Learning by Erasing

Steffen Lange
Division of Electronics and
Information Engineering
Hokkaido University
Sapporo 060, Japan
steffen@db.huee.hokudai.ac.jp

Rolf Wiehagen
Department of Computer Science
University of Kaiserslautern
P.O. Box 3049
67653 Kaiserslautern, Germany
wiehagen@informatik.uni-kl.de

Thomas Zeugmann
Research Institute of
Fundamental Information Science
Kyushu University 33
Fukuoka 812, Japan
thomas@rifis.kyushu-u.ac.jp

Abstract

Learning by erasing means the process of eliminating potential hypotheses from further consideration thereby converging to the least hypothesis never eliminated and this one must be a solution to the actual learning problem.

The present paper deals with learnability by erasing of indexed families $\mathcal{L}$ of languages from both positive data as well as positive and negative data. This refers to the following scenario. A family $\mathcal{L}$ of target languages and a hypothesis space for it are specified. The learner is fed eventually all positive examples (all labeled examples) of an unknown target language L chosen from $\mathcal{L}$. The target language L is learned by erasing if the learner erases some set of possible hypotheses and the least hypothesis never erased correctly describes L.

The capabilities of learning by erasing are investigated in dependence on the requirement of what sets of hypotheses have to be or may be, erased, and in dependence of the choice of the hypothesis space.

Class preserving learning by erasing ($\mathcal{L}$ has to be learned w.r.t. some suitably chosen enumeration of all and only the languages from $\mathcal{L}$), *class comprising* learning by erasing ($\mathcal{L}$ has to be learned w.r.t. some hypothesis space containing at least all the languages from $\mathcal{L}$), and *absolute* learning by erasing ($\mathcal{L}$ has to be learned w.r.t. all class preserving hypothesis spaces for $\mathcal{L}$) are distinguished.

For all these models of learning by erasing necessary and sufficient conditions for learnability are presented. A complete picture of all separations and coincidences of the learning by erasing models is derived. Learning by erasing is compared with standard models of language learning such as learning in the limit, finite learning and learning without overgeneralization. The exact location of these types within the hierarchy of the learning by erasing models is established.

1. Introduction

Learning by erasing means the process of eliminating potential hypotheses from further consideration thereby converging to a hypothesis which will never be eliminated and which has to be a correct solution to the actual learning problem.

This approach is motivated by similarities to both human learning or, more general, human problem solving as well as automated problem solving. Actually, in solving a problem we mostly find out several "non-solutions" to that problem first, contradicting the data we have or explaining them in another unsatisfying way. Of course, we then will exclude these non-solutions from our further consideration and keep only a more or less explicitly given remaining set of potential solutions. Often, at any time of the solving process we have an actual "favored candidate" among all the remaining candidates for a solution which, though, up to now cannot be proved to be really a solution and which also may change from time to time. Then at least the following can happen. Eventually we find a solution to the problem, can even prove its correctness and hence successfully stop the solving process. Or, our "favored candidate" will be stable from some point on, it is really a solution, but we are not absolutely sure of that. The latter case is a version of successful learning in the limit, which is what we do in building theories or, even more real-world, in writing computer programs.

In our approach of learning by erasing we can model both situations of being successful above. However, our main intention is a rigorous study of learning by erasing *in the limit*. In particular, we are mainly interested in characterizing and comparing the general capabilities of such learners, i.e., learners that achieve their goal by erasing non-appropriate hypotheses.

A special case of our approach, the so-called co-learning, was introduced in Freivalds, Karpinski and Smith (1994), and then further studied in Freivalds, Gobleja, Karpinski and Smith (1994) for learning of recursively enumerable classes of recursive functions. In that case the learner has to eliminate all hypotheses but one and this one has to be correct. This approach was then used by Kummer (1995) in order to show that a recursively enumerable class of recursive functions is co-learnable with respect to all of its numberings iff all of these numberings are equivalent (i.e., intercompilable), thus giving a learning-theoretic solution to a longstanding problem of recursion-theoretic numbering theory.

We relax the all-but-one approach by giving the learner more freedom on which sets are allowed to erase eventually. We consider the following possibilities:

- *e-ARB* - an arbitrary set of hypotheses may be erased,
- *e-MIN* - exactly all hypotheses less than the least correct one have to be erased,
- *e-SUB* - only incorrect hypotheses may be erased,
- *e-EQ* - exactly all incorrect hypotheses have to be erased,
- *e-SUPER* - all incorrect hypotheses have to be erased and an arbitrary set of correct hypotheses may be erased, too,
- *e-ALL* - all but one hypotheses have to be erased.

Our objects to be learned are indexed families of languages, i.e., recursively enumerable classes of uniformly recursive languages. We consider both modes of information presentation established in language learning, text (positive data, only) and informant (positive and negative data). And we study class preserving learning (the spaces of hypotheses exactly enumerate the language family to be learned), class comprising learning (the spaces of hypotheses enumerate a possibly proper superset of the family to be learned) and absolute learning (the families have to be learned with respect to all hypothesis spaces enumerating them exactly). Note that the e-ALL-case, more exactly, co-learning of indexed families of languages from text, was already studied in Freivalds and Zeugmann (1995).

Our results can be classified along the lines of characterizations, comparisons inside, and comparisons with known types of language learning.

Characterizations. For all types of learning by erasing we present characterizations, i.e., conditions that are both necessary and sufficient for learnability in the corresponding sense. Mostly these characterizations are stated in terms being independent from learning theory. In several cases the corresponding condition is a purely structural one, namely that the language family may not contain any language together with a proper sublanguage. In other cases, for example for e-SUB both in the class preserving and the class comprising case, the characterization comes to the "granularity" of deriving necessary and sufficient learnability conditions for any given pair of a language family and a hypothesis space. Such granularity results were already derived in language learning theory (cf., e.g., Angluin (1980), Lange and Zeugmann (1992), Kapur and Bilardi (1995), Zeugmann, Lange and Kapur (1995), and Baliga, Case and Jain (1996)). Surprisingly, our characterizations do work without the explicit use of so-called "tell-tales" which were commonly used in all previous characterizations in language learning. Even more surprisingly, up to now no such granularity results are known in Gold's (1967) paradigm of learning recursive functions. There the basic structure of most of the characterizations is the following. Given a class U of recursive functions and some learning type LT; then $U \in LT$ iff *there is a suitable* space of hypotheses such that ... (cf. Wiehagen (1978)). Note that also some characterizations in language learning have this "there is" flavor (cf. Jain and Sharma (1994)).

Comparisons inside. We derive a complete picture containing all separations and coincidences of the types of learning by erasing defined. Fortunately, this picture is of a pretty regular structure and not as sophisticated as sometimes in inductive inference. Several of the separations and coincidences follow from the characterizations above.

Comparisons with known types of language learning. We compare the types of language learning by erasing with well-known "standard" types of learning indexed language families such as learning in the limit, finite learning and conservative learning (or, equivalently, learning without overgeneralization (cf. Lange and Zeugmann (1993b, 1993c)). We present the exact location of these established learning types in the hierarchy of the types of language learning by erasing.

The paper is organized as follows. Section 2 presents notations and definitions. The announced hierarchies for learning by erasing from positive data are established in Section 3, and the subsections therein. Class preserving learning by erasing is studied in Subsection 3.1, while the Subsections 3.2 and 3.3 deal with class comprising

and absolute learning by erasing, respectively. Section 4 studies learning by erasing from informant. The characterization theorems are established in Section 5. Finally, in Section 6 we discuss the results obtained and outline open problems. All references are given in Section 7.

2. Notations and Definitions

Unspecified notations follow Rogers (1967). Let $\mathbb{N} = \{0, 1, 2, ...\}$ be the set of natural numbers. We set $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. By $\langle \cdot, \cdot \rangle : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ we denote ***Cantor's pairing function***, i.e., $\langle x, y \rangle = ((x+y)^2 + 3x + y)/2$ for all $x, y \in \mathbb{N}$. We use $\mathcal{P}^n$ and $\mathcal{R}^n$ to denote the set of all n-ary partial recursive and total recursive functions over $\mathbb{N}$, respectively. The class of all $\{0,1\}$ valued functions $f \in \mathcal{R}^n$ is denoted by $\mathcal{R}^n_{0,1}$. For $n = 1$ we omit the upper index, i.e., we set $\mathcal{P} = \mathcal{P}^1$, $\mathcal{R} = \mathcal{R}^1$, and $\mathcal{R}_{0,1} = \mathcal{R}^1_{0,1}$.

Every function $\psi \in \mathcal{P}^2$ is called a numbering. Moreover, let $\psi \in \mathcal{P}^2$, then we write ψ_j instead of $\lambda x \psi(j, x)$. Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ denote any fixed ***Gödel numbering*** of $\mathcal{P}$, and let $\Phi_0, \Phi_1, \Phi_2, \ldots$ be any associated ***complexity measure*** (cf. Blum (1967)). Furthermore, let $\psi \in \mathcal{R}^2_{0,1}$, then by $L(\psi_j)$ we denote the language generated or described by ψ_j, i.e., $L(\psi_j) = \{x \mid \ \psi_j(x) = 1, \ x \in \mathbb{N}\}$, and by $co\text{-}L(\psi_j)$ its complement, i.e., $\mathbb{N} \setminus L(\psi_j)$. Moreover, we call $\mathcal{L} = (L(\psi_j))_{j \in \mathbb{N}}$ an ***indexed family*** (cf. Angluin (1980)). For the sake of presentation, we restrict ourselves to consider exclusively indexed families of non-empty languages. An indexed family is said to be ***inclusion-free*** iff $L \not\subset \hat{L}$ for all languages L, $\hat{L} \in range(\mathcal{L})$. Let $\mathcal{L}$ be an indexed family. Every numbering $\psi \in \mathcal{R}^2_{0,1}$ is called ***hypothesis space***. A hypothesis space ψ is said to be ***class comprising*** for an indexed family $\mathcal{L}$ iff $range(\mathcal{L}) \subseteq \{L(\psi_j) \mid \ j \in \mathbb{N}\}$. Furthermore, we call a hypothesis space $\psi \in \mathcal{R}^2_{0,1}$ ***class preserving*** for $\mathcal{L}$ iff $range(\mathcal{L}) = \{L(\psi_j) \mid \ j \in \mathbb{N}\}$.

Let $\mathcal{L}$ be any indexed family, let $L \in range(\mathcal{L})$, and let ψ be any class comprising hypothesis space for $\mathcal{L}$. Then we set $min_\psi(L) = \min\{j \mid L(\psi_j) = L\}$.

Let L be a language and let $t = s_0, s_1, s_2, \ldots$ be an infinite sequence of natural numbers such that $content(t) = \{s_k \mid \ k \in \mathbb{N}\} = L$. Then t is said to be a ***text*** for L or, synonymously, a ***positive presentation***. By $text(L)$ we denote the set of all positive presentations of L. Moreover, let t be a text, and let y be a number. Then t_y denotes the initial segment of t of length $y+1$, i.e., $t_y = s_0, \ldots, s_y$. Finally, t_y^+ denotes the ***content*** of t_y, i.e., $t_y^+ = \{s_z \mid \ z \leq y\}$.

Next, we introduce the notion of the ***canonical text*** that turned out to be helpful in proving some characterizations. Let L be any non-empty recursive language, and let $0, 1, 2, \ldots$ be the ordered text of $\mathbb{N}$. The canonical text of L is obtained as follows. Test sequentially whether $z \in L$ for $z = 0, 1, 2, \ldots$ until the first z is found such that $z \in L$. Since $L \neq \emptyset$ there must be at least one z fulfilling the test. Set $t_0 = z$. We proceed inductively. For all $x \in \mathbb{N}$ we define:

$$t_{x+1} = \begin{cases} t_x, z+x+1, & \text{if} \quad z+x+1 \in L, \\ t_x, n, & \text{otherwise, where } n \text{ is the last element in } t_x. \end{cases}$$

As in Gold (1967), we define an ***inductive inference machine*** (abbr. IIM) to be an algorithmic device which works as follows: The IIM takes as its input incrementally

increasing initial segments of a text and it either requests the next input, or it first outputs a hypothesis, i.e., a number, and then it requests the next input.

We interpret the hypotheses output by an IIM with respect to some suitably chosen hypothesis space $\psi \in \mathcal{R}^2_{0,1}$. When an IIM outputs a number j, we interpret it to mean that the machine is hypothesizing the language $L(\psi_j)$.

Furthermore, we define an ***erasing learning machine*** (abbr. ELM) to be an algorithmic device working as follows: The ELM takes as its input incrementally increasing initial segments of a text (as an IIM does) and then it either requests the next input, or it first outputs a number, and then it requests the next input.

However, there is a *major* difference in the *semantics* of the output of an IIM and an ELM, respectively. Let $\psi \in \mathcal{R}^2_{0,1}$ be any hypothesis space. Suppose an ELM M has been successively fed an initial segment t_y of a text t, and it has output numbers $j_0, \ldots, j_z$. Then we interpret $j = \min(\mathbb{N} \setminus \{j_0, \ldots, j_z\})$ as M's ***actual guess***. Intuitively, if an ELM outputs a number j, then it *definitely deletes* j from its list of potential hypotheses.

