagen and Liepe's [29] definition of $\mathcal{T}$-$\mathcal{CONS}$ and obtained $\mathcal{T}$-$\mathcal{CONF}$. Interestingly, when the set of admissible strategies is restricted then conformity does not yield additional learning power, i.e., we have $\mathcal{T}\text{-}\mathcal{CONF} = \mathcal{T}\text{-}\mathcal{CONS}$ (cf. Fulk [9]).

Before proving our hierarchy results we characterize consistent learning with δ-delay in terms of complexity and in terms of computable numberings, since some of the results obtained will be very helpful to achieve the desired separations.

3 Characterizations

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 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. Furthermore, the sufficiency proofs of characterizations provide *uniform* learning methods that are suitable for all classes learnable under a given inference constraint. We start with the complexity theoretic characterizations.

3.1 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 5 (Rogers [23]) *A mapping $\mathfrak{O}\colon \mathfrak{P} \to \mathfrak{P}$ from partial functions to partial functions is called a* partial recursive operator *iff there is a recursively enumerable set $W \subseteq \mathbb{N}^3$ such that for any $y, z \in \mathbb{N}$ it holds that $\mathfrak{O}(f)(y) = z$ if and only if there is an $x \in \mathbb{N}$ such that $(x, y, z) \in W$ and f extends the finite function F_x.*

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

A mapping $\mathfrak{O}\colon \mathcal{P} \to \mathcal{P}$ 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 [12,20,32]. 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 $x \leq n$, $f(x)$ 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 [28] who proved theses theorems for the case $\delta = 0$. Variants of these characterizations for $\delta = 0$ can also be found in Wiehagen and Liepe [29] as well as in Odifreddi [21].

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 [5] we define operator complexity classes as follows. Let $\mathfrak{O}$ be any computable operator; then we set

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

That is, a function $f \in \mathcal{R}$ belongs to the operator complexity class $\mathcal{C}_{\mathfrak{O}}$ if it possesses a program i computing it and the complexity Φ_i of program i is bounded by $\mathfrak{O}(f)$ almost everywhere.

In the following proofs we shall construct *a-priori* bounds for the computational complexity of the functions to be learned. For all functions in the target class, these bounds are uniformly computable by using a computable operator which operates on a uniformly bounded range of the graph of the input function. That is, for computing $\mathfrak{O}(f, n)$ the operator needs only the function values $f(0), \dots, f(n + \delta)$. Then we shall show that the knowledge of such *a-priori* bounds for the computational complexity of the functions to be learned suffices not only for their identification, but also for the synthesis of a program having a complexity less than or equal to the upper bound almost everywhere.

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

Theorem 6 *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 $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.*

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 $n \ge \delta$, 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 $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.

Sufficiency. Let $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$. 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 $\mathcal{C}_{\mathfrak{O}}$ we know that for every $f \in \mathcal{U}$ there are i and k such that $\varphi_i = f$ and $\Phi_i(x) \leq \max\{k,\ \mathfrak{O}(f,x)\}$ for all x. Thus, the desired strategy S searches for the first such 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 $aux \in \mathcal{R}$ such that for all $\alpha \in \mathbb{N}^*$, $\varphi_{aux(\alpha)} = \alpha 0^\infty$.

$S(f^n) =$ "If $n < \delta$ then output $aux(f^n)$.
If $n \geq \delta$ then compute $\mathfrak{O}(f,x)$ for all $x \leq n \dot{-} \delta$. Search for the least $z \leq n$ such that for $c_2(z) = (i,k)$ the conditions
(A) $\Phi_i(x) \leq \max\{k,\ \mathfrak{O}(f,x)\}$ for all $x \leq n \dot{-} \delta$, 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 $S(f^n) = aux(f^n)$."

Assume $n \geq \delta$. 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 7 *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 $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.*

Proof. The necessity is proved *mutatis mutandis* as in the proof of Theorem 6 with the only modification that $\mathfrak{O}(f,n)$ 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 $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$.

