Language Learning Without Overgeneralization

Shyam Kapur*
Gianfranco Bilardi*

TR 90-1168
November 1990

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

*Supported in part by the National Science Foundation under grant MIP-86-02256

LANGUAGE LEARNING WITHOUT OVERGENERALIZATION

Shyam Kapur[1]
Gianfranco Bilardi[1]

Department of Computer Science
Cornell University
Ithaca New York 14853.

14th November, 1990

Abstract. Language learnability is investigated in the Gold paradigm of inductive inference from positive data. Angluin gave a characterization of learnable families in this framework. Here, learnability of families of recursive languages is studied when the learner obeys certain natural constraints. Exactly learnable families are characterized for prudent learners with the following types of constraints: (0) conservative, (1) conservative and consistent, (2) conservative and responsive, and (3) conservative, consistent and responsive. The class of learnable families is shown to strictly increase going from (3) to (2) and from (2) to (1), while it stays the same going from (1) to (0). It is also shown that, when exactness is not required, prudence, consistency and responsiveness, even together, do not restrict the power of conservative learners.

[1]Supported in part by the National Science Foundation under grant MIP-86-02256.

1 Introduction

Gold (1967) presented a basic paradigm of inductive inference. While inference from complete data was shown by him to be powerful, interesting classes of language families were first shown to inferable from *positive data*, i.e., where only strings in the language to be learned are presented, in Angluin (1980a & b). Angluin (1980b) showed that the families of languages learnable from positive data are exactly those for which there is a procedure that, for each language in the family, enumerates a finite set of which that language is a least upper bound.

Within language learning, Angluin (1980b) also initiated the study of *conservative* learners, which do not change a guess unless evidence inconsistent with it is encountered. It can easily be shown that a family can be learned by a conservative learner if and only if it can be learned by a *non overgeneralizing* learner, that is, one that never guesses a subset of a previous guess. Learning without overgeneralization is of interest to research in natural language acquisition since it has been hypothesized (for example, Berwick (1985)) that children learn in this fashion. (However, see Mazurkewich & White (1984) for evidence to the contrary.) Angluin (1980b) established that conservativeness actually restricts the class of learnable families, and derived a sufficient (but not necessary) condition for conservative learning.

In a framework similar to that in Angluin (1980a & b), defined in Section 2, we further investigate conservative learning. We consider the case when the language to be learned is *recursive* and where the learner has to identify the language by a *total Turing machine*. A variety of natural constraints on learners have been proposed in literature (Osherson, Stob & Weinstein, 1986), such as *prudence* (the learner must learn all the languages it ever guesses), *consistency* (at any stage of the learning process, the guessed language must contain all the evidence), and *responsiveness* (a guess must be made for any new piece of evidence). A family is said to be *exactly* learned if the learner learns it and nothing larger. In Section 3, we investigate the impact of all these constraints on conservative learning and demonstrate a nested relation between the classes of families exactly learnable by prudent learners of four types defined by selectively and together imposing responsiveness and consistency.

In Section 4, we first give a uniform characterization of prudent, exact conservative learning of the four types defined in Section 3, based on the concept of least upper bound. While a consistent guess of a conservative learner must always be a least upper bound of the evidence seen, simply guessing a least upper bound does not guarantee learning. We develop a learning algorithm based on the judicious use of a function that computes least upper bounds. One of the corollaries of our characterization is that prudent, exact conservative learnability of a given family of languages F is equivalent to the existence of a recursive enumeration of finite-set/total-machine pairs such that (i) in each pair, the machine accepts a language of F that is a least upper bound in F of the corresponding finite set, and (ii) there is at least one pair for each language in F. Interestingly, if the condition on the enumeration is weakened so that the first component of the pair is a machine that recursively enumerates a finite set, then the existence of the enumeration characterizes general learning.

Finally, in Section 5, we investigate the impact of dropping the requirement of prudence on the learner and of exactness on the family learned. We show that every family that can be learned conservatively, can also be learned conservatively and prudently, although not necessarily exactly. In fact, the hierarchy obtained for prudent, exact conservative learning collapses when exactness is not stipulated, and prudence, consistency, and responsiveness, even together, fail to constrain the conservative learners.

2 Background

Let Σ be a finite alphabet and Σ^* be the set of all finite strings formed by concatenating elements of Σ. Let $M_1, M_2, M_3, \ldots$ be any standard enumeration of all Turing machines over Σ. Let Z_+ be the set of positive integers. For any *index* $I \in Z_+$, let W_I denote the *language* (subset of Σ^*) accepted by the machine M_I. Thus, the W_Is form a complete enumeration of all *recursively enumerable (r.e.)* languages. A collection of *non-empty*, *recursive* languages is called a *family*. An index I is *total* if the corresponding machine M_I is total and accepts a non-empty language. If $I_1, I_2, \ldots$ is a recursive enumeration of total indices, then $F = W_{I_1}, W_{I_2}, \ldots$ is called an *indexed family of non-empty recursive languages* (hereafter, simply an *indexed family*). We denote by Δ_F the class of all non-empty finite subsets of Σ^* that are

contained in some language in family F.