Let M be an IIM or an ELM, let t be a text, and $y \in \mathbb{N}$. Then we use $M(t_y)$ to denote the last number that has been output by M when successively fed t_y. We define convergence of IIMs as usual. Let t be a text, and let M be an IIM. The sequence $(M(t_y))_{y \in \mathbb{N}}$ is said to ***converge*** to a number j iff either $(M(t_y))_{y \in \mathbb{N}}$ is infinite and all but finitely many terms of it are equal to j, or $(M(t_y))_{y \in \mathbb{N}}$ is non-empty and finite, and its last term is j.

An ELM M is said to ***stabilize*** to a number j on a text t iff its sequence of actual guesses converges to j, i.e., $j = \min(\mathbb{N} \setminus \{M(t_y) \mid \ y \in \mathbb{N}\})$.

Now we are ready to define learning and learning by erasing.

Definition 1. (Gold, 1967) *Let $\mathcal{L}$ be an indexed family, let L be a language, and let $\psi \in \mathcal{R}^2_{0,1}$ be a hypothesis space.* ***An IIM M CLIM–identifies L from text with respect to ψ*** *iff for every text t for L, there exists a $j \in \mathbb{N}$ such that the sequence $(M(t_y))_{y \in \mathbb{N}}$ converges to j and $L = L(\psi_j)$.*

Furthermore, M CLIM–identifies $\mathcal{L}$ with respect to ψ iff, for each $L \in range(\mathcal{L})$, M CLIM–identifies L with respect to ψ.

Finally, let CLIM denote the collection of all indexed families $\mathcal{L}$ for which there are an IIM M and a hypothesis space ψ such that M CLIM–identifies $\mathcal{L}$ with respect to ψ.

Since, by the definition of convergence, only finitely many data of L were seen by the IIM upto the (unknown) point of convergence, whenever an IIM identifies the language L, some form of learning must have taken place. For this reason, hereinafter the terms *infer*, *learn*, and *identify* are used interchangeably.

In Definition 1, *LIM* stands for "limit." Furthermore, the prefix C is used to indicate ***class comprising*** learning, i.e., the fact that $\mathcal{L}$ may be learned with respect to some class comprising hypothesis space ψ for $\mathcal{L}$. The restriction of *CLIM* to class preserving hypothesis spaces is denoted by *LIM* and referred to as ***class preserving*** inference. Moreover, we use the prefix A to express the fact that an indexed family $\mathcal{L}$ may be inferred with respect to *all* class preserving hypothesis spaces for $\mathcal{L}$, and we refer to this learning model as to ***absolute*** learning. We adopt this convention in the

definitions of the learning types below.

The following proposition clarifies the relations between absolute, class preserving and class comprising learning in the limit.

Proposition 1. (Lange and Zeugmann, 1993c)

ALIM = LIM = CLIM.

Note that, in general, it is not decidable whether or not an IIM M has already converged on a text t for the target language L. With the next definition, we consider a special case where it has to be decidable whether or not an IIM has successfully finished the learning task.

Definition 2. (Gold, 1967; Trakhtenbrot and Barzdin, 1970) *Let $\mathcal{L}$ be an indexed family, let L be a language, and let $\psi \in \mathcal{R}^2_{0,1}$ be a hypothesis space.* ***An IIM M CFIN–identifies L from text with respect to ψ*** *iff for every text t for L, there exists a $j \in \mathbb{N}$ with $L = L(\psi_j)$ such that M, when successively fed t, outputs the single hypothesis j, and stops thereafter.*

Furthermore, M CFIN–identifies L with respect to ψ iff, for each $L \in range(\mathcal{L})$, M CFIN–identifies L with respect to ψ.

The resulting learning type is denoted by CFIN.

The following proposition states that, if an indexed family $\mathcal{L}$ can be *CFIN*–learned with respect to some hypothesis space ψ for it, then it can be finitely inferred with respect to every class preserving hypothesis space for $\mathcal{L}$.

Proposition 2. (Zeugmann, Lange and Kapur, 1995)

AFIN = FIN = CFIN.

Now, we define ***conservative*** IIMs. Intuitively, conservative IIMs maintain their actual hypothesis at least as long as they have not received data that "provably misclassify" it. Hence, whenever a conservative IIM performs a mind change it is because it has perceived a clear contradiction between its hypothesis and the actual input.

Definition 3. (Angluin, 1980) *Let $\mathcal{L}$ be an indexed family, let L be a language, and let $\psi \in \mathcal{R}^2_{0,1}$ be a hypothesis space.* ***An IIM M CCONSV–identifies L from text with respect to ψ*** *iff*

(1) *M CLIM–identifies L with respect to ψ,*

(2) *for all texts $t \in text(L)$ and for all $y, k \in \mathbb{N}$, if $M(t_y) \neq M(t_{y+k})$ then $t^+_{y+k} \not\subseteq L(\psi_{M(t_y)})$.*

Finally, M CCONSV–identifies $\mathcal{L}$ with respect to ψ iff, for each $L \in range(\mathcal{L})$, M CCONSV–identifies L with respect to ψ.

The resulting collection of sets CCONSV is defined analogously as above.

The following proposition shows that conservative learning is sensitive to the particular choice of the hypothesis space.

Proposition 3. (Lange and Zeugmann, 1993b)

ACONSV $\subset$ CONSV $\subset$ CCONSV $\subset$ ALIM.

Next, we define learning by erasing.

Definition 4. *Let $\mathcal{L}$ be an indexed family, let L be a language, and let $\psi \in \mathcal{R}^2_{0,1}$ be a hypothesis space.* ***An ELM M e-CARB–identifies L from text with respect to ψ*** *iff for every text t for L, there exists a $j \in \mathbb{N}$ with $L = L(\psi_j)$ such that M on t stabilizes to j.*

Furthermore, M e-CARB–identifies $\mathcal{L}$ with respect to ψ iff, for each $L \in range(\mathcal{L})$, M e-CARB–identifies L with respect to ψ.

Finally, let e-CARB denote the collection of all indexed families $\mathcal{L}$ for which there are an ELM M and a hypothesis space ψ such that M e-CARB–identifies $\mathcal{L}$ with respect to ψ.

Definition 5. *Let $\mathcal{L}$ be an indexed family, let L be a language, and let $\psi \in \mathcal{R}^2_{0,1}$ be a hypothesis space.* ***An ELM M is said to***

(A) ***e-CSUB–identify L from text with respect to ψ***

(B) ***e-CEQ–identify L from text with respect to ψ***

(C) ***e-CSUPER–identify L from text with respect to ψ***

(D) ***e-CALL–identify L from text with respect to ψ***

(E) ***e-CMIN–identify L from text with respect to ψ***

iff

M e-CARB–identifies L from text with respect to ψ and, moreover, the following conditions are satisfied

(A) $\{M(t_y) \mid y \in \mathbb{N}\} \subseteq \{j \mid L(\psi_j) \neq L,\ j \in \mathbb{N}\}$, *i.e., M is only allowed to erase hypotheses that are incorrect for L;*

(B) $\{M(t_y) \mid y \in \mathbb{N}\} = \{j \mid L(\psi_j) \neq L,\ j \in \mathbb{N}\}$, *i.e., M has to erase exactly all hypotheses that are incorrect for L;*

(C) $\{M(t_y) \mid y \in \mathbb{N}\} \supseteq \{j \mid L(\psi_j) \neq L,\ j \in \mathbb{N}\}$, *i.e., M has to erase all hypotheses that are incorrect for L but it may additionally erase correct hypotheses for L;*

(D) $\mathrm{card}(\mathbb{N} \setminus \{M(t_y) \mid y \in \mathbb{N}\}) = 1$, *i.e., M has to erase all but one hypothesis;*

(E) $\{M(t_y) \mid y \in \mathbb{N}\} = \{j \in \mathbb{N} \mid j < min_\psi(L)\}$, *i.e., M has to erase exactly all hypotheses prior to the least correct index for L.*

We denote by *e-CSUB*, *e-CEQ*, *e-CSUPER*, *e-CALL*, and *e-CMIN* the collection of all those indexed families $\mathcal{L}$ for which there are a hypothesis space ψ and an ELM M inferring every language of it in the sense of *e-CSUB*, *e-CEQ*, *e-CSUPER*, *e-CALL*, and *e-CMIN* with respect to ψ, respectively.

All the types above have in common that at any step of the learning process the "favored candidate" will always be the *least* hypothesis not yet eliminated. This may seem somewhat arbitrary, but in our opinion it is justified by the following observations. First, by the principle of Occam's razor simple hypotheses should be "favored." Second, in case that even in the limit "many" hypotheses remain uncanceled, we get a distinguished final hypothesis, just the least uncanceled one, and thus one can decide from outside whether or not the learning process was successful. And third, more

formally, in case the learning machine eventually finds a provably correct hypothesis, then it can eliminate all the other hypotheses up to that one (or even all but that one) thereby making that hypothesis the least uncanceled one.

Note that *e-ALL* coincides with co-learning from positive data as defined in Freivalds and Zeugmann (1995). Thus, all our definitions may be regarded as natural variations of this learning type.

As already mentioned in the Introduction, Freivalds, Karpinski and Smith (1994) recently studied co-learnability of recursive functions. On the other hand, in inductive inference functions and languages are usually very different from each other (cf., e.g., Osherson, Stob and Weinstein (1986) and the references therein). Hence, it is only natural to ask whether or not there are major differences between the learnability by erasing of recursive functions and recursive languages, too. In this paper, we provide both similarities and distinctions. However, the overall goal is much far-reaching. In particular, we are mainly interested in the general learning capabilities of learners that achieve their learning goal by erasing non-appropriate hypotheses.

3. Learning from Text

In this section, we compare the learning capabilities of all learning by erasing models from positive data to one another as well as to finite inference, learning in the limit and conservative identification from text. Moreover, we analyze the power of learning by erasing in dependence on the set of admissible hypothesis spaces. Subsection 3.1 is dealing with class preserving learning while Subsections 3.2 and 3.3 are dealing with class comprising and absolute learning, respectively.

3.1. Class Preserving Learning by Erasing

We start our investigations by considering class preserving hypothesis spaces. Our goal is to obtain a complete picture of the learning capabilities of all learning by erasing models. Moreover, we establish lower and upper bounds for the learning by erasing models by comparing them with finite inference, learning in the limit and conservative identification. Our first theorem actually points to similarities and differences of the learning by erasing models defined above.

Theorem 1.

(1) $FIN \subset e\text{-}EQ \subset e\text{-}SUB \subset CONSV$,

(2) *For all* $LT \in \{ARB,\ SUPER,\ ALL\}$, $e\text{-}LT = LIM$.

Proof. First, we show Assertion (1).

Claim A. $FIN \subset e\text{-}EQ$.

Let $\mathcal{L}$ be any indexed family. Hence, there is a class preserving hypothesis space $\psi \in \mathcal{R}^2_{0,1}$ for $\mathcal{L}$ having a decidable equivalence problem, i.e., there is a $p \in \mathcal{R}^2_{0,1}$ such that, for all $j,\ k \in \mathbb{N}$, $p(j,k) = 1$ iff $\psi_j = \psi_k$. By Proposition 2 there exists an IIM M finitely inferring $\mathcal{L}$ with respect to ψ. The desired ELM $\hat{M}$ works with respect to ψ, and is defined as follows. Let $L \in range(\mathcal{L})$, $t \in text(L)$, and let $y \in \mathbb{N}$. On input t_y,

the ELM $\hat{M}$ simulates M on input t_y. Now, two cases are possible. First, M outputs nothing and request the next input. In this case $\hat{M}$ requests the next input, too, and makes no output. Second, M outputs a hypothesis j and stops. Due to the definition of FIN we have $L = L(\psi_j)$. Then $\hat{M}$ outputs, one at a time, all natural numbers k with $p(j,k) = 0$. Clearly, $\hat{M}$ erases all hypotheses that are incorrect for L, and hence it indeed e-EQ–infers L.

In order to separate FIN and e-EQ consider the indexed family $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$ with $L_j = \mathbb{N} \setminus \{j\}$ for all $j \in \mathbb{N}$. Since $\mathcal{L} \notin FIN$ (cf. Freivalds and Zeugmann (1995)), it remains to verify that $\mathcal{L} \in e$-EQ. For that purpose, select the hypothesis space $\psi \in \mathcal{R}^2_{0,1}$ with $L(\psi_j) = L_j$ for all $j \in \mathbb{N}$. The wanted ELM M can be defined as follows. Let $L \in range(\mathcal{L})$, let $t = s_0, s_1, s_2, \ldots$ be any text for L, and let $y \in \mathbb{N}$. On input t_y, M simply outputs the index s_y. By definition, $s_y \notin L(\psi_{s_y})$, but $s_y \in L$. Therefore, M erases only hypotheses that are incorrect for L. Finally, since $L = \mathbb{N} \setminus \{j\}$ for some $j \in \mathbb{N}$, M outputs all natural numbers but j, and therefore it learns L as required.

