Set-Driven and Rearrangement-Independent Learning of Recursive Languages

Steffen Lange*
HTWK Leipzig
FB Mathematik und Informatik
PF 66
04275 Leipzig, Germany
steffen@informatik.th-leipzig.de

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

Abstract

The present paper deals with the learnability of indexed families of uniformly recursive languages from positive data under various postulates of naturalness. In particular, we consider set-driven and rearrangement-independent learners, i.e., learning devices whose output exclusively depends on the range and on the range and length of their input, respectively. The impact of set-drivenness and rearrangement-independence on the behavior of learners to their learning power is studied in dependence on the *hypothesis space* the learners may use. Furthermore, we consider the influence of set-drivenness and rearrangement-independence for learning devices that realize the *subset principle* to different extents. Thereby we distinguish between strong-monotonic, monotonic and weak-monotonic or conservative learning.

The results obtained are twofold. First, rearrangement-independent learning does not constitute a restriction except the case of monotonic learning. Second, we prove that for all but one of the considered learning models set-drivenness is a severe restriction. However, set-driven *conservative* learning is exactly as powerful as unrestricted *conservative* learning provided the *hypothesis space* is appropriately chosen. These results considerably extend previous work done in the field (cf. e.g. Schäfer-Richter (1984) and Fulk (1990)).

*Most of this work has been done while the author was visiting ISIS FUJITSU Labs. at Numazu, Japan. It has been partially supported by the German Ministry for Research and Technology (BMFT) under grant no. 01 IW 101.

1. Introduction

Gold-style formal language learning (cf. Gold (1967)) has attracted a lot of attention during the last decades (cf. e.g. Osherson, Stob and Weinstein (1986) and the references therein). The general situation underlying Gold's model can be described as follows: Given more and more eventually incomplete information concerning a language to be learned, an inference device (an IIM, for short) has to produce, from time to time, a hypothesis about the phenomenon to be inferred. The sequence of hypotheses has to converge to a hypothesis correctly describing the language to be learned. Consequently, the inference process is an ongoing one. Within in the present paper we study exclusively language learning from *positive examples* or, synonymously, from *text*, i.e., exactly all strings belonging to the language which should be recognized will be successively presented. The set of all admissible hypotheses is called *space of hypotheses* or *hypothesis space*, for short.

In this paper we investigate the learning capabilities of learners that *simultaneously* fulfill *various combinations* of desirable properties. A central question directly arising when dealing with Gold's-model of learning in the limit is whether or not the *order* of information presentation does really influence the capabilities of IIMs. We distinguish between two degrees of order-independence. An IIM is said to be *set-driven*, if its output does only depend on the *range* of its input. Schäfer-Richter (1984) and Fulk (1990) proved that set-driven IIMs are less powerful than unrestricted ones. A natural weakening of set-drivenness is rearrangement-independence. An IIM is called *rearrangement-independent* if its output does only depend on the *range* and *length* of its input. As it turned out, any collection of languages that can be learned in the limit may also be learned by a rearrangement-independent IIM (cf. Schäfer-Richter (1984), Fulk (1990)). However, the weakness of set-driven IIMs has been proved in a setting allowing self-referential arguments. This might lead to the impression that this result is far beyond any practical relevance, since self-referential arguments are mainly applicable in settings where the *membership problem* for languages is undecidable in general.

Therefore, we study the power of set-driven and rearrangement-independent IIMs in a more realistic setting with respect to potential applications, i.e., we deal exclusively with indexed families of non-empty and uniformly recursive languages. An indexed family is a recursive enumeration of non-empty languages such that membership is uniformly decidable (cf. Angluin (1980)).

A major problem, one has to deal with when learning from text, is to avoid or to detect *overgeneralization*, i.e., hypotheses that describe proper *supersets* of the target language. The impact of this problem results simply from the fact that a text cannot supply counterexamples to such hypotheses. IIMs that strictly avoid overgeneralized hypotheses are called *conservative* (cf. Definition 6). Several authors proposed the so-called *subset principle* to solve the problem of avoiding overgeneralization (cf. e.g. Berwick (1985), Wexler (1992)). Informally, the subset principle requires the learner to hypothesize the "least" language from the hypothesis space with respect to set inclusion that fits with the data the IIM has read so far. In Lange and Zeugmann (1993a) different notions of monotonic language learning has been introduced. All these notions of monotonicity may be considered as formalizations of learning realizing the subset principle to different extents. Moreover, the power of all the monotonic learning models heavily depends on the choice of the hypothesis space (cf. Lange and Zeugmann (1993b)).

In the sequel we study the impact of set-drivenness and rearrangement-independence on all the models of monotonic learning in dependence on the hypothesis space. The results obtained prove that rearrangement-independent learning does not constitute a restriction in most cases. Note that neither Schäfer-Richter's (1984) nor Fulk's (1990) transformation of an arbitrary IIM into a rearrangement-independent one preserves conservativeness or any other constraint implementing the subset principle. Furthermore, we show that set-drivenness cannot be achieved in general. However, conservative learning is exactly as powerful as set-driven conservative inference, if one may carefully choose a hypotheses space that contains a description for every target language, and, additionally, grammars that do not represent languages contained in the target family of languages to be learned. We regard this result as a particular answer to the question how a "natural" learning algorithm may be designed.

2. Preliminaries

By $\mathbb{N} = \{0, 1, 2, ...\}$ we denote the set of all natural numbers. We set $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. Let $\varphi_0,\ \varphi_1,\ \varphi_2, ...$ denote any fixed ***programming system*** of all (and only all) partial recursive functions over $\mathbb{N}$, and let $\Phi_0,\ \Phi_1,\ \Phi_2, ...$ be any associated ***complexity measure*** (cf. Machtey and Young (1978)). Then φ_k is the partial recursive function computed by program k in the programming system. Furthermore, let $k, x \in \mathbb{N}$. If $\varphi_k(x)$ is defined (abbr. $\varphi_k(x)\downarrow$) then we also say that $\varphi_k(x)$ converges; otherwise, $\varphi_k(x)$ diverges (abbr. $\varphi_k(x)\uparrow$). By $\langle .,. \rangle : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ we denote ***Cantor's pairing function*** i.e., $\langle x, y \rangle = ((x+y)^2 + 3x + y)/2$ for all $x, y \in \mathbb{N}$.