A *text* is an infinite sequence of strings from Σ^*; t ranges over texts, t_n is the nth string in text t, and $\overline{t}_n$ is the initial prefix of length n of the text t; t is *for* L if and only if the set of strings in t equals L. SEQ is the set of finite sequences of strings from Σ^*; *content*(σ) is the set of strings in $\sigma \in SEQ$.

An *inductive inference machine (IIM)* M is an algorithmic procedure (say, a Turing machine) whose input is a text $t_1, t_2, \ldots$ and whose output is a sequence of nonnegative integers $M(\overline{t}_1), M(\overline{t}_2), \ldots$ constrained to be either 0 or total indices. The procedure works in stages, but it may never complete some stage. At the nth stage, t_n is input and $M(\overline{t}_n)$ is output. The intended interpretation is as follows: If $M(\overline{t}_n) = 0$, then the IIM makes no guess; otherwise, it guesses the language $W_{M(\overline{t}_n)}$.

For an IIM M, we define GUESS$(M) = \{W_{M(\sigma)}: \sigma \in SEQ\}$. We observe that GUESS$(M)$ is an indexed family. An IIM M is said to learn the language L if and only if, for each text t for L, there is a k such that $W_{M(\overline{t}_k)} = L$ and, for all $n > k$, $M(\overline{t}_n) \in \{M(\overline{t}_k), 0\}$. (Thus, if t is a text for a language that M learns, then $M(\overline{t}_n)$ is defined for any n.) Intuitively, from some point onward, the value of the guess stabilizes to a total index for the input language. (This is similar to the TXTEX-identification criterion (Gold, 1967).) We denote by LEARN(M) the family of all languages learned by M, and say that M learns a family F if and only if $F \subseteq$ LEARN(M).

An IIM M is *consistent* if it never guesses a language that does not contain all the data on which the guess is based. Formally, if M completes the nth stage and $M(\overline{t}_n) \neq 0$, then *content*$(\overline{t}_n) \subseteq W_{M(\overline{t}_n)}$. M is *conservative* if it does not change its output unless it is inconsistent with the data. Formally, if M completes the nth stage, $M(\overline{t}_n) \neq 0$, and *content*$(\overline{t}_{n+1}) \subseteq W_{M(\overline{t}_n)}$, then $M(\overline{t}_{n+1}) = M(\overline{t}_n)$. M is *responsive* if, for all $n > 0$, whenever *content*$(\overline{t}_n) \in \Delta_{\text{GUESS}(M)}$, then M completes the nth stage and $M(\overline{t}_n) \neq 0$. M is *prudent* if LEARN(M) = GUESS(M). A family F is said to be *exactly* learned by M if F = LEARN(M).

A central role is played by the relationship that holds between a finite set T and a language L in a family F when T is a subset of L but T is not a subset of any language in F properly contained in L. This relation will be expressed by saying either that T is a *tell-tale* subset of L in F, or that L is a *least upper*

bound (l.u.b.) of T in F. As we shall see, various forms of computable mappings from (indices for) languages to tell-tale subsets and from finite sets to (indices for) l.u.b. languages are intimately related to various forms of learning. The importance of tell-tale subsets was first indicated by the following result.

Theorem 1. (Angluin (1980b)) There is an IIM M such that an indexed family $F = \text{LEARN}(M)$ if and only if there is an effective procedure that, given as input a total index I such that $W_I \in F$, recursively enumerates a tell-tale subset T_I of W_I. (The procedure need not signal when T_I has been entirely output.)

3 Variants of Conservative Learning

Angluin (1980b) showed that, in the absence of the conservativeness constraint, if F is learnable it is also learnable by a consistent and responsive algorithm. In this section, we begin to analyze the relationships between the classes of families learnable by various types of conservative learners. It is convenient to introduce the following notation.

Definition 1. An IIM M is said to be of *type 0* if conservative, of *type 1* if conservative and consistent, of *type 2* if conservative and responsive, and of *type 3* if conservative, consistent, and responsive. For $X = 0,1,2,3$, we denote by Φ_X the class of all families learnable by an IIM of type X, Φ_X^p the class of all families learnable by a prudent IIM of type X, and Φ_X^{ep} the class of all families learnable exactly by a prudent IIM of type X.

We next show how consistency and responsiveness constrain the power of exact, prudent learners.