Claim B. e-$EQ \subset e$-SUB.

Since e-$EQ \subseteq e$-SUB (cf. Definition 5), it suffices to show that e-$SUB \setminus e$-$EQ \neq \emptyset$. For all $j \in \mathbb{N}$, let $L_j = \{0, \ldots, j\}$, and set $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$.

First, we verify that $\mathcal{L} \in e$-SUB. Choose the hypothesis space $\psi \in \mathcal{R}^2_{0,1}$ with $L(\psi_j) = L_j$ for all $j \in \mathbb{N}$. The wanted ELM M can be defined as follows. Let $L \in range(\mathcal{L})$, let $t = s_0, s_1, s_2, \ldots$ be any text for L, and let $y \in \mathbb{N}$. We distinguish the cases $s_y = 0$ and $s_y \neq 0$. If $s_y = 0$, then M outputs nothing, and requests the next input. Otherwise, M outputs the hypothesis $s_y - 1$.

Clearly, if $L = L_{\psi_0}$ then M never makes an output, and thus, it stabilizes to 0. Now, suppose that $L = L(\psi_{j+1})$ for some $j \in \mathbb{N}$. Since L is finite, there is a least $z \in \mathbb{N}$ such that $L = t^+_z$. Hence, after having successively read t_z, the ELM M has output all numbers from $\{0, \ldots, j\}$, and the least non-erased hypothesis is $j + 1$. Finally, since $t^+_z = t^+_{z+y}$ for all $y \in \mathbb{N}$, all remaining outputs belong to $\{0, \ldots, j\}$, too. Consequently, M stabilizes to $j + 1$. Since all erased hypotheses are incorrect, M e-SUB–identifies L.

It remains to show that $\mathcal{L} \notin e$-EQ. Suppose the converse, i.e., there are a class preserving hypothesis space $\psi \in \mathcal{R}^2_{0,1}$ for $\mathcal{L}$ and an ELM M witnessing $\mathcal{L} \in e$-EQ with respect to ψ. Consider M when fed the text $t = 0, 0, 0, \ldots$ for L_0. Since M has to e-EQ–identify L_0 from t, it is requested to output all numbers $j \geq 0$ with $L_0 \neq L(\psi_j)$. Let j be M's first output made on t_y for some $y \in \mathbb{N}$. Furthermore, $L(\psi_j) \in range(\mathcal{L})$, and hence $L_0 \subset L(\psi_j)$. Consequently, t_y may be extended to a text $\hat{t}$ for $L(\psi_j)$. Since M, when fed $\hat{t}$, outputs a hypothesis that is correct for $L(\psi_j)$, it fails to e-EQ–learn $L(\psi_j)$, a contradiction.

Claim C. e-$SUB \subset CONSV$.

Let $L_0 = \mathbb{N}$ and for all $j \in \mathbb{N}$, let $L_{j+1} = \{j\}$ as well as $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$. Then, $\mathcal{L} \in CONSV \setminus e$-$SUB$. Obviously, $\mathcal{L} \in CONSV$. Suppose $\mathcal{L} \in e$-SUB. Thus, there are a class preserving hypothesis space ψ and an ELM M witnessing $\mathcal{L} \in e$-SUB with respect to ψ. Let $k = min_\psi(L_0)$; since $range(\mathcal{L})$ is infinite, there must be an $L \in range(\mathcal{L})$ such that $min_\psi(L) > k$. By construction, $L \subset L_0$. Hence, when successively fed the text t for L the ELM M has to output k eventually, say on t_y.

But now, one may again extend t_y to a text for $L(\psi_k)$ on which M outputs the correct guess k, a contradiction.

It remains to prove $e\text{-}SUB \subseteq CONSV$. Let $\mathcal{L} \in e\text{-}SUB$ with respect to some class preserving hypothesis space ψ. Then, for all $L \in range(\mathcal{L})$, and all $j \in \mathbb{N}$,

$$j < min_\psi(L) \text{ implies } L \not\subset L(\psi_j) \qquad (+)$$

For seeing this, suppose the converse, i.e., there are a language $L \in range(\mathcal{L})$ and an index $j < min_\psi(L)$ such that $L \subset L(\psi_j)$. When successively fed any text $t \in text(L)$ the ELM has to output j, say on t_y. Since $L \subset L(\psi_j)$, this initial segment may be extended to a text for $L(\psi_j)$ on which M outputs the correct guess j, a contradiction.

Now, let $L \in range(\mathcal{L})$, $t \in text(L)$, and let $x \in \mathbb{N}$. The desired conservative IIM works with respect to ψ and is defined as follows.

***IIM M*:** "On input t_x determine the least j with $t_x^+ \subseteq L(\psi_j)$. Output j."

By construction, if $M(t_x) \neq M(t_{x+z})$ for some $z \in \mathbb{N}$, then $t_{x+z}^+ \not\subseteq L(\psi_{M(t_x)})$; thus M is conservative. Let $k = min_\psi(L)$. Hence, M's unbounded search always terminates, and M outputs in every stage a hypothesis. Moreover, by $(+)$ we have $L \setminus L(\psi_j) \neq \emptyset$ for all $j < k$. Consequently, every index $j < k$ must be abandoned eventually. Thus, M converges to k. This proves Claim C.

Next, we show Assertion (2). By definition, $e\text{-}ALL \subseteq e\text{-}SUPER \subseteq e\text{-}ARB$. Since $LIM \subseteq e\text{-}ALL$ (cf. Freivalds and Zeugmann (1995)) it suffices to show $e\text{-}ARB \subseteq LIM$.

Claim D. *Let $\mathcal{L}$ be an indexed family, and let $\psi \in \mathcal{R}_{0,1}^2$ be any class comprising hypothesis space for $\mathcal{L}$. Then, $\mathcal{L} \in e\text{-}CARB$ with respect to ψ implies $\mathcal{L} \in CLIM$ with respect to ψ.*

Let M be an ELM witnessing $\mathcal{L} \in e\text{-}CARB$ with respect to ψ. We define the desired IIM $\hat{M}$ as follows. Let $L \in range(\mathcal{L})$, $t \in text(L)$, and let $y \in \mathbb{N}$. $\hat{M}$ simulates M on input t_y, it always outputs the least ψ-number not yet definitely deleted by M, and requests the next input. Obviously, $\hat{M}$ $CLIM$–learns L with respect to ψ. Since Claim D especially holds for any class preserving hypothesis space, Assertion (2) follows. q.e.d.

Theorem 1 and Proposition 3 together allow the following corollary summarizing the inclusions and equalities known so far.

Corollary 2.

$e\text{-}EQ \subset e\text{-}SUB \subset e\text{-}ALL = e\text{-}SUPER = e\text{-}ARB$.

Moreover, Claim D above directly allows the following corollary.

Corollary 3. *For all $LT \in \{SUPER,\ ALL,\ ARB\}$, $e\text{-}LT = e\text{-}CLT = LIM$.*

In the next subsection, we study the relations between class preserving and class comprising learning for the remaining learning types.

3.2. Class Comprising Learning by Erasing

Taking Corollary 3 into account, it remains to investigate the learning power of $e\text{-}CSUB$ and $e\text{-}CEQ$. The following theorem provides the desired complete picture.

Theorem 4.

(1) $e\text{-}EQ = e\text{-}CEQ$,

(2) $e\text{-}SUB \subset e\text{-}CSUB \subset LIM$,

(3) $e\text{-}CSUB \ \# \ CCONSV$.

Proof. First, we verify Assertion (1). Since $e\text{-}EQ \subseteq e\text{-}CEQ$, it suffices to show that $e\text{-}CEQ \subseteq e\text{-}EQ$. The following claim provides general insight into the topological structure of the $e\text{-}CEQ$–identifiable indexed families.

Claim A. *Let $\mathcal{L}$ be an indexed family. If $\mathcal{L} \in e\text{-}CEQ$, then $\mathcal{L}$ is inclusion-free.*

Let $\mathcal{L}$ be any indexed family, and let M be any ELM witnessing $\mathcal{L} \in e\text{-}CEQ$ with respect to some class comprising hypothesis space ψ. Suppose to the contrary that there are L, $\hat{L} \in range(\mathcal{L})$ with $L \subset \hat{L}$. Let $t \in text(L)$; since M $e\text{-}CEQ$–learns L from t, M has to delete sometimes a ψ–index for $\hat{L}$, i.e., there is a least $y \in \mathbb{N}$ such that $M(t_y) = j$ and $L(\psi_j) = \hat{L}$. Because of $L \subset \hat{L}$, t_y can be extended to a text $\hat{t} \in text(\hat{L})$. Moreover, $\hat{L} \in range(\mathcal{L})$. Thus M must $e\text{-}CEQ$–indentify $\hat{L}$ from $\hat{t}$. However, M, when fed the initial segment $\hat{t}_y$, outputs a correct ψ–index for $\hat{L}$, a contradiction, and therefore Claim A follows.

Now we are ready to finish the proof of Assertion (1).

Claim B. *Let $\mathcal{L}$ be an inclusion-free indexed family. Then $\mathcal{L} \in e\text{-}EQ$.*

Let $\mathcal{L}$ be any inclusion-free indexed family. Choose any class preserving hypothesis space $\psi \in \mathcal{R}^2_{0,1}$ for $\mathcal{L}$, and define the desired ELM M as follows. Let $L \in range(\mathcal{L})$, $t \in text(L)$, and let $x \in \mathbb{N}$.

***ELM M*:** "On input t_x proceed as follows:
If $x = 0$, then set $Alive_0 = \mathbb{N}$, output nothing and request the next input.
Otherwise, test for all $k \in Alive_{x-1}$ with $k \leq x$ whether or not $t_x^+ \not\subseteq L(\psi_k)$. If there is at least one index passing this test, output the minimal one, say $\hat{k}$, and set $Alive_x = Alive_{x-1} \setminus \{\hat{k}\}$.
Otherwise, set $Alive_x = Alive_{x-1}$ output nothing and request the next input."

By construction, M never outputs a correct ψ–index for L. It remains to show that M eventually deletes all indices k with $L(\psi_k) \neq L$. This can be seen as follows.

Let $x \in \mathbb{N}$ and let k be the least ψ–index in the set of all remaining candidate hypotheses $Alive_x$ that meets $L(\psi_k) \neq L$. Since $\mathcal{L}$ is inclusion-free, we have $L \setminus L(\psi_k) \neq \emptyset$. Thus, there has to be a least $y \in \mathbb{N}$ such that $t_y^+ \not\subseteq L(\psi_k)$. Now let $z = \max\{x+1, y\}$. Since $t_y^+ \subseteq t_z^+$, we may conclude that $t_z^+ \not\subseteq L(\psi_k)$, too. Consequently, $M(t_z) = k$, and therefore $k \notin Alive_{z+r}$ for all $r \in \mathbb{N}$. Finally, by simply iterating this argumentation it follows that M eventually outputs all ψ–indices k that are incorrect for L. Hence, M $e\text{-}EQ$–identifies L.

Putting Claim A and B together we immediately obtain $e\text{-}CEQ \subseteq e\text{-}EQ$. This finishes the proof of Assertion (1).

We continue in showing Assertions (2) and (3).

Claim C. $CONSV \setminus e\text{-}CSUB \neq \emptyset$.

After a bit of reflection one easily verifies that the indexed family $\mathcal{L}$ used in the proof of Theorem 1, Claim C separates $CONSV$ and $e\text{-}CSUB$, too.

Since $CONSV \subset CCONSV \subset LIM$ (cf. Proposition 3), we obtain $CCONSV \setminus e\text{-}CSUB \neq \emptyset$ and $LIM \setminus e\text{-}CSUB \neq \emptyset$. By definition, any ELM that $e\text{-}CSUB$–identifies an indexed family $\mathcal{L}$ with respect to a hypothesis space ψ witnesses $\mathcal{L} \in e\text{-}CARB$ with respect to ψ as well. Since, additionally, $e\text{-}CARB = LIM$ (cf. Corollary 3), we may conclude that $e\text{-}CSUB \subset LIM$. The following claim provides us the remaining part of Assertions (2) and (3).

Claim D. $e\text{-}CSUB \setminus CCONSV \neq \emptyset$.

First, we define the desired indexed family $\mathcal{L}$ witnessing the claimed separation. For the sake of presentation, we describe $\mathcal{L}$ as a family of languages over the alphabet $\Sigma = \{a, b, c\}$.

For all $k \in \mathbb{N}$, we set $L_{\langle k,0\rangle} = \{a^k b^n \mid n \in \mathbb{N}\}$. Note that $a^0 = \varepsilon$ by convention, where ε denotes the empty word. For all $k \in \mathbb{N}$ and all $j > 0$, we distinguish the following cases:

Case 1. $\neg\ \Phi_k(k) \leq j$

Then we set $L_{\langle k,j\rangle} = L_{\langle k,0\rangle}$.