The only modification for the sufficiency part is to leave $S(f^n)$ undefined if $\mathfrak{O}(f,x)$ is not defined for some $x \leq n \dot{-} \delta$. Note that this can happen if and only if $f \notin \mathcal{U}$. We omit the details. □

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

We finish this subsection by using Theorem 6 to show that $\mathcal{T}\text{-}\mathcal{CONS}^\delta$ 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 8 *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 $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$. Furthermore, a program for S can be found recursively from a program enumerating programs for all and only $(S_i)_{i\in\mathbb{N}}$.*

Proof. The proof of the necessity of Theorem 6 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 $\mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{C}_{\mathfrak{O}_i}$.

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

By construction, we directly obtain $\mathfrak{O} \in \Omega_{id+\delta}$ and $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$. Furthermore, it is easy to see that $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{C}_{\mathfrak{O}}$. Hence, by Theorem 6 we can conclude $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}^\delta(S_i) \subseteq \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$. $\square$

On the other hand, $\mathcal{CONS}^\delta$ and $\mathcal{R}\text{-}\mathcal{CONS}^\delta$ are not even closed under finite union. This is a direct consequence of a more general result Barzdin [2] 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.

3.2 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 [31], next we define these decidable consistency conditions.

Definition 9 *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 [27] and from Wiehagen and Zeugmann [31].

Theorem 10 *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

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

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

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

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 $\alpha \in \mathbb{N}^*$, $i \in \mathbb{N}$, $cons(\alpha, i) = 1$ if and only if $\alpha \sqsubset_\delta \psi_i$. Let $\alpha = (a_0, \ldots, a_x) \in \mathbb{N}^*$ and $i \in \mathbb{N}$. Let $e(i) = (z, n)$. Then define

$$cons(\alpha, i) = \begin{cases} 1, & \text{if } x < \delta \\ 1, & \text{if } \delta \le x \le n \text{ and, for every } y \text{ such that } y + \delta \le x,\ a_y = \psi_i(y) \\ 1, & \text{if } \delta \le x,\ x > n \text{ and, for every } y \le \min\{n,\ x - \delta\},\ a_y = \psi_i(y), \\ & \text{and } S(a_0, \ldots, a_y) = z \text{ for every } y \in \mathbb{N} \text{ such that } n < y + \delta \le x \\ 0, & \text{otherwise.} \end{cases}$$

Since $e(i) = (z, n) \in M$, by construction we know that $\varphi_z(m)$ is defined for all $m \le n$ and $S(\varphi_z^n) = z$. Thus, we have $\psi_i(m) = \varphi_z(m)$ for all $m \le n$. Consequently, if $x \le n$, then for all $y \le x$ it can be effectively tested whether or not $a_y = \psi_i(y)$. Furthermore, $S \in \mathcal{R}$ implies that $S(a_0, \ldots, a_y)$ can be computed for every $y \in \mathbb{N}$ such that $n < y + \delta \le x$. Thus, if $x > n$ the condition $S(a_0, \ldots, a_y) = z$ can be effectively checked for every $y \in \mathbb{N}$ such that $n < y + \delta \le x$. 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 $\alpha = (a_0, \ldots, a_x)$. If $\delta \le x \le n$ then we have $\alpha \sqsubset_\delta \psi_i$, since this has been checked in the second case of the definition of *cons*.