Theorem 2. $\Phi_0^{ep} = \Phi_1^{ep} \supset \Phi_2^{ep} \supset \Phi_3^{ep}$.

Proof. ($\Phi_0^{ep} = \Phi_1^{ep}$) Clearly, $\Phi_0^{ep} \supseteq \Phi_1^{ep}$. To show that $\Phi_1^{ep} \subseteq \Phi_0^{ep}$ we consider that an IIM of type 0 can be modified to refrain from making a guess whenever it would make an inconsistent guess. It is easy to see that the modified IIM learns the original family and in addition is of type 1.

($\Phi_1^{ep} \supset \Phi_2^{ep}$) Clearly, $\Phi_2^{ep} \subseteq \Phi_0^{ep} = \Phi_1^{ep}$. To see that $\Phi_2^{ep} \neq \Phi_1^{ep}$, consider the indexed family F: for all $k \geq 1$, $W_{I_k} = \{0\} \cup \{k, k+1, \ldots\}$. A prudent IIM M of type 1 learns exactly F for which

$M(\bar{t}_n) = I_{\min(content(\bar{t}_n)-\{0\})}$, with $\min(\varnothing)$ defined as 0. However, F can not be learned exactly by a prudent IIM of type 2, because on the input of a text t for any language in F for which $t_1 = 0$, no index for a language in F can be guessed safely at the first stage.

($\Phi_2^{ep} \supset \Phi_3^{ep}$) Clearly, $\Phi_2^{ep} \supseteq \Phi_3^{ep}$. To see that $\Phi_3^{ep} \neq \Phi_2^{ep}$, consider the indexed family F: for all $k > 1$, $W_{I_k} = \{0\} \cup \{k, k+1, ...\}$, and $W_{I_1} = \{5\}$. By the argument given in the preceding paragraph, there is no prudent IIM of type 3 that learns exactly F. We do have a prudent IIM of type 2 that learns exactly F, which behaves as follows: If $t_1 = t_2 = ... = t_n = 0$ (in a text t for any language in F excluding W_{I_1}), or $t_1 = t_2 = ... = t_n = 5$ (in a text t for W_{I_1}), then $M(\bar{t}_n) = I_1$, else $M(\bar{t}_n) = I_{\min(content(\bar{t}_n)-\{0\})}$. □

We shall return to the study of the relations between the classes introduced in Definition 1 after the next section, where some of the tools necessary for the analysis are developed.

4 Characterization of Prudent, Exact Conservative Learning

In the present section, we develop, in a unified way, a characterization of Φ_X^{ep}, for $X = 0, 1, 2$, and 3, in terms of the computability of l.u.b. languages for certain finite sets. We then reformulate these characterizations for some specific cases. It is a simple observation that a consistent guess made by a conservative IIM M in response to a set of input strings S subset of some $L \in \text{LEARN}(M)$, must be a least upper bound for S in the family LEARN(M). (This has been referred to as the *subset principle* by Berwick (1985).) However, systematically guessing an l.u.b. of the available evidence does not guarantee learning, even when F is conservatively learnable. Thus, it is natural to investigate to which extent the ability of computing l.u.b. languages is necessary or sufficient for conservative learning.

We establish that a computable function that associates an l.u.b. to each finite subset of a language in F is sufficient to construct a conservative IIM, provided that each language in F is associated with at least one of its subsets. However, the existence of such an l.u.b. function is not necessary except for the consistent and responsive case. Indeed, two issues have to be confronted when trying to construct an l.u.b. function from a conservative IIM. First, with some input set, a non-responsive IIM may associate

no language and a non-consistent IIM may associate a language that does not contain the set itself. Second, when the family being learned contains a finite language L, even at a stage when all the strings of L have appeared in the input, the IIM's guess may well differ from L. We deal with these issues by introducing, in the formal definition of l.u.b. functions given below, a parameter k that models the possible delay in the IIM's convergence to a correct guess.

Without loss of generality, for the prudent, exact learning case we can assume that the family F to be learned is an indexed family. We recall that Δ_F denotes the class of all non-empty finite sets contained by some language in F, and let $\Delta = \Delta_{\{\Sigma^*\}}$.

Definition 2. A partial recursive function $f : \Delta \times Z_+ \rightarrow Z_+ \cup \{0\}$ is a *least upper bound (l.u.b.)* function of *type* X ($X = 0,1,2,3$) for indexed family F if, for $S \in \Delta_F$ and $k \in Z_+$, $f(S,k)$ is defined, either as 0 or as a total index such that $W_{f(S,k)} \in F$, and satisfies the following conditions:

type 0: if $S \subseteq W_{f(S,k)}$, then $W_{f(S,k)}$ is an l.u.b. of S;
type 1: f is of type 0 and ($f(S,k) = 0$ or $S \subseteq W_{f(S,k)}$);
type 2: f is of type 0 and $f(S,k) \neq 0$;
type 3: f is of type 0 and $S \subseteq W_{f(S,k)}$.