Case 2. $\Phi_k(k) \leq j$

We distinguish the following subcases.

Subcase 2.1. $\Phi_k(k) \leq j \leq 2\Phi_k(k)$

We set $L_{\langle k,j\rangle} = \{a^k b^m \mid m \leq j - \Phi_k(k)\}$.

Subcase 2.2. $j > 2\Phi_k(k)$

Then we set $L_{\langle k,j\rangle} = L_{\langle k,0\rangle}$.

Finally, let $\mathcal{L} = (L_{\langle k,j\rangle})_{k,j\in\mathbb{N}}$. Since $\mathcal{L} \notin CCONSV$ (cf. Lange and Zeugmann (1993a), Theorem 1), it suffices to show that $\mathcal{L} \in e\text{-}CSUB$.

The desired ELM M works with respect to the following class comprising hypothesis space $\mathcal{H} = (H_{\langle k,j\rangle})_{k,j\in\mathbb{N}}$. For all $j,\ k \in \mathbb{N}$, we set:

$$H_{\langle k,j\rangle} = \begin{cases} L_{\langle k,j\rangle} \cup \{a^k c^{\Phi_k(k)}\}, & \text{if } \neg\ \Phi_k(k) \leq j, \\ L_{\langle k,j\rangle}, & \text{otherwise.} \end{cases}$$

Note that $L_{\langle k,j\rangle} \cup \{a^k c^{\Phi_k(k)}\}$ equals $L_{\langle k,j\rangle}$, if $\varphi_k(k)$ is undefined. By definition, $\mathcal{H}$ serves as an class comprising hypothesis space for $\mathcal{L}$. We continue in defining the desired ELM M. Let $L \in range(\mathcal{L})$, let $t \in text(L)$, and let $x \in \mathbb{N}$.

***ELM M*:** "On input t_x do the following:

Determine the unique $k \in \mathbb{N}$ such that $t_x^+ \subseteq \{a^k b^n \mid n \in \mathbb{N}\}$. Test whether or not $\Phi_k(k) \leq x$. If not, goto (α). Otherwise, goto (β).

(α) If $x < \langle k,0\rangle$, then output x. Otherwise, output $\langle k,0\rangle - 1$. Request the next input.

(β) Let $z = M(t_{x-1})$. Fix $j = \max\{m \mid a^k b^m \in t_x^+\}$, and test whether or not $j \leq \Phi_k(k)$. In case it is, goto $(\beta 1)$. Else, goto $(\beta 2)$.

(β1) If $z+1 < \langle k, j+\Phi_k(k)\rangle$, output $z+1$, and request the next input. Otherwise, output z, and request the next input.

(β2) If $z+1 < \langle k, 2\Phi_k(k)+1\rangle$, output $z+1$, and request the next input. Otherwise, output z, and request the next input."

Let $L = L_{\langle k,j\rangle}$ for some $k,\ j \in \mathbb{N}$. In order to verify M's correctness we distinguish the following cases.

Case 1. $\varphi_k(k)$ is undefined.

Hence, $L = L_{\langle k,j\rangle} = H_{\langle k,j\rangle}$ for all $j \in \mathbb{N}$. By construction, M outputs all and only the numbers z with $z < \langle k, 0\rangle$. Let $\hat{z} = \langle m, n\rangle$ be any of these indices. By definition of Cantors pairing function, $\hat{z} = \langle m, n\rangle < \langle k, 0\rangle$ implies $m \neq k$. Thus, M deletes exclusively indices z meeting $H_z \neq L$, and we are done.

Case 2. $\varphi_k(k)$ is defined.

Let $y \in \mathbb{N}$ such that $\Phi_k(k) = y$. We distinguish the following subcases.

Subcase 2.1. L is finite.

By definition of $\mathcal{L}$, there has to be a $j \leq y$ such that $L = L_{\langle k,y+j\rangle}$. Moreover by $\mathcal{H}$'s definition, $L = H_{\langle k,y+j\rangle}$. Since L is finite, there has to be a least $x > y$ such that $t_x^+ = L$. By construction, M, when successively fed $t_{x+\langle k,y+j\rangle}$, outputs all indices $z < \langle k, y+j\rangle$, and it never erases any index exceeding $\langle k, y+j\rangle - 1$.

It remains to show that $H_z \neq L$ for all $z < \langle k, y+j\rangle$. This can be seen as follows. Assume any $\hat{z} = \langle m, n\rangle$ with $\hat{z} < \langle k, y+j\rangle$. Clearly, $m \neq k$ immediately implies $H_{\langle m,n\rangle} \neq L$ for all $n \in \mathbb{N}$. Now, suppose that m equals k, i.e., $\hat{z} = \langle k, n\rangle$. By definition of Cantors pairing function, $\langle k, n\rangle < \langle k, y+j\rangle$ results in $n < y+j$. Since $\varphi_k(k)$ is defined, we know that $H_{\langle k,n\rangle} \notin \mathit{range}(\mathcal{L})$ for all $n < y$. Finally, let $j \geq 1$, and let $n \in \{y, \ldots, y+j-1\}$. Thus, $H_{\langle k,n\rangle} = L_{\langle k,n\rangle}$. By definition of $\mathcal{L}$, $L_{\langle k,n\rangle} \subset L_{\langle k,y+j\rangle} = L$, and thus $H_{\langle k,n\rangle} \neq L$. Hence, we are done.

Subcase 2.2. L is infinite.

Hence, $L = L_{\langle k,0\rangle} = H_{\langle k,2y+1\rangle}$. Since L is infinite and $\varphi_k(k)$ is defined, there has to be a least $x \in \mathbb{N}$ such that both $x \geq y$ and $\max\{m \mid a^k b^m \in t_x^+\} > y$ are fulfilled. Again, by construction, M has deleted all indices z with $z < \langle k, 2y+1\rangle$, when successively fed $t_{x+\langle k,2y+1\rangle}$. Furthermore, M is never outputting any other index.

As above, it remains to show that $H_z \neq L$ for all $z < \langle k, 2y+1\rangle$. However, $H_{\langle k,y+r\rangle}$ is finite for all r with $0 \leq r \leq y$. Therefore, the idea that has been successful used in handling the above subcase applies *mutatis mutandis* to verify that, for all $z < \langle k, 2y+1\rangle$, H_z does not equal L.

To sum up, M *e-CSUB*–learns L, and therefore M witnesses $\mathcal{L} \in$ *e-CSUB* with respect to $\mathcal{H}$.

Clearly, Assertion (3) follows immediately by Claim C and Claim D. Finally, since *e-SUB* $\subseteq$ *e-CSUB* and *e-SUB* $\subset$ *CONSV* (cf. Theorem 1), we obtain *e-SUB* $\subset$ *e-CSUB* via Claim D. Thus Assertions (2) and (3) are proved. q.e.d.

A closer look at the proof of Claim C in the latter theorem clarifies that every inclusion-free indexed family is *e-EQ*–identifiable with respect to every class preserving hypothesis space. Consequently, by Assertion (1) of Theorem 4 we immediately

arrive at the following corollary.

Corollary 5. *e-AEQ = e-EQ = e-CEQ.*

3.3. Absolute Learning by Erasing

Within this subsection, we study the power as well as the limitations of absolute learning for the remaining learning models.

Theorem 6.

(1) *For all LT* $\in \{ARB,\ SUB,\ SUPER\}$, *e-ALT = e-EQ*,

(2) *FIN* $\subset$ *e-AALL* $\subset$ *e-EQ*.

Proof. First, we prove Assertion (1). By Definitions 4 and 5 one immediately obtains $e\text{-}AEQ \subseteq e\text{-}ASUB \subseteq e\text{-}AARB$ as well as $e\text{-}AEQ \subseteq e\text{-}ASUPER \subseteq e\text{-}AARB$. Since $e\text{-}AEQ = e\text{-}EQ$ (cf. Corollary 5), it suffices to show that $e\text{-}AARB \subseteq e\text{-}EQ$. This can be done as follows.

Claim A. *Let* $\mathcal{L}$ *be an indexed family. If* $\mathcal{L} \in e\text{-}AARB$, *then* $\mathcal{L}$ *is inclusion-free.*

Suppose the converse, i.e., $\mathcal{L} \in e\text{-}AARB$, but $\mathcal{L}$ is not inclusion-free. Hence, there are L, $\hat{L} \in range(\mathcal{L})$ with $L \subset \hat{L}$. Now, choose any class preserving hypothesis space ψ for $\mathcal{L}$ such that $L(\psi_0) = \hat{L}$ and $L(\psi_j) \neq \hat{L}$, for all $j > 0$. Clearly, such hypothesis space always exists.

By assumption, $\mathcal{L} \in e\text{-}AARB$, and therefore there is an ELM M which $e\text{-}ARB$–identifies $\mathcal{L}$ with respect to ψ. Now, let t be any text for L. Since M $e\text{-}ARB$–identifies L from t, there has to be an $x \in \mathbb{N}$ such that $M(t_x) = 0$. Otherwise, M would stabilize on t to 0, but $L(\psi_0) \neq L$. Now, fix the least x with $M(t_x) = 0$, and choose any text $\hat{t}$ for $\hat{L} \supset L$ that has the initial segment t_x. However, M, when successively fed $\hat{t}$, deletes the one and only ψ–index for $\hat{L}$. Thus, M does not stabilize on $\hat{t}$ to a correct guess for $\hat{L}$. Since $\hat{L} \in range(\mathcal{L})$, this contradicts our assumption that M $e\text{-}ARB$–identifies $\mathcal{L}$ with respect to ψ, and Claim A follows.

Within the proof of Theorem 4 (cf. Claim B) we have already shown that every inclusion-free indexed family belongs to *e-EQ*. Hence, $e\text{-}AARB \subseteq e\text{-}EQ$, and Assertion (1) follows.

Next, we prove Assertion (2). Since *FIN* $\subset$ *e-AALL* (cf. Freivalds and Zeugmann (1995), Theorems 1 and 13), it remains to show that $e\text{-}AALL \subset e\text{-}EQ$. Clearly, $e\text{-}AALL \subseteq e\text{-}AARB$. Combining this with $e\text{-}AARB = e\text{-}EQ$ (cf. Assertion (1)), we immediately see that the only missing part is the separation of *e-EQ* and *e-AALL*.

Claim C. $e\text{-}EQ \setminus e\text{-}AALL \neq \emptyset$.

We define the desired indexed family $\mathcal{L}$ as follows. For all $k \in \mathbb{N}$, let $L_{2k} = \{2^k, 2^{k+\Phi_k(k)} + 1\}$ and $L_{2k+1} = \{2^k, 2^{k+\Phi_k(k)} + 3\}$. After a bit of reflection one easily verifies that $\mathcal{L}$ is indeed an indexed family. Moreover, $\mathcal{L}$ is inclusion-free, and therefore $\mathcal{L} \in e\text{-}EQ$ (cf. Theorem 4, Claim B).

Next, let us verify that $\mathcal{L} \notin e\text{-}AALL$. For that purpose, we choose the hypothesis space ψ with $L(\psi_k) = L_k$, for all $k \in \mathbb{N}$, and show that $\mathcal{L} \notin e\text{-}ALL$ with respect to ψ.

Since the halting problem is undecidable, the latter statement follows by contraposition of the following lemma.

Lemma. *If there exists an ELM M which e-ALL–identifies $\mathcal{L}$ with respect to ψ, then one can effectively construct an Algorithm $\mathcal{A}$ deciding for all $k \in \mathbb{N}$ whether or not $\varphi_k(k)$ converges.*

Let M be any ELM that witnesses $\mathcal{L} \in e\text{-}ALL$ with respect to ψ. We define the desired algorithm $\mathcal{A}$. On input $k \in \mathbb{N}$, execute the following instructions:

(A1) For $x = 0,\ 1,\ 2,\ \ldots$ generate successively the text $t = 2^k,\ 2^k,\ 2^k,\ \ldots$ until $(\alpha 1)$ or $(\alpha 2)$ happens.

  $(\alpha 1)$ $\Phi_k(k) = x$ has been verified.

  $(\alpha 2)$ $M(t_x) = 2k$ or $M(t_x) = 2k+1$ has been observed.

(A2) In case $(\alpha 1)$ happens first, output '$\varphi_k(k)$ converges' and stop.
Otherwise output '$\varphi_k(k)$ diverges' and stop.

Obviously, Instructions (A1) and (A2) are effectively executable. Next, we argue that $\mathcal{A}$ has to terminate for all $k \in \mathbb{N}$. Clearly, if $\varphi_k(k)$ is defined, then there is some $x \in \mathbb{N}$ with $x = \Phi_k(k)$, and $\mathcal{A}$ stops. Now, suppose that $\varphi_k(k)$ is undefined. Then t is a text for the language $L = \{2^k\} \in \mathcal{L}$. By assumption, M $e\text{-}ALL$–identifies L when fed t. Hence, M has eventually to delete all but one ψ–index, and $(\alpha 2)$ must happen. Thus, $\mathcal{A}$ terminates for all $k \in \mathbb{N}$.