In the sequel we assume familiarity with formal language theory (cf. Hopcroft and Ullman (1969)). By Σ we denote any fixed finite alphabet of symbols. Let Σ^* be the free monoid over Σ. Any subset $L \subseteq \Sigma^*$ is called a language. By $co-L$ we denote the complement of L. Let L be a language and $t = s_0, s_1, s_2, ...$ an infinite sequence of strings from Σ^* such that $range(t) = \{s_k \,|\, k \in \mathbb{N}\} = L$. Then t is said to be a ***text*** for L or, synonymously, a ***positive presentation***. Let L be a language. By $text(L)$ we denote the set of all positive presentations of L. Moreover, let t be a text and let x be a number. Then, t_x denotes the initial segment of t of length $x+1$, and $t_x^+ =_{df} \{s_k \,|\, k \leq x\}$.

Next, we introduce the notion of the ***canonical text*** that turned out to be very helpful in proving several theorems. Let L be any non-empty recursive language, and let $s_0, s_1, s_2, ...$ be the lexicographically ordered text of Σ^*. The canonical text of L is obtained as follows. Test sequentially whether $s_z \in L$ for $z = 0, 1, 2, ...$ until the first z is found such that $s_z \in L$. Since $L \neq \emptyset$ there must be at least one z fulfilling the test. Set $t_0 = s_z$. We proceed inductively. For all $x \in \mathbb{N}$ we define:

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

In the sequel we deal with the learnability of indexed families of uniformly recursive languages defined as follows (cf. Angluin (1980)). A sequence $L_0, L_1, L_2, \ldots$ is said to be an ***indexed family*** $\mathcal{L}$ of uniformly recursive languages provided all L_j are non-empty and there is a recursive function f such that for all numbers j and all strings $s \in \Sigma^*$ we have

$$f(j,s) = \begin{cases} 1, & \text{if} \quad s \in L_j, \\ 0, & \text{otherwise.} \end{cases}$$

In the following we refer to indexed families of uniformly recursive languages as indexed families for short. Moreover, we often denote an indexed family and its range by the same symbol $\mathcal{L}$. The meaning will be clear from the context.

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 larger and larger initial segments of a text t and it either requests the next input string, or it first outputs a hypothesis, i.e., a number encoding a certain computer program, and then it requests the next input string.

At this point we specify the semantics of the hypotheses an IIM outputs. For that purpose we have to clarify what hypothesis spaces we choose. We require the inductive inference machines to output indices of grammars, since this learning goal fits well with the intuitive idea of language learning. Furthermore, since we exclusively deal with indexed families $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$ we always take as space of hypotheses an enumerable family of grammars $G_0, G_1, G_2, \ldots$ over the terminal alphabet Σ satisfying $\mathcal{L} \subseteq \{L(G_j) \mid j \in \mathbb{N}\}$. Moreover, we require that membership in $L(G_j)$ is uniformly decidable for all $j \in \mathbb{N}$ and all strings $s \in \Sigma^*$. When an IIM outputs a number j, we interpret it to mean that the machine is hypothesizing the grammar G_j. Moreover, for notational convenience we use $\mathcal{L}(\mathcal{G})$ to denote $\{L(G_j) \mid j \in \mathbb{N}\}$ for every hypothesis space $\mathcal{G} = (G_j)_{j \in \mathbb{N}}$.

Let t be a text, and $x \in \mathbb{N}$. Then we use $M(t_x)$ to denote the last hypothesis produced by M when successively fed t_x. The sequence $(M(t_x))_{x \in \mathbb{N}}$ is said to ***converge in the limit*** to the number j if and only if either $(M(t_x))_{x \in \mathbb{N}}$ is infinite and all but finitely many terms of it are equal to j, or $(M(t_x))_{x \in \mathbb{N}}$ is non-empty and finite, and its last term is j. Now we define some concepts of learning. We start with learning in the limit.

Definition 1. **(Gold (1967))** *Let $\mathcal{L}$ be an indexed family, $L \in \mathcal{L}$, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}}$ be a hypothesis space.* ***An IIM M CLIM–identifies L from text with respect to $\mathcal{G}$*** *iff for every text t for L, there exists a $j \in \mathbb{N}$ such that the sequence $(M(t_x))_{x \in \mathbb{N}}$ converges in the limit to j and $L = L(G_j)$.*

Furthermore, M $CLIM$–identifies $\mathcal{L}$ with respect to $\mathcal{G}$ if and only if, for each $L \in \mathcal{L}$, M $CLIM$–identifies L from text with respect to $\mathcal{G}$.

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

Suppose, an IIM identifies some language L. That means, after having seen only finitely many data of L the IIM reached its (unknown) point of convergence and it computed a *correct* and *finite* description of a generator for the target language. Hence, some form of learning must have taken place. Therefore, we use the terms *infer* and *learn* as synonyms for identify.

In the above Definition 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 hypothesis space comprising $range(\mathcal{L})$. The restriction of $CLIM$ to ***class preserving*** inference is denoted by LIM. That means LIM is the collection of all indexed families $\mathcal{L}$ that can be learned in the limit with respect to a hypothesis space $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ such that $range(\mathcal{L}) = \{L(G_j)\,|\,j \in \mathbb{N}\}$. Moreover, if a target indexed family $\mathcal{L}$ has to be inferred with respect to the hypothesis space $\mathcal{L}$ itself, then we replace the prefix C by E, i.e., $ELIM$ is the collection of indexed families that can be ***exactly*** learned in the limit. Finally, we adopt this convention in defining all the learning types below.