An l.u.b. function f is *full-range* if each L in F has a subset S such that, for some $k \in Z_+$, $W_{f(S,k)} = L$.

In terms of the concepts just introduced, we next state our main theorem.

Theorem 3. Let $F = W_{I_1}, W_{I_2}, \ldots$ where $I_1, I_2, \ldots$ is a recursive enumeration of total indices. For $X = 0,1,2,3$, $F \in \Phi_X^{ep}$ if and only if there exists a full-range l.u.b. function of type X for F.

Proof. The proof is organized in three parts. In part (a) we show how to obtain a full-range l.u.b. function of type X from a prudent IIM M of type X that learns exactly F, for $X = 0,1,2,3$. In part (b), we prove that any full-range l.u.b. function can effectively be converted into one with a certain special property. Finally, in part (c), we show how an l.u.b. function of type X of the special kind obtained in part (b) can be used to construct a prudent IIM M' of type X that learns exactly F, thereby completing the proof.

(a) *From M to a full-range l.u.b. function.* The following procedure defines a function f : $\Delta \times Z_+ \rightarrow Z_+ \cup \{0\}$. Let $S \in \Delta$ and $k \in Z_+$. Scan $F = W_{I_1}, W_{I_2}, \ldots$ until an I_j is found such that $S \subseteq W_{I_j}$. (This scan terminates if and only if $S \in \Delta_F$.) Then let $f(S,k)$ be the output of M when the input is the lexicographic enumeration $s_1, s_2, \ldots, s_n$ of S ($n = |S|$) followed by $k-1$ repetitions of s_n. Since F is learned, for any $S \in \Delta_F$, this output must be defined. It is a simple exercise to show that if M is an IIM of type X, then f is an l.u.b. function of type X.

To establish the full-range property of f, we argue as follows. If $L \in F$ is infinite, let $t = t_1, t_2, t_3, \ldots$ be a text for L corresponding to its lexicographic presentation. Consider the behavior of M on t. At some stage n, M must output an index of L. Therefore $W_{f(\{content(\bar{t}_n)\},1)} = L$. If $L \in F$ is finite, let $s_1, s_2, \ldots, s_n$ be the lexicographic enumeration of L, and let $s_1, s_2, \ldots, s_n, s_n, s_n, \ldots$ be a text for L presented to M. At some stage h, M must output an index for L. If $h \leq n$, then $W_{f(\{s_1, \ldots, s_h\},1)} = L$, else, $W_{f(\{s_1, \ldots, s_n\}, h-n+1)} = L$.

(b) *From a full-range l.u.b. function to a special l.u.b. function.* Let f be an l.u.b. function for an indexed family F. Let $L \in F$ and R be a finite subset of L. We say that R is *reserved* for L beyond m if, for every finite S such that $R \subseteq S \subseteq L$, and every $k \geq m$, $W_{f(S,k)} = L$. A *special* l.u.b. function is one that reserves a set for every L in F. Thus, if f is special, then it is also full-range.

Given any full-range l.u.b. function f, we can construct another one, $\bar{f}$, which is special. To define $\bar{f}$, consider a recursive enumeration $(D_1, h_1), (D_2, h_2), \ldots$ of $\Delta \times Z_+$ such that if $h < h'$ then (D,h) appears before (D,h'). First, F is scanned to determine whether $S \in \Delta_F$. If so, consider the set $J(S,k) \stackrel{\Delta}{=} \{j : j \leq k, D_j \subseteq S, \text{ and } S \subseteq W_{f(D_j,h_j)}\}$ (where, for convenience, we let $W_0 = \emptyset$). Observe that whenever $D_j \subseteq S \in \Delta_F$, $f(D_j, h_j)$ must be defined. If $J(S,k)$ is empty, we let $\bar{f}(S,k) = f(S,k)$. Otherwise, we let $i = \min J(S,k)$ and $\bar{f}(S,k) = f(D_i, h_i)$. In this case, since $D_i \subseteq S \subseteq W_{f(D_i,h_i)}$, we have that $W_{\bar{f}(S,k)} = W_{f(D_i,h_i)}$ is an l.u.b. of D_i and hence of S. Clearly, $\bar{f}$ is computable with the help of the algorithm that computes f and of the recursive enumeration of indices $I_1, I_2, \ldots$. It is also simple to see that $\bar{f}$ is an l.u.b. function of the same type as f. We now show that $\bar{f}$ is special.