Finally, we verify $\mathcal{A}$'s correctness. Clearly, if $\mathcal{A}$ outputs '$\varphi_k(k)$ converges', then $(\alpha 1)$ has happened. Thus, $\Phi_k(k) = x$ for some $x \in \mathbb{N}$, and hence $\varphi_k(k)$ is defined. Now, suppose that $\mathcal{A}$ outputs '$\varphi_k(k)$ diverges,' but $\varphi_k(k)$ is defined. Let $y \in \mathbb{N}$ be such that $y = \Phi_k(k)$. Without loss of generality, we may, additionally, assume that M, when fed t_x for some $x < y$, was striking off the ψ–index $2k$, i.e., $M(t_x) = 2k$. By definition of ψ, we know that $L_{2k} = \{2^k, 2^{k+y}+1\}$. Moreover, $2k$ is the one and only ψ–index that is correct for L_{2k}. Clearly, that immediately implies that M cannot $e\text{-}ALL$–identify L_{2k} when fed L_{2k}'s text $\hat{t} = t_x \cdot 2^{k+y}+1,\ 2^{k+y}+1,\ \ldots$, a contradiction.

Consequently, $\mathcal{A}$ solves the halting problem, and the lemma follows. Hence, Assertion (2) is proved. q.e.d.

The following corollary summarizes the results obtained.

Corollary 7.

(1) *For all $LT \in \{ARB,\ SUPER,\ ALL\}$, $e\text{-}ALT \subset e\text{-}LT = e\text{-}CLT$,*

(2) *$e\text{-}ASUB \subset e\text{-}SUB \subset e\text{-}CSUB$.*

Figure 1 displays the achieved separations and coincidences of the learning by erasing models and the ordinary learning types defined. Each learning type is represented as a vertex in a directed graph. A directed edge (or path) from vertex A to vertex B indicates that A is a proper subset of B, and no edge (or path) between these vertices imply that A and B are incomparable. Finally, LT stands for *ARB, SUB, EQ* and *SUPER*, respectively.

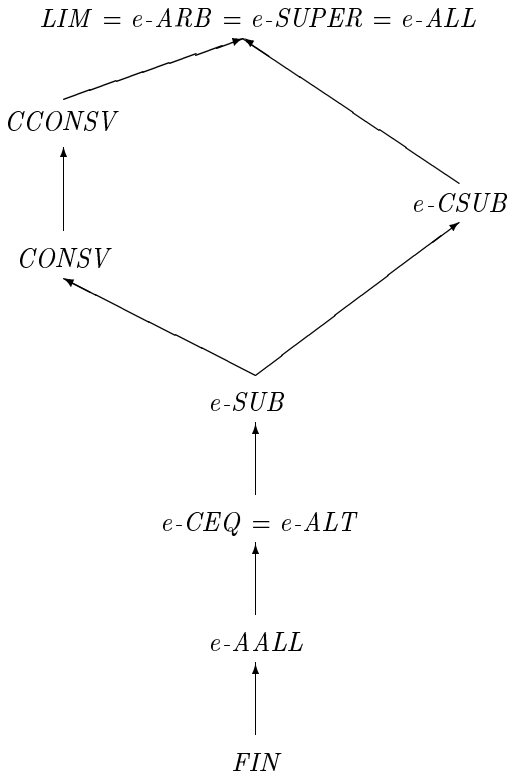

Figure 1. ***Learning by erasing versus ordinary inference***

4. Learning from Informant

In this section we study learning by erasing from both positive and negative data. Thus, we have to introduce some more notations and definitions. Let L be a language, and let $i = (s_0, b_0), (s_1, b_1), \ldots$ be an infinite sequence of elements of $\mathbb{N} \times \{+, -\}$ such that $content(i) = \{s_k \mid k \in \mathbb{N}\} = \mathbb{N}$, $i^+ = \{s_k \mid (s_k, b_k) = (s_k, +), k \in \mathbb{N}\} = L$ and $i^- = \{s_k \mid (s_k, b_k) = (s_k, -), k \in \mathbb{N}\} = co\text{-}L$. Then we refer to i as an ***informant*** for L. If L is classified via an informant then we also say that L is represented by ***positive and negative data***. By $info(L)$ we denote the set of all informants for L. We use i_x to denote the initial segment of i of length $x + 1$, and define $i_x^+ = \{s_k \mid (s_k, +) \in i, k \leq x\}$ and $i_x^- = \{s_k \mid (s_k, -) \in i, k \leq x\}$. Furthermore, $CLIM.INF$ and $FIN.INF$ are defined analogously as in Definitions 1 and 2, respectively by replacing everywhere text by informant. Finally, we extend all definitions of learning by erasing in the same way, and denote the resulting learning types by $e\text{-}CLT.INF$ for all $LT \in \{ARB, SUB, EQ, SUPER, MIN\}$.

Freivalds, Karpinski and Smith (1994) originally introduced the learning types $e\text{-}ALL.INF$ and implicitely $e\text{-}AALL.INF$, and referred to them as to *co–learning* (abbr. *co-FIN*). Furthermore, they considered the co–learnability of arbitrary recursively enumerable classes of total recursive functions. This contrasts our scenario, since we exclusively study the learnability of $\{0, 1\}$ valued functions. Nevertheless, their results easily translate into our setting. The following proposition displays the results obtained.

Proposition 4. (Freivalds, Karpinski and Smith, 1994)

FIN.INF $\subseteq$ *e-AALL.INF* $\subset$ *e-ALL.INF* = *LIM.INF*.

Taking into account that *CLIM.INF* = *ALIM.INF*, one easily verifies *e-ALL.INF* = *e-CALL.INF*. Moreover, Freivalds, Gobleja *et al.* (1994) could improve Proposition 4 to *FIN.INF* $\subset$ *co-FIN* by using a deep result by Selivanov (1976). Note, however, that the separating function class is *not* $\{0,1\}$ valued. The latter result directly raises two questions. First, which indexed families belong to *e-AALL.INF*, and second, whether or not *e-AALL.INF* $\setminus$ *FIN.INF* $\neq \emptyset$, too.

The first question has been completely answered by Kummer (1995) as the next proposition shows.

Proposition 5. (Kummer, 1995)

Let $\mathcal{L}$ be any indexed family. Then $\mathcal{L} \in$ e-AALL.INF if and only if every class preserving hypothesis space for $\mathcal{L}$ has a recursive equality problem.

Moreover, Kummer (1995) proved that every indexed family $\mathcal{L} \in$ *e-AALL.INF* must be discrete. An indexed family $\mathcal{L} = (L(\psi_j))_{j\in\mathbb{N}}$ is said to be ***discrete*** iff for every $k \in \mathbb{N}$, there is a finite function $\delta_k \subseteq \psi_k$ such that for all $j \in \mathbb{N}$, if $\delta_k \subseteq \psi_j$ then $\psi_k = \psi_j$. We refer to δ_k as to a ***separating function*** for ψ_k. An indexed family $\mathcal{L} = (L(\psi_j))_{j\in\mathbb{N}}$ is said to be ***effectively discrete*** if there exists an algorithm computing for every $k \in \mathbb{N}$ a separating function δ_k for ψ_k.

Our next theorem completely answers the second question posed above. Again, the proof is based on the Selivanov's (1976) result already used by Freivalds, Gobleja *et al.* (1994).

Theorem 8. *e-AALL* $\setminus$ *FIN.INF* $\neq \emptyset$.

Proof. Selivanov (1976) showed that there is a recursively enumerable class $\mathcal{U}_{se}$ of total recursive functions fulfilling the following requirements:

(1) every numbering $\tau \in \mathcal{R}^2$ for $\mathcal{U}_{se}$ has a recursive equivalence problem,

(2) $\mathcal{U}_{se}$ is not effectively discrete.

Since $\mathcal{U}_{se}$ is not $\{0,1\}$ valued, some transformation of it is in order. Using $\mathcal{U}_{se}$ we define an indexed family $\mathcal{L}_{se}$ that is well-suited to separate *e-AALL* and *FIN.INF*.

Let $\tau \in \mathcal{R}^2$ be any numbering for $\mathcal{U}_{se}$. For all $j, x, y \in \mathbb{N}$ we set:

$$\psi_j(\langle x, y\rangle) = \begin{cases} 1, & \text{if } \tau_j(x) = y, \\ 0, & \text{otherwise.} \end{cases}$$

Finally, set $\mathcal{L}_{se} = (L(\psi_j))_{j\in\mathbb{N}}$. Clearly, $\mathcal{L}_{se}$ is an indexed family.

Claim A. $\mathcal{L}_{se} \in$ *e-AALL*.

Applying our characterization of *e-AALL* (cf. Theorem 18) it suffices to show that $\mathcal{L}_{se}$ is inclusion-free and, furthermore, every class preserving hypothesis space for $\mathcal{L}_{se}$ has a recursive equality problem.

This can be verified as follows. Let $j \in \mathbb{N}$. Since τ is a numbering of total recursive functions, we may easily conclude that, for every $x \in \mathbb{N}$, there is exactly one $y \in \mathbb{N}$ with $\langle x, y\rangle \in L(\psi_j)$. Thus, one immediately sees that $\mathcal{L}_{se}$ is inclusion-free.

Now, assume any class preserving hypothesis space $\hat{\psi}$ for $\mathcal{L}_{se}$. For all $j,\ x \in \mathbb{N}$, set $\hat{\tau}(j,x) = y$, where y is the uniquely determined number with $\langle x,y\rangle \in L(\hat{\psi}_j)$. Obviously, $\hat{\tau} \in \mathcal{R}^2$. Since $\hat{\psi}$ is class preserving for $\mathcal{L}_{se}$, $\hat{\tau}$ is a numbering for $\mathcal{U}_{se}$. Clearly, $\hat{\tau}_j = \hat{\tau}_k$ implies $L(\hat{\psi}_j) = L(\hat{\psi}_k)$. By Property (1) $\hat{\tau}$ has a recursive equality problem, and thus we are done.

Claim B. $\mathcal{L}_{se} \notin$ *FIN.INF*.

Suppose the converse, i.e., $\mathcal{L}_{se} \in$ *FIN.INF*. Since finite inference is invariant with respect to choice of the hypothesis space (cf. Lange and Zeugmann (1994)), we may assume that there is an IIM M witnessing $\mathcal{L}_{se} \in$ *FIN.INF* with respect to the hypothesis space ψ defined above.

Given M, we define an algorithm $\mathcal{A}$ that assigns to every τ_k a separating function δ_k. $\mathcal{A}$ is defined as follows. On input $k \in \mathbb{N}$, execute the following instructions:

($A1$) For $z = 0,\ 1,\ 2,\ \ldots$ generate successively the lexicographically ordered informant i^k for $L(\psi_k)$ until M outputs a guess, say on input $i^k_{\hat{z}}$.

($A2$) Set $\delta_k = \{(x,\tau_k(x)) \mid x \leq \hat{z}\}$, and stop.

Since M *FIN.INF*–identifies $\mathcal{L}_{se}$, we may conclude that Instruction ($A1$) terminates for every $k \in \mathbb{N}$, and thus $\mathcal{A}$ is recursive. It suffice to show that, for all $j \in \mathbb{N}$, $\delta_k \subseteq \tau_j$ implies $\tau_k = \tau_j$.

Suppose any $j \in \mathbb{N}$ with $\delta_k \subseteq \tau_j$. Clearly, $\tau_k(x) = \tau_j(x)$ for all $x \leq \hat{z}$, and therefore $\psi_k(\langle x,y\rangle) = \psi_k(\langle x,y\rangle)$ for all $y \in \mathbb{N}$ and all $x \leq \hat{z}$. Thus, $i^k_{\hat{z}}$ is an initial segment of the lexicographically ordered informant i^j for $L(\psi_j)$. By Definition 2, M, when successively fed i^k and i^j, respectively, is only allowed to generate a single, but correct hypothesis. Since M, when fed $i^k_{\hat{z}} = i^j_{\hat{z}}$, has ouput its one and only hypothesis, we obtain $M(i^k_{\hat{z}}) = M(i^j_{\hat{z}})$, and hence $L(\psi_k) = L(\psi_j)$. Consequently, $\tau_k = \tau_j$, too.

Thus, $\mathcal{U}_{se}$ is effectively discrete, a contradiction. q.e.d.

Thus, it remains to clarify the relations with respect to inclusion between the remaining learning by erasing models. This is done by the following theorem.

Theorem 9. *For all* $LT \in \{ARB,\ SUB,\ EQ,\ SUPER,\ MIN\}$ *we have*

e-*ALT.INF* $= e$-*LT.INF* $= e$-*CLT.INF* $=$ *LIM.INF*.

Proof. First, we prove that every indexed family belongs to e-*AEQ.INF* and e-*AMIN.INF*, respectively.

Claim A. $\mathcal{L} \in e$-*AEQ.INF for every indexed family* $\mathcal{L}$.