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

$$S(a_0, \ldots, a_y) = z \text{ for every } y \in \mathbb{N} \text{ such that } n < y + \delta \le x\ . \tag{1}$$

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

$$\varphi_z(m) = a_m \text{ for all } m \text{ such that } m + \delta \le x\ . \tag{2}$$

By construction $\psi_i(x) = \varphi_z(x)$ for all $x \le n$. By (2) we can conclude that $\varphi_z(y)$ is defined for all y such that $n < y + \delta \le x$. Finally, (1) implies that $S(\varphi_z^y) = z$ for all y such that $n < y + \delta \le x$. Thus, $\psi_i(m) = \varphi_z(m)$ for all m such that $m + \delta \le x$. Therefore, by (2) 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 $\delta \le x$ and $x + \delta \le n$ 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 $m + \delta \le x$, since otherwise $\psi_i(m)$

is not defined for $n < m + \delta \leq x$. Additionally, $S(\varphi_z^y) = z$ for all y with $n < y + \delta \leq x$. Thus, $S(a_0, \ldots, a_y) = z$ 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 $\mathcal{U} \subset \mathcal{P}_\psi$. For learning any function $f \in \mathcal{U}$ consistently with δ-delay, 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 $\varphi \in$ *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 $\mathcal{U} \in \mathcal{T}\text{-}\mathcal{CONS}_\varphi^\delta$. $\square$

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

Theorem 11 *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 10. 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 $\alpha \sqsubset_\delta \psi_i$. $\square$

Finally, we characterize $\mathcal{CONS}^\delta$. Again the proof is analogous to the one of Theorem 10 and therefore omitted.

Theorem 12 *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.*

4 Hierarchy Results

In 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 [5] and Minicozzi [18]. Intuitively, a learner M is reliable provided it converges if and only if it learns.

Definition 13 (Blum and Blum [5], Minicozzi [18]) *Let $\mathcal{U} \subseteq \mathcal{R}$, $\mathcal{M} \subseteq \mathfrak{T}$ and let $\varphi \in Göd$. 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) *$\mathcal{U} \in \mathcal{LIM}_\varphi(S)$, 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 14 *The following statements hold for all $\delta \in \mathbb{N}$:*

(1) $\mathcal{T}$-$\mathcal{CONS}^\delta \subset \mathcal{T}$-$\mathcal{CONS}^{\delta+1}$,
(2) $\bigcup_{\delta\in\mathbb{N}} \mathcal{T}$-$\mathcal{CONS}^\delta \subset \mathfrak{T}$-$\mathcal{REL}$,
(3) $\mathcal{NUM} \cap \wp(\mathcal{R}_{\{0,1\}}) = \mathcal{T}$-$\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{\{0,1\}}) = \mathfrak{T}$-$\mathcal{REL} \cap \wp(\mathcal{R}_{\{0,1\}})$,
(4) $\mathcal{T}$-$\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{\{0,1\}}) \subset \mathcal{R}$-$\mathcal{REL} \cap \wp(\mathcal{R}_{\{0,1\}})$.

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

The desired strategy S is defined as follows. Let $aux \in \mathcal{R}$ be the function defined in the sufficiency proof of Theorem 6. For all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$ we set

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

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

In order to show this claim, it is technically advantageous to have the following notations. For two finite sequences of natural numbers σ and τ, let $\sigma \diamond \tau$ denote the concatenation of σ and τ. For all $g \in \mathcal{P}$, $m \in \mathbb{N}$, such that, for all $k < m$, $g(k)$ is defined, let $g[m]$ denote $g(0) \diamond \ldots \diamond g(m-1)$. Note that $g[0]$ denotes the empty sequence. For all $g \in \mathcal{P}$, $m < n \in \mathbb{N}$, such that, for all k with $m \leq k \leq n$, $g(k)$ is defined, let $g[m,n]$ denote $g(m) \diamond \ldots \diamond g(n-1)$.

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 f belonging to $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$ but on which S fails. By Kleene's recursion theorem (cf. Rogers [23], Exercise 11-4) find $e \in \mathbb{N}$ such that for all $x \in \mathbb{N}$

$$\varphi_e(x) = \begin{cases} e, & \text{if } x < \delta + 1\ , \\ \Phi_e(x-\delta-1)+1, & \text{else if } x \text{ is divisible by } \delta+1 \text{ and} \\ & S(\varphi_e[x]) = S(\varphi_e[x] \diamond \Phi_e[x-\delta-1, x])\ , \\ \Phi_e(x-\delta-1), & \text{otherwise.} \end{cases}$$

We set $f = \varphi_e$. By induction it is easy to see that f is total and $f \in \mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$. By the definition of a complexity measure, we conclude $\Phi_e \in \mathcal{R}$, too.