Since f is full-range, for any given $L \in F$, we can properly define the positive integer $m = \min\{i : W_{f(D_i,h_i)} = L\}$. We now observe that, if $j<m$ and $D_j \subseteq L$, then $W_{f(D_j,h_j)} \not\supseteq L$. In fact, $W_{f(D_j,h_j)} = L$ would violate the definition of m, since $j<m$. Moreover, $W_{f(D_j,h_j)} \supset L$ would violate the basic l.u.b. property of f, since $W_{f(D_j,h_j)} \supset L \supseteq D_j$ implies that $W_{f(D_j,h_j)}$ is not an l.u.b. of D_j. Thus, there exists a finite set V containing a string $x_j \in (L - W_{f(D_j,h_j)})$, for every $j<m$ such that $D_j \subseteq L$. We claim that the set $R = D_m \cup V$ is reserved by $\bar{f}$ for L, beyond m. In fact, let $S \in \Delta_F$ be such that $R \subseteq S \subseteq L$ and $k \geq m$. Then, the set $J(S,k) = \{j : j \leq k, D_j \subseteq S, \text{ and } S \subseteq W_{f(D_j,h_j)}\}$ has minimum m. Indeed, it contains m (since $D_m \subseteq S \subseteq W_{f(D_m,h_m)} = L$) and no $j<m$ (since if $j<m$ and $D_j \subseteq S$ then $S \not\subseteq W_{f(D_j,h_j)}$, because $x_j \in S$ but $x_j \notin W_{f(D_j,h_j)}$). In conclusion, $\bar{f}(S,k) = \min J(S,k) = m$ and R is reserved for L, as claimed.

(c) *From a special l.u.b. function to M'.* Given a special l.u.b. function $\bar{f}$ of type X for F, consider the IIM M' that works as follows: At stage 1, first scan F to determine whether $\{t_1\} \in \Delta_F$. If so, let $M'(\bar{t}_1) = \bar{f}(\{t_1\}, 1)$. At stage $n>1$, maintain the previous guess if consistent with the previous input as well as with t_n. Otherwise, scan the family F to determine whether $content(\bar{t}_n) \in \Delta_F$. If so, let $M'(\bar{t}_n) = \bar{f}(content(\bar{t}_n), n)$. It is easy to see that M' is an IIM of type X and that GUESS$(M') \subseteq F$. We next establish that M' is prudent and learns exactly the indexed family F.

Let $L \in F$, and let R be reserved beyond m for L. Given a text $t_1, t_2, \ldots$ for L, let h be the smallest integer such that $h \geq m$ and $R \subseteq content(\bar{t}_h)$. Then, either $W_{M'(\bar{t}_h)} = L$, in which case M' has already reached the correct guess at stage h, or there is an integer $r>h$ such that t_r is not in $W_{M'(\bar{t}_h)}$. Then $M'(\bar{t}_r) = \bar{f}(content(\bar{t}_r), r)$ is an index for L because R is reserved by $\bar{f}$ beyond m. In conclusion, L is learned by M. □

A simple adaptation of part (a) of the proof of Theorem 3 yields the following corollary.

Corollary 1. If $F \in \Phi_3^{ep}$ or if $F \in \Phi_X^{ep}$ $(X = 0,1,2)$ and does not contain any finite language, then there is a full-range l.u.b. function for F that does not depend on k.

There is an interesting analogy between the second case of the above corollary and the situation for general *set-driven* learning, which is always achievable for learnable families of infinite languages, but not always for learnable families that contain finite languages. (See Osherson, Stob & Weinstein, 1986.)

For Φ_1^{ep}, the characterizing condition of Theorem 3 admits the following useful reformulation.

Corollary 2. An indexed family $F \in \Phi_1^{ep}$ if and only if there is a recursive enumeration $(T_1,I_1),(T_2,I_2),...$ of pairs such that, for each $h \in Z_+$, T_h is a finite set, I_h is a total index, W_{I_h} is an l.u.b. of T_h in F, and $F = \{W_{I_1},W_{I_2},...\}$.

Proof. From Theorem 3, $F \in \Phi_1^{ep}$ if and only if there is a full-range l.u.b. function f of type 1 for F.

Given such an f, scan a recursive enumeration $(D_1,h_1),(D_2,h_2),...$ of $\Delta_F \times Z_+$ and, for each i such that $f(D_i,h_i) \neq 0$, output the pair $(D_i,f(D_i,h_i))$. It is easy to check that the pairs enumerated by this procedure satisfy all required properties.