Moreover, an IIM is required to learn the target language from every text for it. This might lead to the impression that an IIM mainly extracts the range of the information fed to it, thereby neglecting the length and order of the data sequence it reads. IIMs really behaving thus are called set-driven. More precisely, we define:

Definition 2. **(Wexler and Culicover, Sec. 2.2 (1980))** ***An IIM is said to be set-driven*** *iff its output depends only on the range of its input; that is, iff* $M(t_x) = M(\hat{t}_y)$ *for all* $x, y \in \mathbb{N}$, *all texts* t, $\hat{t}$ *provided* $t_x^+ = \hat{t}_y^+$.

Schäfer-Richter (1984) as well as Fulk (1990), later, and independently proved that set-driven IIMs are less powerful than unrestricted ones. Fulk (1990) interpreted the weakening in the learning power of set-driven IIMs by the need of IIMs for time to "reflect" on the input. However, this time cannot be bounded by any a priorily fixed computable function depending exclusively on the size of the range of the input, since otherwise set-drivenness would not restrict the learning power. Indeed, Osherson, Stob and Weinstein (1986) proved that any *non-recursive* IIM M may be replaced by a *non-recursive* set-driven IIM $\hat{M}$ learning at least as much as M does. With the next definition we consider a natural weakening of Definition 2.

Definition 3. **(Schäfer-Richter (1984), Osherson et al. (1986))** ***An IIM is said to be rearrangement-independent*** *iff its output depends only on the range and on the length of its input; that is, iff* $M(t_x) = M(\hat{t}_x)$ *for all* $x \in \mathbb{N}$, *all texts* t, $\hat{t}$ *provided* $t_x^+ = \hat{t}_x^+$.

We make the following convention. For all the learning models in this paper we use the prefix s-, and r- to denote the learning model restricted to set-driven and rearrangement-independent IIMs, respectively. For example, $s-LIM$ denotes the collection of all indexed families that are LIM–inferable by some set-driven IIM. Next we formalize the other inference models that we have mentioned in the introduction.

Definition 4. **(Gold (1967))** *Let* $\mathcal{L}$ *be an indexed family,* $L \in \mathcal{L}$, *and let* $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ *be a hypothesis space.* ***An IIM M CFIN–identifies* L *from text*** *iff for every text* t *for* L, *there exists a* $j \in \mathbb{N}$ *such that* M, *when successively fed* t, *outputs the single hypothesis* j, $L = L(G_j)$, *and stops thereafter.*

Furthermore, M $CFIN$*–identifies* $\mathcal{L}$ *with respect to* $\mathcal{G}$ *if and only if, for each* $L \in \mathcal{L}$, M $CFIN$*–identifies* L *from text with respect to* $\mathcal{G}$.

The resulting learning type is denoted by $CFIN$.

Consequently, every hypothesis produced by a finitely working IIM has to be a correct guess.

The next definition formalizes the different notions of monotonicity.

Definition 5. (Jantke (1991), Wiehagen (1991)) *Let $\mathcal{L}$ be an indexed family of languages, $L \in \mathcal{L}$ and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}}$ be a space of hypotheses.* ***An IIM M is said to identify a language*** **L** ***from text with respect to*** $\mathcal{G}$

(A) ***strong-monotonically***

(B) ***monotonically***

(C) ***weak-monotonically***

iff

M LIM–identifies L from text with respect to $\mathcal{G}$ and for any text $t \in text(L)$ as well as for any two consecutive hypotheses j_x, j_{x+k} which M has produced when fed t_x and t_{x+k} where $k \in \mathbb{N}^+$ the following conditions are satisfied:

(A) $L(G_{j_x}) \subseteq L(G_{j_{x+k}})$

(B) $L(G_{j_x}) \cap L \subseteq L(G_{j_{x+k}}) \cap L$

(C) *if* $t^+_{x+k} \subseteq L(G_{j_x})$ *then* $L(G_{j_x}) \subseteq L(G_{j_{x+k}})$.

By $CSMON$, $CMON$, and $CWMON$, we denote the set of all indexed families $\mathcal{L}$ for which there is an IIM M and a hypothesis space $\mathcal{G}$ such that M infers $\mathcal{L}$ strong-monotonically, monotonically, and weak-monotonically, respectively, with respect to the hypothesis space $\mathcal{G}$.

Definition 6. (Angluin (1980)) *Let $\mathcal{L}$ be an indexed family, $L \in \mathcal{L}$, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}}$ be a space of hypotheses.* ***An IIM M CCONSERVATIVE–identifies*** **L** ***from text with respect to*** $\mathcal{G}$ *iff*

(1) *M CLIM–identifies L from text with respect to $\mathcal{G}$,*

(2) *for every text t the following condition is satisfied:*
if M on input t_x makes the guess j_x and then outputs the hypothesis $j_{x+k} \neq j_x$ at some subsequent step, then $t^+_{x+k} \not\subseteq L(G_{j_x})$.

Finally, M CCONSERVATIVE–identifies $\mathcal{L}$ with respect to $\mathcal{G}$ if and only if, for each $L \in \mathcal{L}$, M CCONSERVATIVE–identifies L from text with respect to $\mathcal{G}$.

The collection of sets $CCONSERVATIVE$ is defined in an analogous manner as above. Note that $\lambda WMON = \lambda CONSERVATIVE$ for all $\lambda \in \{C, \epsilon, E\}$, where ϵ denotes the empty string (cf. Lange and Zeugmann (1993b)).

3. Learning with Set-driven IIMs.

In this section we study the question under what circumstances set-drivennes does restrict the power of the learning models defined above. We start with finite learning. The next theorem in particular states that finite learning is invariant with respect to the specific choice of the hypothesis space. Moreover, for every hypothesis space comprising the target indexed family $\mathcal{L}$ there is a *set-driven* IIM that finitely learns $\mathcal{L}$.

Theorem 1. $EFIN = FIN = CFIN = s{-}EFIN$

As we have already mentioned, the examples of Schäfer-Richter (1984) and Fulk (1990) witnessing the restriction of set-driven learners are not indexed families. Hence, we ask whether the uniform recursiveness of all target languages may compensate the impact to learn with set-driven IIMs. The answer is no as the following theorem impressively shows.