By our supposition, S learns $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)}$ and S is δ-delayed consistent. Consequently, there must exist a $t_0 \in \mathbb{N}$ such that (i) t_0 is divisible by $\delta+1$, (ii) $\varphi_{S(f[t_0])} = f$, and (iii) $S(f[m]) = S(f[t_0])$ for all $m > t_0$. We distinguish the following cases.

Case 1. $f(t_0) = \Phi_e(t_0 - \delta - 1) + 1$

Then by (ii) we have

$$\varphi_{S(f[t_0])}(t_0) = f(t_0) = \Phi_e(t_0 - \delta - 1) + 1 \; . \tag{3}$$

Furthermore, by construction we have $S(f[t_0]) = S(f[t_0] \diamond \Phi_e[t_0 - \delta - 1, t_0])$. Since S is δ-delayed consistent we get

$$f(t_0) = \varphi_{S(f[t_0] \diamond \Phi_e[t_0-\delta-1, t_0])}(t_0) = \Phi_e(t_0 - \delta - 1) \; ,$$

a contradiction to (3). Thus, this case cannot happen.

Case 2. $f(t_0) = \Phi_e(t_0 - \delta - 1)$

We show

$$f[t_0] \diamond \Phi_e[t_0 - \delta - 1, t_0] = f[t_0 + \delta + 1] \; . \tag{4}$$

Since both sequences have the same length, the following shows the equality.

For $0 \le x < t_0$ the equality is obvious. For $x = t_0$ we have the equality by the assumption of Case 2.

Finally, let $t_0 < x \le t_0 + \delta$. By (i) we know that t_0 is divisible by $\delta + 1$. Thus, x is not divisible by $\delta + 1$. Hence, by the definition of f we know that $f(x) = \Phi_e(x - \delta - 1)$, and the equality is shown.

On the other hand, by (4) and (iii) we get $S(f[t_0]) = S(f[t_0] \diamond \Phi_e[t_0 - \delta - 1, t_0])$. Consequently, by construction of f we must have $f(t_0) = \Phi_e(t_0 - \delta - 1) + 1$, a contradiction.

So, Case 2 cannot happen either, and thus $\mathcal{U}_{\delta+1}^{(\varphi,\Phi)} \notin \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$.

This proves Claim 2. Assertion (1) now follows from Claim 1 and 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 [18], $\mathfrak{T}\text{-}\mathcal{REL}$ 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 (2).

For showing Assertion (3), 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 $\mathfrak{O}(f, x) \le \hat{\mathfrak{O}}(f, x)$ 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 $\forall x[f(x) \le g(x)]$ implies $\forall x[\mathfrak{O}(f, x) \le \mathfrak{O}(g, x)]$. This can be seen as follows. Let $\alpha = (\alpha_0, \dots, \alpha_m)$ and $\beta = (\beta_0, \dots, \beta_m)$ be any tuples of length $m + 1$ from $\mathbb{N}^*$. We write $\alpha \le \beta$ if $\alpha_i \le \beta_i$ for all $i = 0, \dots, m$. Now, let $\mathfrak{O} \in \Omega_{id+\delta}$.

We define

$$\hat{\mathfrak{O}}(\beta, x) = \max\{\mathfrak{O}(\alpha, x) \mid |\alpha| = |\beta| = x + \delta + 1,\ \alpha \leq \beta\} .$$

Since the operator $\mathfrak{O}$, for computing the value $\mathfrak{O}(\alpha, x)$, just needs the values $\alpha_0, \ldots, \alpha_{x+\delta}$, we see that $\hat{\mathfrak{O}}$ 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 6, 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 $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$. 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 [4]).