Given a recursive enumeration $(T_1,I_1),(T_2,I_2),...$ satisfying the properties listed in the statement of Corollary 2, we define a function $f: \Delta \times Z_+ \rightarrow Z_+ \cup \{0\}$ as follows. For an $(S,k) \in \Delta \times Z_+$, an h is searched such that $S \subseteq W_{I_h}$. If no such h is found, let $f(S,k)$ diverge. Otherwise, let the machine that enumerates $(T_1,I_1),(T_2,I_2),...$ run for k steps. If no pair of the form (S,I) is output, then let $f(S,k) = 0$, else let $f(S,k) = I$, where (S,I) is the last pair output in k steps whose first component equals S. It is easy to check that f is a full-range l.u.b. function of type 1 for F. □

The characterization of general learning stated at the end of Section 2 as Theorem 1 is now restated for a direct comparison with Corollary 2. The equivalence between the two formulations is not immediate, but is a corollary of the developments of Angluin (1980b).

Theorem 1′. There is an IIM M such that an indexed family $F = \text{LEARN}(M)$ if and only if there is a recursive enumeration $(E_1,I_1),(E_2,I_2),...$ of pairs such that, for each $h \in Z_+$, E_h is a procedure that recursively enumerates a finite set T_h, I_h is a total index, W_{I_h} is an l.u.b. of T_h in F, and $F = \{W_{I_1},W_{I_2},...\}$.

5 Prudence does not Restrict Conservative Learners

In this section, we study conservative learnability when prudence or exactness are not required. In the case of general learning of r.e. languages, Fulk (1990) showed that prudence does not restrict the class of learnable families. Specifically, for every IIM M, there is a TXTEX-prudent IIM M' with LEARN(M)$\subseteq$LEARN(M'). In the same spirit, although with different techniques, we have:

Theorem 4. $\Phi_1 = \Phi_1^p$.

Proof. Consider a family $F \in \Phi_1$ and a conservative IIM of type 1 M that learns F. For $D \in \Delta$ and $k \in Z_+$, let $<D,h>$ denote the lexicographic ordering of D followed by $h-1$ repetitions of the last string. Let $(S_1,J_1),(S_2,J_2),...$ be the recursive enumeration, called the S-enumeration, obtained as follows. Scan a recursive enumeration $(D_1,h_1,k_1),(D_2,h_2,k_2),...$ of $\Delta \times Z_+ \times Z_+$, and, if for some i M halts on input $<D_i,h_i>$ in at most k_i steps with $M(<D_i,h_i>) \neq 0$, then output the pair $(D_i,M(<D_i,h_i>))$.

As M is not necessarily prudent, W_{J_i} need not be an l.u.b. of S_i. However, based on the S-enumeration, we will construct another enumeration of the type considered in Corollary 2. To this end, let us say that (S_r,J_r) is *certified* with *certificate* V_r if there is a finite set $V_r \subseteq W_{J_r}$ such that, for all $p<r$ for which $S_p \subseteq W_{J_r}$, $S_r \cup V_r \subseteq W_{J_p}$. It is straightforward to construct a certification procedure that, on input r, will output a certificate V_r if there is one, and run for ever otherwise. Using this procedure in a dove-tail fashion, we can produce a recursive enumeration $(T_1,I_1),(T_2,I_2),...$, called the T-enumeration, containing the pair $(S_r \cup V_r,J_r)$ exactly when the pair (S_r,J_r) is certified with certificate V_r. We now state and prove two claims for the T-enumeration.

Claim 1: For each $h \in Z_+$, W_{I_h} is an l.u.b. of T_h in $\{W_{I_1},W_{I_2},...\}$. Let $(S_1',J_1'),(S_2',J_2'),...$ be the subsequence of the S-enumeration containing the certified pairs, and let V_i' be the certificate produced for pair (S_i',J_i'). Then, the sequence $(S_1' \cup V_1',J_1'),(S_2' \cup V_2',J_2'),...$ is a rearrangement of the T-enumeration. Therefore, the claim can be reformulated as follows assuming the T-enumeration to be infinite (the finite case can be argued similarly): For each $r \in Z_+$, and each k such that $1 \leq k \leq r$, $W_{J_k'}$ is an l.u.b. of $S_k' \cup V_k'$ in

$\{W_{J_1'}, \ldots, W_{J_r'}\}$. For $r = 1$, the latter statement is trivially true. Let us inductively assume it holds for $r-1$ and establish it for r. If the statement fails for r, then for some $p < r$ either (a) $(S_p' \cup V_p') \subseteq W_{J_r'} \subset W_{J_p'}$ or (b) $(S_r' \cup V_r') \subseteq W_{J_p'} \subset W_{J_r'}$. In case (a), we have $S_p' \subseteq W_{J_r'}$ (by hypothesis) and $(S_r' \cup V_r') \subset W_{J_p'}$ (since $(S_r' \cup V_r') \subset W_{J_r'}$ and, by hypothesis, $W_{J_r'} \subset W_{J_p'}$). In case (b), we have $S_p' \subseteq W_{J_r'}$ (since $S_p' \subseteq W_{J_p'}$ and, by hypothesis, $W_{J_p'} \subset W_{J_r'}$) and $(S_r' \cup V_r') \subseteq W_{J_p'}$ (by hypothesis). In either case, we reach a contradiction with the fact that V_r' is a certificate for (S_r', J_r') and the claim remains established.