Theorem 2. $s{-}CLIM \subset ELIM = LIM = CLIM$

Proof. The part $ELIM = LIM = CLIM$ is due to Lange and Zeugmann (1993b). It remains to show that $s{-}CLIM \subset ELIM$.

The desired indexed family $\mathcal{L}$ is defined as follows. For all $k \in \mathbb{N}$ we set $L_{\langle k,0\rangle} = \{a^k b^n \,|\, n \in \mathbb{N}^+\}$. For all $k \in \mathbb{N}$ and all $j \in \mathbb{N}^+$ 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. $\Phi_k(k) \leq j$

Let $d = 2 \cdot \Phi_k(k) - j$. Now, we set:

$$L_{\langle k,j\rangle} = \begin{cases} \{a^k b^m \,|\, 1 \leq m \leq d\}, & \text{if } d \geq 1, \\ \{a^k b\}, & \text{otherwise.} \end{cases}$$

$\mathcal{L} = (L_{\langle k,j\rangle})_{j,k\in\mathbb{N}}$ is an indexed family of recursive languages, since the predicate "$\Phi_i(y) \leq z$" is uniformly decidable in i, y, and z.

Claim A. $\mathcal{L} \notin s{-}CLIM$

Since the halting problem is undecidable, Claim A follows by contraposition of the following Claim B.

Claim B. If there exists an IIM M witnessing $\mathcal{L} \in s$–$CLIM$, then one can effectively construct an algorithm deciding for all $k \in \mathbb{N}$ whether or not $\varphi_k(k)$ converges.

Let M be any IIM that learns $\mathcal{L}$ in the limit with respect to some hypothesis space $\mathcal{G}$ comprising $\mathcal{L}$. We define an algorithm $\mathcal{A}$ that solves the halting problem.

Algorithm $\mathcal{A}$: "On input k execute (A1) and (A2).

(A1) For $z = 0, 1, 2, \ldots$ generate successively the canonical text t of $L_{\langle k,0\rangle}$ until M on input t_z outputs for the first time a hypothesis j such that $t_z^+ \cup \{a^k b^{z+2}\} \subseteq L(G_j)$.

(A2) Test whether $\Phi_k(k) \leq z+1$. In case it is, output "$\varphi_k(k)$ converges." Otherwise output "$\varphi_k(k)$ diverges."

Since M has to infer $L_{\langle k,0\rangle}$ in particular from t, there has to be a least z such that M on input t_z computes a hypothesis j satisfying $t_z^+ \cup \{a^k b^{z+2}\} \subseteq L(G_j)$. Moreover, the test whether or not $t_z^+ \cup \{a^k b^{z+2}\} \subseteq L(G_j)$ can be effectively performed, since membership in $L(G_j)$ is uniformly decidable. By the definition of a complexity measure, instruction (A2) is effectively executable. Hence, $\mathcal{A}$ is an algorithm.

It remains to show that $\varphi_k(k)$ diverges, if $\neg\ \Phi_k(k) \leq z+1$. Suppose the converse; then there exists a $y > z+1$ with $\Phi_k(k) = y$. In accordance with the definition of $\mathcal{L}$, we obtain $L = t_z^+ \in \mathcal{L}$. Hence, t_z is also an initial segment of a text $\hat{t}$ for L. Due to the definition of $\mathcal{A}$, we have $L(G_j) \neq L$. Since M is a set-driven IIM, $L = t_z^+$ implies $M(\hat{t}_{x+r}) = j$ for all $r \in \mathbb{N}$. Therefore, M fails to infer L from its text $\hat{t}$. This contradicts our assumption that M is a set-driven IIM which $CLIM$ infers $\mathcal{L}$ with respect to $\mathcal{G}$. Hence, Claim B is proved.

The remaining part $\mathcal{L} \in ELIM$ is omitted. The reader is referred to Lange and Zeugmann (1993d).

q.e.d.

As the latter theorem shows, sometimes there is no way to design a set-driven IIM. However, with the following theorems we mainly intend to show that the careful choice of the hypothesis space deserves special attention whenever set-drivenness is desired.

Theorem 3. *There is an indexed family $\mathcal{L}$ such that*

(1) $\mathcal{L} \in r\text{--}ESMON$,

(2) $\mathcal{L} \notin s\text{--}LIM$,

(3) *there is a set-driven IIM M and a hypothesis space $\mathcal{G}$ such that M $CSMON$–identifies $\mathcal{L}$ with respect to $\mathcal{G}$.*

As we have seen, set-drivenness constitutes a severe restriction. While this is true in general as long as exact and class preserving learning is considered, the situation looks differently in the class comprising case. On the one hand, learning in the limit cannot always be achieved by set-driven IIMs (cf. Theorem 2). On the other hand, conservative learners may always be designed to be set-driven, if the hypothesis space is appropriately chosen.

Theorem 4. $s\text{--}CCONSERVATIVE = CCONSERVATIVE$

Proof. We only sketch the main ideas of the proof, and refer the interested reader to Lange and Zeugmann (1993d) for any detail. The proof is partitioned into two parts. The first part establishes the equality of class comprising conservative and class comprising, rearrangement-independent conservative learning. The main ingredients into this proof are the characterization of *CCONSERVATIVE* (cf. Lange and Zeugmann (1993b)) as well as a technically simple, but powerful modification of the corresponding tell-tale family.

Let $\mathcal{L} \in CCONSERVATIVE$. Then there exists a space $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ of hypotheses and a recursively generable tell-tale family $(T_j)_{j\in\mathbb{N}}$ of finite and non-empty sets such that

(1) $range(\mathcal{L}) \subseteq \mathcal{L}(\mathcal{G})$,

(2) for all $j \in N$, $T_j \subseteq L(G_j)$,

(3) for all $j, k \in \mathbb{N}$, if $T_j \subseteq L(G_k)$, then $L(G_k) \not\subset L(G_j)$.