The same ideas can be used to show[1] the remaining part for $\mathfrak{T}\text{-}\mathcal{REL}$ (cf. Grabowski [11]). Hence, Assertion (3) is shown.

Finally, Assertion (4) is an immediate consequence of Assertion (3) and Theorems 2 and 3 from Stephan and Zeugmann [25] 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. $\square$

Together with Theorem 8 the proof of Theorem 14 allows for a nice corollary.

Corollary 15 *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 $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1}$ from the proof of Theorem 14 and $\mathcal{V} = \{\alpha 0^\infty \mid \alpha \in \mathbb{N}^*\}$. Clearly, $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1}$, $\mathcal{V} \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1}$ and hence, by Theorem 8 we also have $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1} \cup \mathcal{V} \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1}$. Consequently, $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1} \cup \mathcal{V} \in \mathcal{R}\text{-}\mathcal{CONS}^{\delta+1}$ and $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1} \cup \mathcal{V} \in \mathcal{CONS}^{\delta+1}$. It remains to argue that $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1} \cup \mathcal{V} \notin \mathcal{CONS}^\delta$. 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 $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1} \cup \mathcal{V} \in \mathcal{CONS}^\delta(S)$. 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 implies $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1} \cup \mathcal{V} \in \mathcal{T}\text{-}\mathcal{CONS}^\delta(S)$, a contradiction to $\mathcal{U}^{(\varphi,\Phi)}_{\delta+1} \notin \mathcal{T}\text{-}\mathcal{CONS}^\delta$. $\square$

[1] Of course, Grabowski's [11] result that $\mathfrak{T}\text{-}\mathcal{REL} \cap \wp(\mathcal{R}_{\{0,1\}}) = \mathcal{NUM} \cap \wp(\mathcal{R}_{\{0,1\}})$ directly implies Assertion (3) by using Assertion (2). 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 proof of Corollary 15 shows that we have even proved the following corollary shedding some light on the power of our notion of δ-delay.

Corollary 16 *$\mathcal{T}$-$\mathcal{CONS}^{\delta+1} \setminus \mathcal{CONS}^{\delta} \neq \emptyset$ for all $\delta \in \mathbb{N}$.*

On the one hand, the Corollary 16 shows the strength of δ-delay. On the other hand, the δ-delay cannot compensate for all the learning power that is provided by the different consistency demands on the domain of the strategies.

Theorem 17 *$\mathcal{R}$-$\mathcal{CONS} \setminus \mathcal{T}$-$\mathcal{CONS}^{\delta} \neq \emptyset$ for all $\delta \in \mathbb{N}$.*

Proof. The proof uses the class $\mathcal{U} = \{f \mid f \in \mathcal{R},\ \varphi_{f(0)} = \ f\}$ 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 15) by Theorem 8. But this is a contradiction to $\mathcal{U} \cup \mathcal{V} \notin \mathcal{LIM}$ as shown in Barzdin [2]. $\square$

Corollary 16 and Theorem 17 together imply the following incomparabilities.

Corollary 18 *$\mathcal{T}$-$\mathcal{CONS}^{\delta}$ # $\mathcal{CONS}^{\mu}$ and $\mathcal{T}$-$\mathcal{CONS}^{\delta}$ # $\mathcal{R}$-$\mathcal{CONS}^{\mu}$ for all δ, $\mu \in \mathbb{N}$ provided $\delta > \mu$.*

Next we show the analogue to Theorem 17 for $\mathcal{R}$-$\mathcal{CONS}^{\delta}$ and $\mathcal{CONS}$.

Theorem 19 *$\mathcal{CONS} \setminus \mathcal{R}$-$\mathcal{CONS}^{\delta} \neq \emptyset$ 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 [31]. As shown in [31], $\mathcal{U} \in \mathcal{CONS}$.