Claim 2: $F \subseteq \{W_{I_1}, W_{I_2}, \ldots\}$. For $L \in F$, let (S_i, J_i) be the first pair in the S-enumeration such that $W_{J_i} = L$. The existence of such a pair can be established by arguments similar to those used in part (a) of the proof of Theorem 3. It suffices to show that (S_i, J_i) gets certified. If not, there is another pair (S_k, J_k), with $k < i$, such that $S_k \subseteq W_{J_i}$ and $W_{J_i} \subset W_{J_k}$. By the construction of the S-enumeration, this implies that for some $h \in Z_+$, $M(\langle S_k, h \rangle) = J_k$. Thus, if t is a text for $W_{J_i} = L$ with $\langle S_k, h \rangle$ as a prefix, for $n = |S_k| + h - 1$ $M(\overline{t}_n) = J_k$. Moreover, since $L \subset W_{J_k}$ and M is conservative, for $n > |S_k| + h - 1$, $M(\overline{t}_n) = J_k$. In conclusion, M does not learn $L \in F$, against our assumptions.

By Claim 1 and Corollary 2, there is a prudent IIM M' such that LEARN $(M') = \{W_{I_1}, W_{I_2}, \ldots\}$. By Claim 2, $F \subseteq$ LEARN(M') so that $F \in \Phi_1^p$. Thus $\Phi_1 \subseteq \Phi_1^p$. As $\Phi_1^p \subseteq \Phi_1$, we conclude that $\Phi_1 = \Phi_1^p$. □

We next show that, when exactness is not required, responsiveness and consistency do not constrain the power of prudent learners.

Theorem 5. $\Phi_0^p = \Phi_1^p = \Phi_2^p = \Phi_3^p$.

Proof. That $\Phi_0^p = \Phi_1^p$ follows from the fact that a prudent IIM M of type 0 can be modified to refrain from making an inconsistent guess. The modified IIM is still prudent and learns the original family. To show that $\Phi_1^p = \Phi_2^p = \Phi_3^p$, we observe that, by definition, $\Phi_1^p \supseteq \Phi_2^p \supseteq \Phi_3^p$ and prove that $\Phi_1^p = \Phi_3^p$.

Let $F \in \Phi_1^p$ and let M be a prudent IIM of type 1 that learns F. We shall prove that $F \in \Phi_3^p$ by constructing a full-range l.u.b. function of type 3, f, for a family $F' \supseteq F$ and invoking Theorem 3. We will use

the fact that, since M is prudent, $\Delta_{\text{LEARN}(M)}$ is recursively enumerable. For $S \in \Delta$, let $<S>$ denote the lexicographic ordering of S and, for a finite language L, let *index* (L) denote a total index such that $W_{index(L)} = L$. With these preliminaries, for $S \in \Delta$, let $f(S)$ be defined (recursively) as follows.

If $S \notin \Delta_{\text{LEARN}(M)}$ then (a) let $f(S)$ diverge. Else $(S \in \Delta_{\text{LEARN}(M)})$ if $M(<S>) \neq 0$ then (b) let $f(S) = M(<S>)$. Else $(M(<S>) = 0)$ if there is any non-empty Q such that $Q \subset S \subseteq W_{f(Q)}$, then (c) let $f(S) = f(Q)$; if more than one Q qualifies, choose a minimal one arbitrarily. Else (d) let $f(S) = index(S)$.

We observe that, for any $S \in \Delta_{\text{LEARN}(M)}$, $f(S)$ is a total index and consider the family $F' = \{W_{f(S)} : S \in \Delta_{\text{LEARN}(M)}\}$. We show that $\text{LEARN}(M) \subseteq F'$ and hence $F \subseteq F'$. Consider any $L \in \text{LEARN}(M)$.

If L is infinite, then for a suitable prefix σ of the lexicographic presentation of L, $W_{M(\sigma)} = L$. Since $content(\sigma) \in \Delta_{\text{LEARN}(M)}$ and $M(<content(\sigma)>) = M(\sigma) \neq 0$, f gets defined in clause (b) as $f(content(\sigma)) = M(\sigma)$, so that $L = W_{f(content(\sigma))} \in F'$.