Using this tell-tale family, we define a new recursively generable family $(\hat{T}_j)_{j\in\mathbb{N}}$ of finite and non-empty sets that allows the design of a rearrangement-independent IIM inferring $\mathcal{L}$ conservatively with respect to $\mathcal{G}$. But surprisingly enough, we can even do better, namely, we can define an IIM witnessing $\mathcal{L}(\mathcal{G}) \in r-ECONSERVATIVE$. For all $j \in \mathbb{N}$, we set $\hat{T}_j = \bigcup_{n\leq j} T_n \cap L(G_j)$. Note that the new tell-tale family fulfills Properties (1) through (3) above.

Now, the wanted IIM can be defined as follows: Let $L \in \mathcal{L}(\mathcal{G})$, $t \in text(L)$, and $x \in N$.

$M(t_x) =$ "Generate $\hat{T}_k$ for all $k \leq x$ and test whether $\hat{T}_k \subseteq t_x^+ \subseteq L(G_k)$. In case there is one k fulfilling the test, output the minimal one, and request the next input. Otherwise, output nothing and request the next input."

Obviously, M is rearrangement-independent. We omit the proof that the IIM M *ECONSERVATIVE*-identifies $\mathcal{L}(\mathcal{G})$.

The second part of the proof establishes set-drivenness. For that purpose, we define a new hypothesis space $\tilde{\mathcal{G}} = (\tilde{G}_j)_{j\in\mathbb{N}}$ as well as a new IIM $\tilde{M}$. The basis for these definitions are the hypothesis space $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$, and the IIM M described above. The hypothesis space $\tilde{\mathcal{G}}$ is the canonical enumeration of all grammars from $\mathcal{G}$ and all finite languages over the underlying alphabet Σ. Before defining the IIM $\tilde{M}$, we introduce the notion of *repetition free* text $rf(t)$. Let $t = s_0, s_1, ...$ be any text. We set $rf(t_0) = s_0$ and proceed inductively as follows: For all $x \geq 1$, $rf(t_{x+1}) = rf(t_x)$, if $s_{x+1} \in rf(t_x)^+$, and $rf(t_{x+1}) = rf(t_x), s_{x+1}$ otherwise. Obviously, given any initial segment t_x of a text t one can effectively compute $rf(t_x)$. Now we are ready to present the definition of $\tilde{M}$. Let $L \in \mathcal{L}(\mathcal{G})$, $t \in text(L)$, and $x \in \mathbb{N}$.

$\tilde{M}(t_x) =$ "Compute $rf(t_x)$. If M on input $rf(t_x)$ outputs a hypothesis, say j, then output the canonical index of j in $\tilde{\mathcal{G}}$ and request the next input.
Otherwise, output the canonical index of t_x^+ in $\tilde{\mathcal{G}}$ and request the next input."

Intuitively, it is clear that $\tilde{M}$ is set-driven. The proof that $\tilde{M}$ conservatively infers $\mathcal{L}(\mathcal{G})$ with respect to $\tilde{\mathcal{G}}$ is omitted.

q.e.d.

The latter theorem allows a nice corollary that we present next.

Corollary 5. *Let $\mathcal{L} \in CCONSERVATIVE$. Then, there is a hypothesis space $\hat{\mathcal{G}} = (\hat{G}_j)_{j\in\mathbb{N}}$ comprising $\mathcal{L}$ such that $\mathcal{L}(\hat{\mathcal{G}}) \in s-ECONSERVATIVE$.*

Proof. Let $\mathcal{L} \in CCONSERVATIVE$. Furthermore, due to the latter theorem, there is a set-driven IIM $\tilde{M}$ and a hypothesis space $\tilde{\mathcal{G}}$ such that $\tilde{M}$ conservatively infers $\mathcal{L}$ with respect to $\tilde{\mathcal{G}}$.

Recall that $\tilde{\mathcal{G}}$ is a canonical enumeration of $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ satisfying $\mathcal{L} \subseteq \mathcal{L}(\mathcal{G})$ and of all finite languages over the underlying alphabet. Without loss of generality we may assume that $\tilde{\mathcal{G}}$ fulfills the following property. If j is even, then $L(\tilde{G}_j) \in \mathcal{L}(\mathcal{G})$. Hence, $\tilde{M}$ infers $L(\tilde{G}_j)$ from text. Otherwise, $L(\tilde{G}_j)$ is a finite language.

We start with the definition of the desired hypothesis space $\hat{\mathcal{G}} = (\hat{G}_j)_{j\in\mathbb{N}}$ If j is even, then we set $\hat{G}_j = \tilde{G}_j$. Otherwise, we distinguish the following cases. If M

when fed the lexicographically ordered enumeration of all strings in $L(\tilde{G}_j)$ outputs the hypothesis j, then we set $\hat{G}_j = \tilde{G}_j$. In case it does not, we set $\hat{G}_j = \tilde{G}_{j-1}$.

Now we are ready to define the desired IIM M which witnesses $\mathcal{L}(\hat{\mathcal{G}}) \in s-ECONSERVATIVE$. Let $L \in \mathcal{L}(\hat{\mathcal{G}})$, $t \in text(L)$, and $x \in \mathbb{N}$.

$M(t_x) =$ "Simulate $\tilde{M}$ on input t_x. If $\tilde{M}$ does not output any hypothesis, then output nothing and request the next input.
Otherwise, let $\tilde{M}(t_x) = j$. Output j and request the next input."

Since $\tilde{M}$ is a conservative and set-driven IIM, M behaves thus. It remains to show that M learns L. Obviously, if $L = L(\hat{G}_{2k})$ for some $k \in \mathbb{N}$, then $\tilde{M}$ infers L. Therefore, since M simulates $\tilde{M}$, we are done.