Let $\psi \in \mathcal{R}^2_{0,1}$ be any class preserving hypothesis space for $\mathcal{L}$. The desired ELM M can be defined as follows. Let $L \in \mathit{range}(\mathcal{L})$, let $i \in \mathit{info}(L)$, and let $x \in \mathbb{N}$.

***ELM M*:** "On input i_x proceed as follows:
If $x = 0$, then set $\mathit{Alive}_0 = \mathbb{N}$, output nothing and request the next input.
Otherwise, test for all $k \in \mathit{Alive}_{x-1}$ with $k \leq x$, whether or not $i^+_x \not\subseteq L(\psi_k)$ or $i^-_x \not\subseteq$ *co*-$L(\psi_k)$. If there is at least one index passing this test, output the minimal one, say $\hat{k}$, update $\mathit{Alive}_x = \mathit{Alive}_{x-1} \setminus \{\hat{k}\}$, and request the next input. Otherwise, set $\mathit{Alive}_x = \mathit{Alive}_{x-1}$ output nothing and request the next input."

By construction, M never outputs a correct ψ–index for L. It remains to argue, that M eventually deletes all indices k with $L(\psi_k) \neq L$. This can be shown similarly as in the proof of Theorem 4, Claim B. Note that $L(\psi_k) \neq L$ implies $i_x^+ \not\subseteq L(\psi_k)$ or $i_x^- \not\subseteq \textit{co-}L(\psi_k)$ for almost all $x \in \mathbb{N}$. We omit the details.

Claim B. $\mathcal{L} \in \textit{e-AMIN.INF}$ *for every indexed family* $\mathcal{L}$.

Let $\psi \in \mathcal{R}_{0,1}^2$ be any class preserving hypothesis space for $\mathcal{L}$. The desired ELM M can be defined as follows. Let $L \in \textit{range}(\mathcal{L})$, let $i \in \textit{info}(L)$, and let $x \in \mathbb{N}$. Initialize $\textit{Guess}_0 = 0$. On input i_x, $x > 0$, M behaves as follows. Let $k = \textit{Guess}_{x-1}$. If $i_x^+ \subseteq L(\psi_k)$ and $i_x^- \subseteq \textit{co-}L(\psi_k)$, it outputs nothing, sets $\textit{Guess}_x = \textit{Guess}_{x-1}$, and requests the next input. Otherwise, it updates $\textit{Guess}_x = k + 1$, and outputs k.

One directly verifies that M exactly outputs all and only the indices j less than the least ψ–index for L. We omit the details.

By definition, $\textit{e-AEQ.INF} \subseteq \textit{e-ALT.INF}$ for all $LT \in \{\textit{ARB}, \textit{SUB}, \textit{SUPER}\}$, and thus, by Claim A, every indexed family is contained in $\textit{e-ALT.INF}$, too. On the other hand, every indexed family belongs to $\textit{LIM.INF}$ (cf. Gold (1967)). Finally, taking into account that $\textit{e-ALT.INF} \subseteq \textit{e-LT.INF} \subseteq \textit{e-CLT.INF}$ for any learning type $LT \in \{\textit{ARB}, \textit{SUB}, \textit{EQ}, \textit{SUPER}, \textit{MIN}\}$ the theorem directly follows. q.e.d.

So far we have studied separately learning from text and learning from informant. Now we focus our attention to another interesting aspect, namely the interplay between information presentation and learnability constraints. The first known result along this line of research relates finite learning from informant to conservative inference from text.

Proposition 6. (Lange and Zeugmann, 1993a)

$\textit{FIN.INF} \subset \textit{CONSV}$.

Since $\textit{FIN.INF} \subseteq \textit{e-AALL.INF}$, the question arises whether or not Proposition 6 generalizes to $\textit{e-AALL.INF} \subset \textit{CONSV}$ or at least to $\textit{e-AALL.INF} \subset \textit{LIM}$.

Our next result establishes a nice consequence of Kummer's (1995) characterization of $\textit{e-AALL.INF}$. We prove that discreteness implies $\textit{LIM}$-learnability from text.

Theorem 10. *Let* $\mathcal{L}$ *be any indexed family. If* $\mathcal{L}$ *is discrete, then* $\mathcal{L} \in \textit{LIM}$.

Proof. Let $\mathcal{L}$ be an indexed family that is discrete, and let ψ be any class preserving hypothesis space for $\mathcal{L}$ having a recursive equality problem. Note that such a hypothesis space always exists as shown in Lange, Nessel and Wiehagen (1996) (cf. proof of Theorem 1). Based on ψ, we assign to every $L(\psi_j)$ a recursively enumerable set T_j, a so-called "tell–tale," that satisfies the following requirements:

(1) $T_j \subseteq L(\psi_j)$,

(2) T_j is finite,

(3) for all $k \in \mathbb{N}$, $T_j \subseteq L(\psi_k)$ implies $L(\psi_k) \not\subset L(\psi_j)$.

Thus, applying Angluin's characterization of $\textit{LIM}$ (cf. Angluin (1980), Theorem 1) we may conclude that $\mathcal{L}$ is $\textit{LIM}$–identifiable with respect to ψ.

It remains to construct the sets T_j. For all $j \in \mathbb{N}$, we set $T_j = \bigcup_{z \in \mathbb{N}} T_j^{(z)}$, and define the corresponding subsets $T_j^{(z)}$ as follows. Let $j,\ z \in \mathbb{N}$, we set:

$$T_j^{(z)} = \begin{cases} \emptyset, & \text{if } \psi_j = \psi_k, \\ \{x \mid x \leq n,\ x \in L(\psi_j)\}, & \text{otherwise, where } n = \min\{x \mid \psi_j(x) \neq \psi_z(x)\}. \end{cases}$$

Since ψ is a hypothesis space that has a recursive equality problem, one easily verifies that $T_j^{(z)}$ is finite and recursive. By definition $T_j^{(z)} \subseteq L(\psi_j)$ and therefore T_j is a recursively enumerable subset of $L(\psi_j)$.

Next, we verify (2). Let $j \in \mathbb{N}$ be arbitrarily fixed. Since $\mathcal{L}$ is discrete and ψ is class preserving for $\mathcal{L}$, there has to be a separating function δ_j for $L(\psi_j)$. Let $\hat{n} = \max(Arg(\delta_j))$. Hence, $T_j^{(z)} \subseteq \{x \mid x \leq \hat{n},\ x \in L(\psi_j)\}$ for all $z \in \mathbb{N}$, and thus T_j is indeed finite.

Finally, we show (3). Suppose the converse, i.e., there are $j,\ k \in \mathbb{N}$ such that $T_j \subseteq L(\psi_k)$ and $L(\psi_k) \subset L(\psi_j)$. Clearly, $\psi_j \neq \psi_k$. Let $n_k = \min\{x \mid \psi_j(x) \neq \psi_k(x)\}$. Since $L(\psi_k) \subset L(\psi_j)$, we obtain $n_k \notin L(\psi_k)$ and $n_k \in L(\psi_j)$. By $T_j^{(k)}$'s definition $n_k \in T_j^{(k)}$, and therefore $n_k \in T_j$, too. However, $n_k \notin L(\psi_k)$, and thus $T_j \not\subseteq L(\psi_k)$, a contradiction. q.e.d.

Corollary 11. *e-AALL.INF* $\subset$ *LIM*.

Proof. Let $\mathcal{L} \in$ *e-AALL.INF*. Hence, $\mathcal{L}$ is discrete (cf. Kummer (1995), Theorem 10 and Fact 5). Thus, $\mathcal{L} \in$ *LIM* follows by Theorem 10. On the other hand, let $\mathcal{L}_{fin}$ denote the indexed family canonically enumerating all finite sets of natural numbers. Obviously, $\mathcal{L}_{fin}$ is not discrete, and thus, $\mathcal{L} \notin$ *e-AALL.INF* (cf. Kummer (1995), Lemma 9). Finally, $\mathcal{L}_{fin} \in$ *LIM*, and the corollary follows. q.e.d.

By Corollary 3 we may easily conclude:

Corollary 12. *For all LT* $\in \{$*ARB*, *SUPER*, *ALL*$\}$, *e-AALL.INF* $\subset$ *e-LT*.

The following theorem enables us to clarify the relation between the remaining models of learning by erasing from text and informant, respectively.

Theorem 13.

(1) *FIN.INF* $\setminus$ *e-CSUB* $\neq \emptyset$,

(2) *e-EQ* $\setminus$ *e-AALL.INF* $\neq \emptyset$.

Proof. For verifying Assertion (1) recall the definition of the indexed family $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$ used in the proof of Theorem 1, i.e., $L_0 = \mathbb{N}$ and $L_{j+1} = \{j\}$. Obviously, $\mathcal{L}$ is *FIN.INF*–identifiable, and since $\mathcal{L} \notin$ *e-CSUB* (cf. Theorem 4, Claim C), Assertion (1) follows.

Next, we show Assertion (2). For that purpose, we choose the indexed family $\mathcal{L}$ introduced in the proof of Theorem 6, Claim C, i.e., $L_{2k} = \{2^k, 2^{k+\Phi_k(k)} + 1\}$ and $L_{2k+1} = \{2^k, 2^{k+\Phi_k(k)} + 3\}$ for all $k \in \mathbb{N}$. We already know that $\mathcal{L} \in$ *e-EQ*.

It remains to show that $\mathcal{L} \notin$ *e-AALL.INF*. Let ψ be the hypothesis space defined by $L(\psi_k) = L_k$ for all $k \in \mathbb{N}$. We show $\mathcal{L} \notin$ *e-AALL.INF* by reducing the halting

problem to $\mathcal{L} \in e\text{-}AALL.INF$. If $\mathcal{L} \in e\text{-}AALL.INF$ with respect to ψ would hold, then, by Proposition 5, ψ has a recursive equality problem. Consequently, $\varphi_k(k)$ is defined iff $L_{2k} \neq L_{2k+1}$, and the halting problem would be recursive. Hence, $\mathcal{L} \notin e\text{-}AALL.INF$. q.e.d.

Taking into account that $LIM \subset LIM.INF$ (cf. Gold (1967)), we directly arrive at the following corollary displaying the consequences of the latter theorem.

Corollary 14. *For all* $LT \in \{ARB, SUB, EQ, SUPER, ALL\}$ *and for all* $\lambda \in \{A, \varepsilon, C\}$, *we have* $e\text{-}\lambda LT \subset e\text{-}\lambda LT.INF$.

Putting Theorem 13 together with Corollary 5 we can easily conclude:

Corollary 15.

(1) $e\text{-}AEQ \# e\text{-}AALL.INF$,

(2) $e\text{-}SUB \# e\text{-}AALL.INF$,

(3) $e\text{-}CSUB \# e\text{-}AALL.INF$,

(4) $e\text{-}AALL \# FIN.INF$.

Figure 2 summarizes the estabilshed relations of learning by erasing from text and informant, respectively. The semantics of Figure 2 is analogous to that of Figure 1. Again LT stands for ARB, $SUPER$, SUB and EQ, respectively.

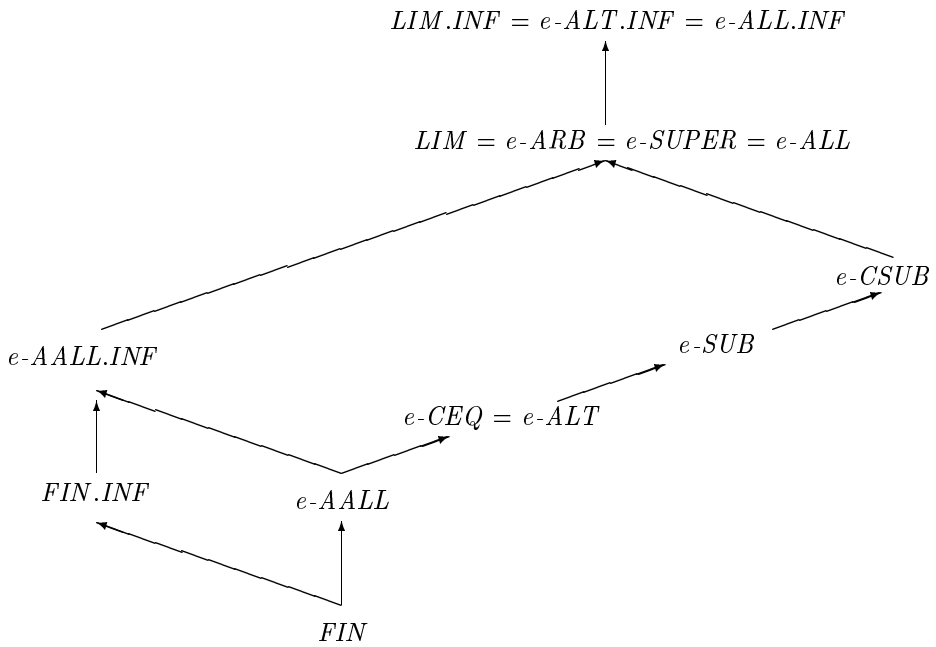

Figure 2. ***Relations between learning by erasing from text and informant***

5. Characterizations

In this section we present characterizations of all the learning by erasing models. These characterizations may help to gain a better understanding of what the defined learning models have in *common* and what their *differences* are. Our first result characterizes *e-ARB*, *e-EQ*, *e-SUB* and *e-SUPER* in purely topological terms.