If L is finite, $f(L)$ can get defined in any of the clauses (b), (c), and (d). If $f(L)$ is defined in (b), then $M(<L>) \neq 0$ and $W_{M(<L>)} = L$, since M must guess an l.u.b.. If $f(L)$ is defined in (c), then $f(L) = f(Q)$ for some minimal Q such that $Q \subset L \subseteq W_{f(Q)}$. The minimality of Q implies that $f(Q)$ itself has been defined either in clause (b) or clause (d). The latter case is excluded because it would imply $W_{f(Q)} = Q$, whereas by assumption $Q \subset W_{f(Q)}$. Therefore, $W_{f(Q)} = W_{M(<Q>)}$ must be an l.u.b. of Q in F. Since $L \in F$ and $Q \subset L \subseteq W_{f(Q)}$, this implies $W_{f(Q)} = L$. Finally, if $f(L)$ is defined in (d), $W_{f(L)} = W_{index(L)} = L$. In any case, $L \in F'$.

We now show that f is an l.u.b. function of type 3 for F'. Otherwise, there must be a finite set S and a language $L \in F'$ such that $S \subseteq L \subset W_{f(S)}$. Further, $f(S)$ must have been defined in clause (b) or clause (c), since in clause (d) $S = W_{f(S)}$.

If $f(S)$ is defined in clause (b) as $f(S) = M(<S>)$, then $W_{f(S)} \in \text{LEARN}(M)$, since M is prudent, and $L \notin \text{LEARN}(M)$, since M is conservative. Therefore, $L \in (F' - \text{LEARN}(M))$. By definition of f and F', $L \in (F' - \text{LEARN}(M))$ implies that $f(L)$ is defined in clause (d). However, this leads to a contradiction.

In fact, either $L = S$ and $f(L)$ is defined in clause (b), by assumption, or $L \supset S$ and $f(L)$ is defined in clause (c), because the set $\{Q : Q \subset L \subseteq W_{f(Q)}\}$ contains at least S.

If $f(S)$ is defined in clause (c) as $f(S) = f(Q)$, we observe that, due to the minimality of Q, $f(Q)$ is defined in clause (b) and the argument made for S in the previous case can now be made for Q.

Having established that f is a type 3 l.u.b. function for F', we conclude by observing that, by the definition of F', f is also full-range. □

Considering that $\Phi_0 = \Phi_1$ (by an easy argument), $\Phi_1 \supseteq \Phi_2 \supseteq \Phi_3$ (by a simple consequence of the definition), $\Phi_3 \supseteq \Phi_3^p$ (again, by definition), $\Phi_3^p = \Phi_2^p = \Phi_1^p = \Phi_0^p$ (by Theorem 5), and $\Phi_1^p = \Phi_1$ (by Theorem 4), we obtain the following result.

Theorem 6. For $X, Y \in \{0,1,2,3\}$, $\Phi_X = \Phi_Y^p$.

In summary, when exactness is not required, any combination of prudence, consistency, and responsiveness does not restrict conservative learnability.

It is natural to wonder, for any $X \in \{0,1,2,3\}$, what is the relation between Φ_X^{ep} and Φ_X. Clearly, $\Phi_X^{ep} \subset \Phi_X$ since Φ_X contains non-r.e. families whereas Φ_X^{ep} does not. However, it remains open the question whether any indexed family F that belongs to Φ_X also belongs to Φ_X^{ep}. (An analogous result does hold for TXTEX identification (Osherson, Stob & Weinstein, 1982).)

In conclusion, we would like to emphasize that the characterizations of Section 4 have been instrumental in establishing the results of the present section. We expect our characterization of conservative learning of various types to provide a means for exploring further the consequences of various constraints on learners in the Gold (1967) model of inductive inference.

Acknowledgments. We are greatly indebted to the two referees for their constructive criticisms of a previous version of this manuscript. In particular, Section 5 was motivated by the suggestions of one of the referees. We also thank Kieran Herley for his comments on an earlier draft.

References

ANGLUIN, D. (1980a) Finding patterns common to a set of strings. *Journal of Computer System Sciences*, 21, 46-62.

ANGLUIN, D. (1980b) Inductive inference of formal languages from positive data. *Information and Control*, 48, 117-135.

ANGLUIN, D & SMITH, C. H. (1983) Inductive inference: theory and methods. *Computing Surveys*, 15(3):237-269, September 1983.

BERWICK, R. (1985) *The Acquisition of Syntactic Knowledge*. MIT press, Cambridge.

FULK, M. A. (1990) Prudence and Other Conditions on Formal Language Learning. *Information and Computation*, 85, 1-11.

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

MAZURKEWICH, I. & WHITE, L. (1984) The acquisition of dative-alternation: unlearning overgeneralizations. *Cognition*, 16(3), 261-283.

OSHERSON, D. N., STOB, M. & WEINSTEIN, S. (1982) Learning Strategies. *Information and Control*, 53, 32-51.

OSHERSON, D. N., STOB, M. & WEINSTEIN, S. (1986) *Systems that Learn: An Introduction to Learning*. MIT press, Cambridge.