Now, let us suppose, $L \neq L(\hat{G}_{2k})$ for some $k \in \mathbb{N}$. By definition of $\hat{\mathcal{G}}$, we know that L is finite. Moreover, since t is a text for L, there exists an x such that $t_y^+ = L$ for all $y \geq x$. Recalling the definition of $\hat{\mathcal{G}}$, and by assumption, we obtain the following. There is a number j such that $\tilde{M}(t_x) = j$, $L = t_x^+ = L(\tilde{G}_j) = L(\hat{G}_j)$. Hence, $M(t_x) = j$, too. Finally, since M is set-driven, we directly get $M(t_y) = j$ for all $y \geq j$. Consequently, M learns L.

q.e.d.

The next theorem gives some more evidence that set-drivenness is not that restrictive as it might seem.

Theorem 6.

(1) $s-SMON \setminus EWMON \neq \emptyset$,

(2) $s-CSMON \setminus WMON \neq \emptyset$,

(3) $s-EWMON \setminus MON \neq \emptyset$.

Proof. First of all, we show Assertion (1). Let us consider the following indexed family $\mathcal{L}_{sm} = (L_{\langle k,j\rangle})_{j,k\in\mathbb{N}}$. For all $k \in \mathbb{N}$, we set $L_{\langle k,0\rangle} = \{a^k b^n \mid n \in \mathbb{N}^+\}$. For all $k \in \mathbb{N}$ and all $j \in \mathbb{N}^+$, we distinguish the following cases:

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

We set: $L_{\langle k,j\rangle} = L_{\langle k,0\rangle}$.

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

Then, we set: $L_{\langle k,j\rangle} = \{a^k b^m \mid 1 \leq m \leq \Phi_k(k)\}$.

In Lange and Zeugmann (1993b) it was already shown that the family $\mathcal{L}_{sm}$ is witnessing $SMON \setminus EWMON \neq \emptyset$. Hence, it remains to show the following claim.

Claim A. $\mathcal{L}_{sm} \in s-SMON$.

We have to show that there is a hypothesis space $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ which satisfies $range(\mathcal{L}_{sm}) = \mathcal{L}(\mathcal{G})$ and a set-driven IIM M such that M does strong-monotonically infer $\mathcal{L}$ with respect to $\mathcal{G}$.

First of all, we define the hypothesis space $\mathcal{G}$. For all $k \in \mathbb{N}$, we set $L(G_{2k}) = \bigcap_{j\in\mathbb{N}} L_{\langle k,j\rangle}$ and $L(G_{2k-1}) = L_{\langle k,0\rangle}$.

Since $\mathcal{L}_{sm}$ is an indexed family, it is easy to verify that membership is uniformly decidable for $\mathcal{G}$. Moreover, we have $range(\mathcal{L}_{sm}) = \mathcal{L}(\mathcal{G})$.

Let $L \in \mathcal{L}_{sm}$, let t be any text for L, and let $x \in \mathbb{N}$. The desired IIM M is defined as follows.

$M(t_x) =$ "Determine the unique k such that $t_0 = a^k b^m$ for some $m \in \mathbb{N}$. Test whether or not $t_x^+ \in L(G_{2k})$. In case it is, output $2k$. Otherwise, output $2k-1$."

Obviously, M changes its mind at most once. Since $L(G_{2k}) \subseteq L(G_{2k-1})$, this mind change satisfies the strong-monotonicity requirement. Furthermore, M converges to a correct hypothesis for L. Accordingly to the definition, it is easy to see that M is indeed a set-driven IIM. This proves Claim A, and therefore (1) follows.

In order to prove Assertion (2), we use the following indexed family $\mathcal{L}_{csm} = (L_{\langle k,j\rangle})_{j,k\in\mathbb{N}}$. For all $k \in \mathbb{N}$ we set $L_{\langle k,0\rangle} = \{a^k b^n \,|\, n \in \mathbb{N}^+\}$. For all $k \in \mathbb{N}$ and all $j \in \mathbb{N}^+$ we distinguish the following cases:

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

We set: $L_{\langle k,j\rangle} = L_{\langle k,0\rangle}$

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

Let $d = j - \Phi_k(k)$. Then, we set:

$L_{\langle k,j\rangle} = \{a^k b^m \,|\, 1 \leq m \leq \Phi_k(k)\} \cup \{a^k b^{\Phi_k(k)+2(d+m)} \,|\, m \in \mathbb{N}^+\}$

By reducing the halting problem to $\mathcal{L}_{csm} \in WMON$, one may prove that $\mathcal{L}_{csm} \notin WMON$. An IIM M witnessing $\mathcal{L}_{csm} \in s\!-\!CSMON$ can be easily designed, if one choose the following space of hypotheses $\mathcal{G} = (G_{\langle k,j\rangle})_{j,k\in\mathbb{N}}$. For all $k,\ j \in \mathbb{N}$, we set $L(G_{\langle k,0\rangle}) = \bigcap_{j\in\mathbb{N}} L_{\langle k,j\rangle}$ and $L(G_{\langle k,j+1\rangle}) = L_{\langle k,j\rangle}$. We omit further details.

The remaining part can be easily shown. One has simply to choose the same indexed family as used in Lange and Zeugmann (1993a) in order to separate $WMON$ and MON.

q.e.d.

4. Learning with Rearrangement-Independent IIMs.

In this section we deal with rearrangement-independent learning. The first theorem summarizes the known results.

Theorem 7. (Angluin (1980), Schäfer-Richter (1984), Fulk (1990))

$$r\!-\!ELIM = ELIM = LIM = CLIM$$

A closer look to the proof of the latter theorem shows that neither Schäfer-Richter's (1984) nor Fulk's (1990) transformation of an arbitrary, unrestricted IIM into a rearrangement-independent one preserves any of the monotonicity constraints defined. And indeed, the situation is much more subtle as the following theorems show.

Theorem 8.

(1) $r-ESMON = ESMON$,

(2) $r-SMON = SMON$.

Proof. First, we prove Assertion (2).

Let $\mathcal{L} \in SMON$. Applying the characterization theorem for $SMON$ (cf. Lange and Zeugmann (1992)), we know that there exists a class preserving space of hypothesis $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ as well as a recursively generable family $(T_j)_{j\in\mathbb{N}}$ of finite non-empty sets such that

(i) for all $j \in \mathbb{N}$, $T_j \subseteq L(G_j)$,

(ii) for all $j, k \in \mathbb{N}$, if $T_j \subseteq L(G_k)$, then $L(G_j) \subseteq L(G_k)$.