Theorem 16. *Let $\mathcal{L}$ be any indexed family. $\mathcal{L} \in$ e-EQ if and only if $\mathcal{L}$ is inclusion-free.*

Proof. Necessity: Let $\mathcal{L} \in$ *e-EQ*. Hence, $\mathcal{L} \in$ *e-CEQ*, and therefore $\mathcal{L}$ is inclusion-free (cf. Theorem 4, Claim A).

Sufficiency: This part has been already shown within the proof of Theorem 4 (cf. Claim B), and the theorem follows. q.e.d.

Taking Corollary 5 and Theorem 6, Assertion (1) into consideration we may easily conclude:

Corollary 17. *Let LT $\in$ {ARB, SUB, EQ, SUPER}, and let $\mathcal{L}$ be an indexed family. $\mathcal{L} \in$ e-ALT if and only if $\mathcal{C}$ is inclusion-free.*

For characterizing *e-AALL* we had to combine the topological approach with the numbering theoretical one used by Kummer (1995).

Theorem 18. *Let $\mathcal{L}$ be any indexed family. $\mathcal{L} \in$ e-AALL if and only if*

(1) *$\mathcal{L}$ is inclusion-free, and*

(2) *every class preserving hypothesis space for $\mathcal{L}$ has a recursive equality problem.*

Proof. Necessity: Let $\mathcal{L} \in$ *e-AALL*. By definition $\mathcal{L} \in$ *e-AARB*, and thus $\mathcal{L}$ is inclusion-free (cf. Corollary 17). On the other hand, *e-AALL* $\subset$ *e-AALL.INF*. Hence, by Proposition 5 we obtain that every class preserving hypothesis space for $\mathcal{L}$ has a recursive equality problem.

Sufficiency: Let $\mathcal{L}$ be any inclusion-free indexed family, and let ψ be any class preserving hypothesis space for $\mathcal{L}$ having a recursive equality problem. Thus, there is a $p \in \mathcal{R}^2_{0,1}$ such that for all $j,\ k \in \mathbb{N}$, $p(j,k) = 1$ iff $\psi_j = \psi_k$. We design an ELM M witnessing $\mathcal{L} \in$ *e-ALL* with respect to ψ. Let $L \in \mathit{range}(L)$, $t \in \mathit{text}(L)$, and $x \in \mathbb{N}$.

***ELM M*:** "On input t_x proceed as follows:
If $x = 0$, then set $\mathit{ToErase}_0 = \emptyset$, output nothing and request the next input.
Otherwise, test whether or not $\mathit{ToErase}_{x-1} = \emptyset$.
In case it is, execute Instruction ($A1$). Otherwise, goto ($A2$).

($A1$) Determine the least index k with $t_x^+ \subseteq L(\psi_k)$, and fix $z = \max\{x, k\}$. Set $\mathit{ToErase}_x = \{j \mid j \leq z,\ t_x^+ \not\subseteq L(\psi_j)\} \cup \{j \mid k < j \leq z,\ p(k,j) = 1\}$. Output nothing, and request the next input.

($A2$) Determine $j = \min(\mathit{ToErase}_{x-1})$. Update $\mathit{ToErase}_x = \mathit{ToErase}_{x-1} \setminus \{j\}$, output j, and request the next input."

Since $t \in \mathit{text}(L)$ for some $L \in \mathit{range}(\mathcal{L})$, the unbounded search performed within Instruction $(A1)$ terminates for every $x \in \mathbb{N}$, and thus, M is an ELM. Let $\hat{k} = \mathit{min}_{\psi}(L)$. We show that M eventually outputs all natural numbers but $\hat{k}$.

Claim A. $\hat{k} \notin \mathit{ToErase}_x$ *for all* $x \in \mathbb{N}$.

Suppose the converse. Hence, there exists a least $x \in \mathbb{N}$ such that $\hat{k} \in \mathit{ToErase}_x$. By definition, M includes $\hat{k}$ into $\mathit{ToErase}_x$ iff either $t_x^+ \not\subseteq L(\psi_{\hat{k}})$ or an index $k < \hat{k}$ has been found that meets $p(k, \hat{k}) = 1$. Since $L(\psi_{\hat{k}}) = L$, $t_x^+ \not\subseteq L(\psi_{\hat{k}})$ cannot be observed. Moreover, $p(k, \hat{k}) = 1$ implies $L(\psi_k) = L$ contradicting that $\hat{k}$ is the least ψ–index for L. The claim follows.

Since M is exclusively outputting numbers $j \in \mathit{ToErase}_x$ for some $x \in \mathbb{N}$, the index $\hat{k}$ is never deleted. It remains to show that M eventually outputs all ψ–indices that are different from $\hat{k}$.

Claim B. *M, when successively fed t, outputs all* $k \in \mathbb{N} \setminus \{\hat{k}\}$.

We distinguish the following cases.

Case 1. $L(\psi_k) \neq L$.

Since ψ is class preserving and $\mathcal{L}$ is inclusion-free, we know that $L \setminus L(\psi_k) \neq \emptyset$. Because of $t \in \mathit{text}(L)$, there must be a minimal y such that $t_{y+\ell}^+ \not\subseteq L(\psi_k)$ for all $\ell \in \mathbb{N}$. Thus, if $\mathit{ToErase}_{y-1} = \emptyset$ then $k \in \mathit{ToErase}_y$. Otherwise, there must be an $r \in \mathbb{N}$ such that $k \in \mathit{ToErase}_{y+r}$. Consequently, k is output eventually.

Case 2. $L(\psi_k) = L$.

Since $\hat{k}$ is the least ψ–index of L, the inclusion-freeness of $\mathcal{L}$ implies $L(\psi_{\hat{k}}) \setminus L(\psi_j) \neq \emptyset$ for all $j < \hat{k}$. Hence, there exists a $y \in \mathbb{N}$ such that $t_y^+ \not\subseteq L(\psi_j)$ for all $j < \hat{k}$. Therefore, for all $x \geq y$ the unbounded search in Instruction (A1) terminates at $\hat{k}$. Moreover, since $L(\psi_k) = L$ we have $p(\hat{k}, k) = 1$. Consequently, there must be an $x \geq \max\{y, k\}$ such that $k \in \mathit{ToErase}_x$. Consequently, k is again eventually output.

Thus, Claim B follows, and the theorem is proved. q.e.d.

Next, we characterize *e-CSUB* and *e-SUB*. Now we derive necessary and sufficient conditions for any given pair of an indexed family and a hypothesis space for it. Again, the characterization is mainly based on the topological properties of the relevant hypothesis spaces. However, we had to add a recursive component to these topological properties. Within the next definition we provide the necessary framework for establishing the desired characterization theorems.

Definition 6. *Let* $\mathcal{L}$ *be any indexed family, and let* ψ *be any class comprising hypothesis space for* $\mathcal{L}$. *Then we set:*

(1) $\mathit{Bad}(\mathcal{L}, \psi) = \{j \mid L \subset L(\psi_j)$, and $j < \mathit{min}_{\psi}(L)$ for some $L \in \mathit{range}(\mathcal{L})\}$,

(2) $\mathit{Comp}(\mathcal{L}, \psi) = \{j \mid L(\psi_j) \notin \mathit{range}(\mathcal{L})\}$.

Theorem 19. *Let* $\mathcal{L}$ *be an indexed family.* $\mathcal{L} \in$ *e-CSUB if and only if there are a class comprising hypothesis space* ψ *for* $\mathcal{L}$ *and a recursively enumerable set* W *such that* $\mathit{Bad}(\mathcal{L}, \psi) \subseteq W \subseteq \mathit{Comp}(\mathcal{L}, \psi)$.

Proof. Necessity: Let $\mathcal{L} \in$ *e-CSUB*. Hence, there are a class comprising hypothesis space $\hat{\psi}$ for $\mathcal{L}$ and an ELM $\hat{M}$ that witnesses $\mathcal{L} \in$ *e-CSUB* with respect to $\hat{\psi}$.

Next, we use M to define $f \in \mathcal{P}$ such that $W = \mathit{range}(f)$. For every $k \in \mathbb{N}$, let t^k denote the canonical text for the language $L(\psi_k)$. For every $k,\ x \in \mathbb{N}$, we set:

$$f(\langle k, x\rangle) = \begin{cases} M(t_x^k), & \text{if } \mathit{content}(t_x^k) \subseteq L(\psi_{M(t_x^k)}), \\ \text{not defined}, & \text{otherwise.} \end{cases}$$

Using the convention that, if M on input t_x^k does not output any hypothesis then $f(\langle k, x\rangle)$ is also not defined, we obviously have $f \in \mathcal{P}$. It remains to show that $\mathit{Bad}(\mathcal{L}, \psi) \subseteq W \subseteq \mathit{Comp}(\mathcal{L}, \psi)$.

Claim A. $W \subseteq \mathit{Comp}(\mathcal{L}, \psi)$.

If $W = \emptyset$, we are done. Now, let $z = f(\langle k, x\rangle)$ for some $k,\ x \in \mathbb{N}$. By definition of f, we have $M(t_x^k) = z$ and $\mathit{content}(t_x^k) \subseteq L(\psi_z)$. Suppose, $L(\psi_z) \in \mathit{range}(\mathcal{L})$. Since $\mathit{content}(t_x^k) \subseteq L(\psi_z)$, t_x^k is an initial segment of some text $\hat{t}$ for $L(\psi_z)$. Thus M, when fed the text $\hat{t}$ for $L(\psi_z) \in \mathit{range}(\mathcal{L})$, outputs the correct ψ–index z for $L(\psi_z)$. This contradicts our assumption that M *e-CSUB*–infers $\mathcal{L}$ with respect to ψ. Thus, Claim A follows.

Claim B. $\mathit{Bad}(\mathcal{L}, \psi) \subseteq W$.

Suppose the converse, i.e., there is a $z \in \mathit{Bad}(\mathcal{L}, \psi) \setminus W$. Hence, $z < \mathit{min}_\psi(L)$ for some $L \in \mathit{range}(\mathcal{L})$ with $L \subset L(\psi_z)$. We distinguish the following cases.

Case 1. $L(\psi_z) \in \mathit{range}(\mathcal{L})$

Consider M when fed any text t for L. Because of $z < \mathit{min}_\psi(L)$, M eventually outputs z, say on input t_x, since otherwise it would stabilize on t to some $z' \leq z$ with $L(\psi_{z'}) \neq L$. However, since $L \subset L(\psi_z)$, the initial segment t_x may be extended to a text for $L(\psi_z)$ on which M outputs z. This is a contradiction to M *e-CSUB*–identifies $\mathcal{L}$ with respect to ψ.

Case 2. $L(\psi_z) \notin \mathit{range}(\mathcal{L})$

Let k be any ψ–index for L. Consider M when fed the canonical text t^k for L. Since M *e-CSUB*–identifies L from text, M must stabilize on t^k to $\mathit{min}_\psi(L)$. Because of $z < \mathit{min}_\psi(L)$, there has to be an $x \in \mathbb{N}$ with $M(t_x^k) = z$. Thus, $f(\langle k, x\rangle) = z$, and hence $z \in W$, again a contradiction.

Claim B follows, and we are done.

Sufficiency: Let $\mathcal{L}$ be any indexed family, let ψ be a class comprising hypothesis space for $\mathcal{L}$, and let W be a recursively enumerable set with $\mathit{Bad}(\mathcal{L}, \psi) \subseteq W \subseteq \mathit{Comp}(\mathcal{L}, \psi)$. Let $\ell \in \mathbb{N}$ be such that $W = \mathit{range}(\varphi_\ell)$, and let W^z be the elements of W, if any, enumerated after z steps of computation of φ_ℓ. We define an ELM M that witnesses $\mathcal{L} \in \mathit{e\text{-}CSUB}$ with respect to ψ. So, let $L \in \mathit{range}(\mathcal{L})$, $t \in \mathit{text}(L)$, and let $x \in \mathbb{N}$.

***ELM M*:** "On input t_x proceed as follows:

If $x = 0$, initialize $\mathit{ToErase}_0 = \emptyset$. Output nothing and request the next input. Otherwise, test whether or not $\mathit{ToErase}_{x-1} = \emptyset$. In case it is, goto $(A1)$. Otherwise, goto $(A2)$.

$(A1)$ Determine the least index k that satisfies both $t_x^+ \subseteq L(\psi_k)$ and $k \notin W^x$. Update $\mathit{ToErase}_x = \{j \mid j < k\}$, output nothing, and request the next input.

(*A*2) Determine $j = \min(\mathit{ToErase}_{x-1})$. Update $\mathit{ToErase}_x = \mathit{ToErase}_{x-1} \setminus \{j\}$, output j, and request the next input."