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 [24], we construct a function $f \in \mathcal{U}$ such that either S changes its mind infinitely often when successively fed f^n 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 [24], 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 $\varphi_k(x+1) \notin \{0,1\}$.

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 and $\varphi_k(x+1) \in \{0,1\}$.

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 $\mathcal{U} \in \mathcal{R}\text{-}\mathcal{CONS}_\varphi^\delta(S)$, and thus there is no strategy $S \in \mathcal{R}$ such that $\mathcal{U} \in \mathcal{R}\text{-}\mathcal{CONS}_\varphi^\delta(S)$. $\square$

Finally, putting Theorem 17 and 19 together we directly arrive at the following corollary.

Corollary 20 $\mathcal{T}\text{-}\mathcal{CONS}^\delta \subset \mathcal{R}\text{-}\mathcal{CONS}^\delta \subset \mathcal{CONS}^\delta$ *for all* $\delta \in \mathbb{N}$.

5 Coherence and Consistency

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.

As already mentioned at the end of Section 2 another relaxation of consistency is conformity. A conform strategy is only allowed to make errors of ommission for the function values it has already seen. But of course, as Wiehagen's [28] result $\mathcal{CONS} \subset \mathcal{CONF} \subset \mathcal{LIM}$ shows, in general it is impossible for a strategy to find out whether or not it makes an error of ommission for the function values it has already received.

Therefore, in this section we look at the other extreme. What is the knowledge worth to know that always a *particular* function value already seen is correctly reflected. Clearly, if it is always the same one, then this knowledge is useless, since this value can also be patched in the sequence of hypotheses without altering its convergence. The same trick works *mutatis mutandis* for any finite fixed set of positions. This observation led to our definition of coherent learning with δ-delay presented below.

Definition 21 *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* $n \geq \delta$.

$\mathcal{COH}_\psi^\delta(S)$, $\mathcal{COH}_\psi^\delta$ *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 $\mathcal{R}$-$\mathcal{COH}^\delta$ and $\mathcal{T}$-$\mathcal{COH}^\delta$, 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 $LT \in \{\mathcal{COH}^\delta,\ \mathcal{R}\text{-}\mathcal{COH}^\delta,\ \mathcal{T}\text{-}\mathcal{COH}^\delta\}$ we have $LT_\varphi = LT$ for every $\varphi \in$ *Göd*.

Next, 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.

If we look again at the teacher student scenario described in the Introduction, then the answer should be intuitively clear. A coherent student always correctly remembers the content of the last lecture. Supposing we have access to coherent students $s_0, s_1, \ldots, s_n$ having attended lecture $0, 1, 2, \ldots, n$, respectively, we can correctly reconstruct the content of all these n lectures by asking student s_i about lecture i, $i = 0, \ldots, n$. So, coherent learning should be exactly as powerful as consistent learning. Now, it should also be clear that this intuition extends when the δ-delay is included. Consequently, the main technical problem is then the simulation of the coherent students.

Theorem 22 *Let $\delta \in \mathbb{N}$ be arbitrarily fixed. Then we have*

(1) $\mathcal{CONS}^\delta = \mathcal{COH}^\delta$,
(2) $\mathcal{R}$-$\mathcal{CONS}^\delta = \mathcal{R}$-$\mathcal{COH}^\delta$,
(3) $\mathcal{T}$-$\mathcal{CONS}^\delta = \mathcal{T}$-$\mathcal{COH}^\delta$.

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

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 $\hat{S}(f^0), \ldots, \hat{S}(f^n)$ 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 $f(0), \ldots, f(n^*)$. If $n - n^* \le \delta$, 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 $k = 1, \ldots, n - n^*$. Thus, for all $k \in \{1, \ldots, n - n^* - \delta\}$ we get

$$\varphi_{S(f^n)}(n^* + k) = \varphi_{\hat{S}(f^{n^*})}(n^* + k) = \varphi_{\hat{S}(f^{n^*+\delta+k})}(n^* + k) = f(n^* + k) \ . \quad (5)$$

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. Moreover, by assumption we know that $\hat{S}$ learns f and thus S also learns f. 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 (5) is true for all $f \in \mathcal{R}$. This completes the proof. $\square$

6 Conclusions and Future Work

Looking for possible relaxations for the demand to learn consistently we have introduced the notions δ-delay and of coherent learning. 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.