On the basis of this family $(T_j)_{j\in\mathbb{N}}$ we define an IIM M witnessing $\mathcal{L} \in r-SMON$. So let $L \in \mathcal{L}$, $t \in text(L)$, and $x \in \mathbb{N}$.

$M(t_x) =$ "Search for the least $j \leq x$ for which $T_k \subseteq t_x^+ \subseteq L(G_k)$. If it is found, output j and request the next input.
Otherwise, output nothing and request the next input."

Obviously, M is a rearrangement-independent IIM. It remains to show that M $SMON$–infers $\mathcal{L}$ with respect to the hypothesis space $\mathcal{G}$.

Claim 1. M infers L on text t.

Let $j = \mu z[L(G_z) = L]$. Hence, there is a least x such that $T_j \subseteq t_x^+$. Therefore, M will output sometimes a hypothesis. For all $k < j$ with $T_k \subseteq L$ we may conclude that $L(G_k) \subset L$. Otherwise, we obtain $L(G_j) = L(G_k) = L$, because of $T_k \subseteq L(G_k)$ and $T_j \subseteq L(G_j)$ (cf. (ii)). Hence, there exists a y such that $t_y^+ \not\subseteq L(G_k)$ for all $k < j$ with $T_k \subseteq L$. Therefore, $M(t_{y+r}) = j$ for all $r \in \mathbb{N}$. This proves the claim.

Claim 2. M works strong-monotonically.

Let $M(t_x) = j$ and $M(t_{x+r}) = k$ for some $x \in \mathbb{N}$ and $r \in \mathbb{N}^+$. Due to the definition of M, we have $T_j \subseteq t_x^+ \subseteq L(G_k)$. Therefore, $L(G_j) \subseteq L(G_k)$ (cf. (ii)). This proves the claim.

To sum up, M is witnessing $\mathcal{L} \in r-SMON$. Thus, Assertion (2) is shown.

Next, we prove Assertion (1). Let $\mathcal{L} \in ESMON$. Because of $ESMON \subseteq SMON$ as well as of Assertion (2), there exists a rearrangement-independent IIM $\hat{M}$ as well as a class preserving hypothesis space $\mathcal{G}$ such that $\hat{M}$ $SMON$–identifies $\mathcal{L}$ with respect to the hypothesis space $\mathcal{G}$.

Applying Theorem 4 of Lange and Zeugmann (1993b), we know that there exists some total recursive function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ satisfying

(i) for all $j \in \mathbb{N}$, $lim_{x\to\infty} f(j,x) = k$ exists and satisfies $L(G_j) = L_k$,

(ii) for all $j, x \in \mathbb{N}$, $L_{f(j,x)} \subseteq L_{f(j,x+1)}$.

That means, f is a limiting recursive strong-monotonic compiler from $\mathcal{G}$ into $\mathcal{L}$.

Given the IIM $\hat{M}$, the hypothesis space $\mathcal{G}$ as well as the limiting recursive strong-monotonic compiler f, we define an IIM M witnessing $\mathcal{L} \in r-ESMON$. So, let $L \in \mathcal{L}$, $t \in text(L)$, and $x \in \mathbb{N}$.

$M(t_x) =$ "Simulate $\hat{M}$ on input t_x. If $\hat{M}$ when successively fed t_x does not output any guess, then output nothing and request the next input.
Otherwise, let $j = \hat{M}(t_x)$. If $t_x^+ \subseteq L(G_j)$, then execute (A1). Otherwise, output nothing and request the next input.

(A1) Find the least $y \in \mathbb{N}$ for which $t_x^+ \subseteq L_{f(j,y)}$. Output $f(j,y)$ and request the next input."

Since the membership problem for $\mathcal{G}$ is uniformly decidable, the test "$t_x^+ \subseteq L(G_j)$" can be effectively performed. Additionally, since $\mathcal{L}$ is an indexed family, the test within instruction (A1) can be effectively accomplished, too. Furthermore, by Property (i) of f and since $t_x^+ \subseteq L(G_j)$, instruction (A1) has to terminate for every $j \in \mathbb{N}$. Hence, M is indeed an IIM. Due to its definition, M is a rearrangement-independent IIM, since the IIM $\hat{M}$ simulated by M is rearrangement-independent by assumption.

It remains to show that M strong-monotonically infers L from text t. Since $\hat{M}$ infers L from text t and by Property (i) of f, M converges to a correct hypothesis for L. Finally, we show that M fulfills the strong-monotonicity constraint. Let $f(j,y)$ and $f(k,z)$ denote two successively hypotheses generated by M. Hence, $M(t_x) = f(j,y)$ and $M(t_{x+r}) = f(k,z)$ for some $x \in \mathbb{N}$, $r \in \mathbb{N}^+$. We distinguish the following cases.

Case 1. $j = k$

Due to the definition of M, we may conclude $y \le z$. Hence, Property (ii) guarantees $L_{f(j,y)} \subseteq L_{f(j,z)}$.

Case 2. $j \neq k$

Since f satisfies (i) and (ii), we obtain $L_{f(j,y)} \subseteq L(G_j)$. Furthermore, M's definition implies $t_{x+r}^+ \subseteq L_{f(k,z)}$. Hence, the given IIM $\hat{M}$ has generated the hypothesis j on an initial segment of a text for $L_{f(k,z)} \in \mathcal{L}$. Since $\hat{M}$ works strong-monotonically on every text for every language $L \in \mathcal{L}$, we may conclude that $L(G_j) \subseteq L_{f(k,z)}$. Together with $L_{f(j,y)} \subseteq L(G_j)$, we get $L_{f(j,y)} \subseteq L_{f(k,z)}$.

Thus, M is rearrangement-independent and it works strong-monotonically. This proves the theorem.

q.e.d.

Theorem 9.