Let $\hat{k} = \mathit{min}_\psi(L)$. Clearly, $\hat{k} \notin W$ and $t_x^+ \subseteq L(\psi_{\hat{k}})$ for all $x \in \mathbb{N}$. Hence, the unbounded search performed within Instruction (*A*1) terminates for all $x \in \mathbb{N}$. Furthermore, the same arguments imply that M never outputs $\hat{k}$.

It remains to show that M outputs all numbers $k < \hat{k}$. This can be done inductively. Let $0 < k < \hat{k}$, and suppose that y has been selected in a way such that M has already output all $j < k$ when fed t_y. We distinguish the following cases.

Case 1. $L \setminus L(\psi_k) \neq \emptyset$.

Hence, there is a least $x > y$ such that $t_x^+ \not\subseteq L(\psi_k)$. By M's definition there has to be a $z \geq x$ such that M, when fed t_z, starts the execution of Instruction (*A*1), and includes k into $\mathit{ToErase}_z$. Hence, M eventually outputs k.

Case 2. $L \subset L(\psi_k)$.

Now, we may conclude that $k \in \mathit{Bad}(\mathcal{L}, \psi)$, and therefore $k \in W$, too. Consequently, there is some $x \in \mathbb{N}$ such that $k \in W^x$. Now, select $z \geq \max\{y, x\}$ in a way such that M, when fed t_z, executes Instruction (*A*1). Consequently, k appears in $\mathit{ToErase}_z$, and will therefore be output in some subsequent step.

Clearly, the above argumentation applies *mutatis mutandis* to handle the induction base, i.e., the $k = 0$ case. Hence, M stabilizes on t to the least ψ–index $\hat{k}$ for L, thereby outputting only indices $k < \hat{k}$. q.e.d.

Theorem 20. *Let $\mathcal{L}$ be any indexed family. $\mathcal{L} \in$ e-SUB if and only if there is a class preserving hypothesis space ψ for $\mathcal{L}$ such that $\mathit{Bad}(\mathcal{L}, \psi) = \emptyset$.*

Proof. Necessity: Let $\mathcal{L}$ be any *e-SUB*–learnable indexed family. Thus, there are an ELM M and a class preserving hypothesis space ψ for $\mathcal{L}$ such that M *e-SUB*–infers $\mathcal{L}$ with respect to ψ. Suppose, $\mathit{Bad}(\mathcal{L}, \psi) \neq \emptyset$. Hence, there exist a language $L \in \mathcal{L}$ and an index $j < \mathit{min}_\psi(L)$ such that $L \subset L(\psi_j)$. Now, when fed any text $t \in \mathit{text}(L)$, M has to output j sometime, say on t_x. Again, t_x can be extended to a text $\hat{t} \in \mathit{text}(L(\psi_j))$ on which M also outputs j and hence a correct number for $L(\psi_j)$, a contradiction.

Sufficiency: Let $\mathcal{L}$ be any indexed family possessing a class preserving hypothesis space ψ with $\mathit{Bad}(\mathcal{L}, \psi) = \emptyset$. Then $\mathcal{L}$ can be *e-SUB*–inferred with respect to ψ by an ELM M working as follows. Let $L \in \mathcal{L}$, $t \in \mathit{text}(L)$, and let $x \in \mathbb{N}$. Initialize $S_0 = \emptyset$. On input t_x M behaves as follows. It searches the least k such that $t_x^+ \subseteq L(\psi_k)$. Then, it computes $E = \{1, \ldots, k-1\} \setminus S_{x-1}$. If $E \neq \emptyset$, it outputs $j = \min E$, updates $S_x = S_{x-1} \cup \{j\}$, and requests the next input. Otherwise, it outputs nothing, sets $S_x = S_{x-1}$, and requests the next input.

One directly verifies that M exactly outputs the indices $j < \mathit{min}_\psi(L)$. q.e.d.

A closer look at the characterizations above shows that the ELMs constructed in the sufficiency part share a nice property. Namely, given any text t for any language L in the target class, the ELMs precisely erase all hypotheses less than the least correct one for L. Using this insight, we directly obtain all the remaining characterizations for learning by erasing.

Theorem 21. *For all* $\lambda \in \{\varepsilon,\ A,\ C\}$, $e\text{-}\lambda MIN = e\text{-}\lambda SUB$.

Proof. By definition, $e\text{-}\lambda MIN \subseteq e\text{-}\lambda SUB$ for all $\lambda \in \{\varepsilon,\ A,\ C\}$. Furthermore, $e\text{-}SUB \subseteq e\text{-}MIN$ and $e\text{-}CSUB \subseteq e\text{-}MIN$ has been shown within the proof of the sufficiency part of Theorems 20 and 19, respectively.

It remains to verify that $e\text{-}ASUB \subseteq e\text{-}AMIN$. However, this inclusion follows immediately from the proof of the sufficiency part in Theorem 20. q.e.d.

To sum up, since $e\text{-}EQ = e\text{-}ASUB = e\text{-}AMIN$, we obtain the missing characterizations for the learning types $e\text{-}AMIN$, $e\text{-}MIN$ and $e\text{-}CMIN$ by Theorems 16, 20 and 19, respectively.

6. Conclusions and Open Problems

Different models of learning by erasing have been defined, and the learning power of all the resulting learning types has been related to one another as well as to learning in the limit, conservative identification, and finite inference. As it turned out, all but the $e\text{-}EQ$ learning model are sensitive with respect to the particular choice of the hypothesis space, thus nicely contrasting learning in the limit and finite learning. Moreover, the learning power of the $e\text{-}SUB$ model is even very dependent on the set of admissible hypothesis spaces.

A further interesting aspect is provided by Theorem 1, Corollaries 2 and 7. That is, these results show that the process of elimination cannot be restricted to *incorrect* hypotheses for achieving its full learning power. On the other hand, all learning by erasing models that are allowed to erase *correct* hypotheses, too, are as powerful as learning in the limit provided the hypothesis space is appropriately chosen (cf. Theorem 1). Consequently, in order to decide whether or not a particular indexed family can be $e\text{-}LT$-learned, $LT \in \{ARB,\ SUPER,\ ALL\}$, one can apply any of the known criteria for LIM-inferability (cf., e.g., Angluin (1980), Sato and Umayahara (1992)).

These differences almost vanish if *absolute* learning is considered. Now, we have a somehow opposite effect. Erasing all but one guess turns out to be most restrictive with respect to the resulting learning capabilities.

The phenomena described above find their natural explanation in our characterization theorems. All models $e\text{-}ALT$ of absolute learning by erasing are constraint by the topological properties of the indexed families to be learned, i.e., they must be inclusion-free ($LT \in \{ARB,\ SUB,\ EQ,\ SUPER,\ ALL\}$), and in case of $e\text{-}AALL$, additionally, all hypothesis spaces must be equivalent with respect to reducibility, i.e., they must have a recursive equality problem.

Moreover, in Section 4 we studied the problem whether or not information presentation may be traded versus learnability. The results obtained put the strength of $e\text{-}AALL.INF$ learning into the right perspective as displayed in Figure 2, pp. 21. However, it remained open whether or not $e\text{-}AALL.INF \subset LIM$ can be strengthened to $e\text{-}AALL.INF \subset CCONSV$. Note that Theorem 10 cannot be sharpened to discreteness implies conservative learnability, since the index family $\mathcal{L}$ defined in the proof of Theorem 4, Claim D is discrete but $\mathcal{L} \notin CCONSV$ (cf. Lange and Zeugmann (1993a)).

Finally, we want to point to a further possible line of research. In our opinion, it may also be sufficiently interesting to investigate the *complexity* of learning by erasing. This includes the comparison of the complexity of both the different models of learning by erasing as well as of learning by erasing with standard learning. As a result of the first type we have the following comparison of the complexity of hypothesis spaces for e-ALL-learning and e-ARB-learning. That is, there is an infinite indexed family $\mathcal{L}$ such that:

(1) for every $\psi \in \mathcal{R}^2_{0,1}$ such that $\mathcal{L} \in e\text{-}ALL$ with respect to ψ all but one language $L \in range(\mathcal{L})$ must have infinitely many ψ-numbers,

(2) there exists $\psi \in \mathcal{R}^2_{0,1}$ such that $\mathcal{L} \in e\text{-}ARB$ with respect to ψ and every language from $\mathcal{L}$ has exactly one ψ-number.

This can be easily verified using the indexed family $\mathcal{L}$ defined in the proof of Theorem 1, Claim B, thus Property (2) follows. Property (1) is an immediate consequence of Theorem 9 in Freivalds and Zeugmann (1995).

7. References

Angluin, D. (1980), Inductive inference of formal languages from positive data, *Information and Control* **45**, 117 – 135.

Baliga, G., Case, J., and Jain, S. (1996), Synthesizing enumeration techniques for language learning, *eCOLT*, eC-TR-96-003.

Blum, M. (1967), A machine independent theory of the complexity of recursive functions, *Journal of the Association for Computing Machinery* **14**, 322 – 336.

Freivalds, R., Karpinski, M., and Smith, C.H. (1994), Co-learning of total recursive functions, *in* "Proceedings 7th Annual ACM Conference on Computational Learning Theory," New Brunswick, July 1994, pp. 190 – 197, ACM Press, New York.

Freivalds, R., Gobleja, D., Karpinski, M., and Smith, C.H. (1994), Co-learnability and FIN-identifiability of enumerable classes of total recursive functions, *in* "Proceedings 4th International Workshop on Analogical and Inductive Inference, AII'94," (S. Arikawa and K.P. Jantke, Eds.), Lecture Notes in Artificial Intelligence Vol. 872, pp. 100 – 105, Springer-Verlag, Berlin.

Freivalds, R., and Zeugmann, T. (1995), Co–learning of recursive languages from positive data, RIFIS Technical Report RIFIS-TR-CS-110, RIFIS, Kyushu University 33, 1995.

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

Jain, S., and Sharma, A. (1994), Characterizing language learning by standardizing operations, *Journal of Computer and System Sciences* **49**, 96 – 107.

Kapur, S., and Bilardi, G. (1995), Language learning without overgeneralization, *Theoretical Computer Science* **141**, 151 – 162.

Kummer, M. (1995), A learning-theoretic characterization of classes of recursive functions, *Information Processing Letters* **54**, 205 – 211.

Lange, R., Nessel, J., and Wiehagen, R. (1996), Learning recursive languages from good examples, *submitted to Annals of Mathematics and Artificial Intelligence.*

Lange, S., and Zeugmann, T. (1992), Types of monotonic language learning and their characterization, *in* "Proceedings 5th Annual ACM Workshop on Computational Learning Theory," July 1992, Pittsburgh, (D. Haussler, Ed.), pp. 377 – 390, ACM Press, New York.

Lange, S., and Zeugmann, T. (1993a), Monotonic versus non-monotonic language learning, *in* "Proceedings 2nd International Workshop on Nonmonotonic and Inductive Logic," December 1991, Reinhardsbrunn, (G. Brewka, K.P. Jantke and P.H. Schmitt, Eds.), Lecture Notes in Artificial Intelligence Vol. 659, pp. 254 – 269, Springer-Verlag, Berlin.

Lange, S., and Zeugmann, T. (1993b), Language learning in dependence on the space of hypotheses, *in* "Proceedings 6th Annual ACM Conference on Computational Learning Theory," July 1993, Santa Cruz, (L. Pitt, Ed.), pp. 127 – 136, ACM Press, New York.

Lange, S., and Zeugmann, T. (1993c), Learning recursive languages with bounded mind changes, *International Journal of Foundations of Computer Science* **4**, 157 – 178.

Lange, S., and Zeugmann, T. (1994), Characterization of language learning from informant under various monotonicity constraints, *Journal of Experimental & Theoretical Artificial Intelligence* **6**, 73 – 94.

Osherson, D., Stob, M., and Weinstein, S. (1986), "Systems that Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists," MIT Press, Cambridge, Massachusetts.

Rogers, H.Jr. (1967), "Theory of Recursive Functions and Effective Computability", McGraw–Hill, New York.

Sato, M., and Umayahara, K. (1992), Inductive inferability for formal languages from positive data, *IEICE Transactions on Information and Systems* **E–75D**, 415 – 419.

Selivanov, V.L. (1976), Enumerations of families of general recursive functions, *Algebra i Logika* **15**, 205 – 226; English Translation, *Algebra and Logic* **15**, 128 – 141.

Trakhtenbrot, B.A., and Barzdin, J. (1970) "Конечные Автоматы (Поведение и Синтез)", Наука, Москва,
English translation: "Finite Automata – Behavior and Synthesis, Fundamental Studies in Computer Science 1", North-Holland, Amsterdam, 1973.

Wiehagen, R. (1978), Characterization problems in the theory of inductive inference, *in* "Proceedings Automata, Languages and Programming", International Colloquium, Udine, (C. Böhm, Ed.), Lecture Notes in Computer Science Vol. 62, pp. 494 – 504, Springer-Verlag, Berlin.

Zeugmann, T., Lange, S., and Kapur, S. (1995), Characterizations of monotonic and dual monotonic language learning, *Information and Computation* **120**, 155 – 173.