Moreover, we showed characterization theorems for $\mathcal{CONS}^\delta$ and $\mathcal{T}$-$\mathcal{CONS}^\delta$ in terms of complexity and in terms of computable numberings. These theorems provide a first explanation for the increase in learning power caused by the δ-delay. Looking at the characterizations in terms of computable numberings, we see that differences between $\mathcal{CONS}^\delta$, $\mathcal{R}$-$\mathcal{CONS}^\delta$ and $\mathcal{T}$-$\mathcal{CONS}^\delta$ have been traced back to the decidability of different consistency-related decision problems.

Our characterizations in terms of complexity express the difference between $\mathcal{CONS}^\delta$ and $\mathcal{T}$-$\mathcal{CONS}^\delta$ by different sets of admissible operators, i.e., $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$ versus $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$. Moreover, the power of the δ-delay is nicely reflected by

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

Fig. 1. Hierarchies of consistent learning with δ-delay

the amount of data needed to compute $\mathfrak{O}(f, n)$, i.e., $\mathfrak{O} \in \Omega_{id+\delta}$. Furthermore, 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. The difficulty here is that the conditions $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$ and $\mathfrak{O} \in \Omega_{id+\delta}$ cannot be changed. Varying $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ to $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$ explains the difference between $\mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ and $\mathcal{CONS}^{\delta}$. Consequently, the only parameter allowing a further variation is the domain of the admissible operators. In order to characterize $\mathcal{R}\text{-}\mathcal{CONS}^{\delta}$ one should demand $dom(\mathfrak{O}) = \mathcal{R}$, $\mathfrak{O}(\mathcal{U}) \subseteq \mathcal{R}$ and $\mathcal{U} \subseteq \mathcal{C}_{\mathfrak{O}}$. While these requirements can easily be shown to be necessary, it is hard too see their sufficiency, since an $\mathcal{R}$-consistent learner with δ-delay must output a hypothesis on every input.

Acknowledgements

The authors heartily thank the anonymous referees for their careful reading and the many valuable comments made. In particular we are grateful to one referee for proposing a simplified proof for Claim 2 in Theorem 14.

References

[1] M. Anthony and N. Biggs. *Computational Learning Theory.* Cambridge Tracts in Theoretical Computer Science (30). Cambridge University Press, Cambridge,

1992.

[2] 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.

[3] 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).

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

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

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

[7] R. P. Daley. On the error correcting power of pluralism in BC-type inductive inference. *Theoret. Comput. Sci.*, 24(1):95–104, 1983.

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

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

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

[11] 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.

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

[13] 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.

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

[15] Y. Kodratoff and R. Michalski, editors. *Machine Learning: An Artificial Intelligence Approach*, volume III. Morgan Kaufmann, Los Altos, California, 1990.

[16] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors. *Machine Learning: An Artificial Intelligence Approach*, volume I. Morgan Kaufmann, Los Altos, California, 1983.

[17] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors. *Machine Learning: An Artificial Intelligence Approach*, volume II. Morgan Kaufmann, Los Altos, California, 1986.

[18] E. Minicozzi. Some natural properties of strong identification in inductive inference. *Theoret. Comput. Sci.*, 2(3):345–360, 1976.

[19] T. M. Mitchell. *Machine Learning.* WCB/McGraw-Hill, Boston, Massachusetts, 1997.

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

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

[22] 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.

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

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

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

[26] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, Berlin, 2nd edition, 2000.

[27] 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.

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

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

[30] 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.

[31] 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.

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