(1) $s-EMON \subset r-EMON \subset EMON$,

(2) $s-MON \subset r-MON \subset MON$.

Proof. First of all, we show $r-EMON \setminus s-MON \neq \emptyset$. By definition, this yields immediately $s-EMON \subset r-EMON$ as well as $s-MON \subset r-MON$.

Lemma 1. $r-EMON \setminus s-MON \neq \emptyset$

By Theorem 3 we already know that $r-ESMON \setminus s-LIM \neq \emptyset$. It is easy to verify that $r-ESMON \subseteq r-EMON$. By definition, $s-MON \subseteq s-LIM$. Hence, we may conclude $r-EMON \setminus s-MON \neq \emptyset$. This proves the lemma.

It remains to show $EMON \setminus r-MON \neq \emptyset$. This statement directly implies $r-EMON \subset EMON$ and $r-MON \subset MON$, and hence, the theorem will be proved.

Lemma 2. $EMON \setminus r-MON \neq \emptyset$

We only present an indexed family $\mathcal{L} = (L_k)_{k\in\mathbb{N}}$ which witnesses the desired separation. A detailed proof can be found in Lange and Zeugmann (1993d). For all $k \in \mathbb{N}$ and all $z \in \{0, \dots, 3\}$ we define:

$$L_{4k+z} = \begin{cases} \{a^k b\} \cup A_k, & \text{if } z = 0, \\ \{a^k c\} \cup B_k, & \text{if } z = 1, \\ \{a^k b, a^k c\} \cup A_k, & \text{if } z = 2, \\ \{a^k b, a^k c\} \cup B_k, & \text{if } z = 3. \end{cases}$$

The remaining languages A_k and B_k will be defined via their characteristic functions f_{A_k} and f_{B_k}, respectively. For all $k \in \mathbb{N}$ and all strings $s \in \{a, b, c\}^+$ we set:

$$f_{A_k}(s) = \begin{cases} 1, & \text{if } s = b^k a^m \text{ and } \Phi_k(k) = m, \\ 0, & \text{otherwise.} \end{cases}$$

$$f_{B_k}(s) = \begin{cases} 1, & \text{if } s = c^k a^m \text{ and } \Phi_k(k) = m, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that $\mathcal{L}$ is indeed an indexed family.

q.e.d.

Finally, we consider rearrangement-independence in the context of exact and class preserving conservative learning. Since conservative learning is exactly as powerful as weak-monotonic one, by the latter theorem one might expect that rearrangement-independence is a severe restriction under the weak-monotonic constraint, too. On the other hand, looking at Theorem 4 we see that conservative learning has its peculiarities. And indeed, exact and class preserving learning can always be performed by rearrangement-independent IIMs.

Theorem 10.

(1) $r-ECONSERVATIVE = ECONSERVATIVE$,

(2) $r-CONSERVATIVE = CONSERVATIVE$.

The following figure summarizes the results obtained and points to the questions that remain open.

	exact learning	class preserving learning	class comprising learning
FIN	*set drivenness* +	*set drivenness* +	*set drivenness* +
SMON	*rearrangement independence* +	*rearrangement independence* +	?
MON	*rearrangement independence* −	*rearrangement independence* −	?
CONSERVATIVE	*rearrangement independence* +	*rearrangement independence* +	*set drivenness* +
LIM	*rearrangement independence* +	*rearrangement independence* +	*rearrangement independence* +

For every mode of learning ID mentioned "*rearrangement-independence* +" indicates $r{-}ID = ID$ as well as $s{-}ID \subset ID$. "*Rearrangement-independence* −" implies $s\text{–}ID \subset r\text{–}ID \subset ID$ whereas "*set-drivenness* +" should be interpreted as $s\text{–}ID = ID$ and, therefore, $r{-}ID = ID$, too.

5. References

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

BERWICK, R. (1985), "The Acquisition of Syntactic Knowledge," Cambridge, Mass.: MIT Press.

FULK, M. (1990), Prudence and other restrictions in formal language learning, *Information and Computation* **85**, 1 - 11.

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

HOPCROFT, J.E., AND ULLMAN, J.D (1969), "Formal Languages and their Relation to Automata," Addison-Wesley, Reading, Massachusetts.

JANTKE, K.P. (1991) Monotonic and non-monotonic inductive inference, *New Generation Computing* **8**, 349 - 360.

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 27 - 29, Pittsburgh, pp. 377 - 390, ACM Press.

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 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," Santa Cruz, July 1993, pp. 127 - 136, ACM Press, New York.

LANGE, S., AND ZEUGMANN, T. (1993c), The learnability of recursive languages in dependence on the hypothesis space, GOSLER–Report 20/93, FB Mathematik, Informatik und Naturwissenschaften, HTWK Leipzig.

LANGE, S., AND ZEUGMANN, T. (1993d), On the impact of order independence to the learnability of recursive languages, ISIS-RR-93-17E, ISIS FUJITSU Labs., Numazu.

MACHTEY, M., AND YOUNG, P. (1978) "An Introduction to the General Theory of Algorithms," North-Holland, New York.

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.

SCHÄFER-RICHTER, G. (1984), "Über Eingabeabhängigkeit und Komplexität von Inferenzstrategien." Dissertation, Rheinisch Westfälische Technische Hochschule Aachen.

WEXLER, K. (1992), The subset principle is an intensional principle, *in* "Knowledge and Language: Issues in Representation and Acquisition," (E. Reuland and W. Abraham, Eds.), Kluwer Academic Publishers.

WEXLER, K., AND CULICOVER, P. (1980), "Formal Principles of Language Acquisition," Cambridge, Mass.: MIT Press.

WIEHAGEN, R. (1991), A thesis in inductive inference, *in* "Proceedings First International Workshop on Nonmonotonic and Inductive Logic," December 1990, Karlsruhe, (J. Dix, K.P. Jantke and P.H. Schmitt, Eds.), Lecture Notes in Artificial Intelligence Vol. 543, pp. 184 - 207, Springer-Verlag, Berlin.