A Survey on Recent Results on Partial Learning

Ziyuan Gao[1], Sanjay Jain[2], Frank Stephan[3] and Sandra Zilles[4]

[1] Department of Computer Science, University of Regina,
Regina, SK, Canada S4S 0A2
Email: gao257@cs.uregina.ca

[2] Department of Computer Science, National University of Singapore,
Singapore 117417, Republic of Singapore
Email: sanjay@comp.nus.edu.sg

[3] Department of Mathematics and Department of Computer Science,
National University of Singapore, Singapore 119076, Republic of Singapore
Email: fstephan@comp.nus.edu.sg

[4] Department of Computer Science, University of Regina,
Regina, SK, Canada S4S 0A2
Email: zilles@cs.uregina.ca

Abstract. In the context of recursion-theoretic inductive inference, a partial learner identifies an r.e. language L if and only if it outputs one index e of the language L infinitely often and outputs all other indices only finitely often. Osherson, Stob and Weinstein showed in 1986 that there is a partial learner that succeeds to learn every r.e. language under this learning criterion. Therefore, subsequent investigations refined the question to what happens if the partial learner has to satisfy additional constraints like consistency, conservativeness, confidence and reliability. The present paper gives an overview of recent investigations on this topic.

1 Introduction

Gold [15] introduced a learning paradigm to model the nature of language acquisition. His model featured a learnability criterion known as "identification in the limit", whereby a learner, as it is presented piecewise with information about an unknown target language L, outputs a sequence of conjectures based on a pre-assigned system of naming all the languages to be learnt, and after some finite time must settle on exactly one correct description of L. The present paper deals almost exclusively with recursively enumerable (r.e.) languages; accordingly, languages will be described using a fixed Gödel numbering of all r.e. languages of natural numbers. The underlying model is mainly learning from positive examples, that is, the learner processes an infinite sequence of numbers whose range equals the target language L. Any such sequence is called a text for L.

The main elements of Gold's model were later adapted to a wide variety of alternative learning paradigms [2, 7, 8]. Notably, many of these paradigm variations, such as behaviourally correct learning [3, 4, 7, 8] and vacillatory learning

[5, 8], impose some sort of strict convergence criterion on the learner, in the sense that the learner must never make a wrong conjecture after processing a sufficient amount of finite information about the target language. However, as Osherson, Stob and Weinstein [25, Section 7.5] observed, if a learning paradigm is viewed as a model of child language acquisition, then such a requirement on the learner cannot be justified merely on *a priori* grounds. It would thus not be unreasonable to consider other plausible learning paradigms in which the convergence requirement is relaxed. In this regard, Osherson, Stob and Weinstein [25, Exercise 7.5A] introduced *partial learning* as an alternative to the paradigm of "identification in the limit". According to this stronger learning model, a learner M is judged to have successfully identified a language L if and only if, on every text for L, M outputs exactly one index e infinitely often, and e is a Gödel number for L. Osherson, Stob and Weinstein observed that partial learning is so powerful that the whole class of r.e. languages is identifiable under this learning criterion. Yet their discovery did not spell the end of partial learning: the notion was later picked up, combined with the constraint of consistency[5], and studied comprehensively for the case of learning graphs of recursive functions [18]. This study uncovered quite an interesting connection between partial learning and Gold's more stringent model of learning in the limit, specifically when consistency is imposed in both models. It was shown that if the order of data presentation of a target recursive function is arbitrary, then consistent partial learning and consistent learning in the limit coincide. Moreover, the precise inference degrees of these various learning criteria were determined; a non-trivial fact established by way of these results was that successful consistent partial identification of recursive functions may, under specific conditions, depend on the order of data presentation.

A similar line of research was also extensively pursued in [9, 10, 12, 13]; the object of study in these papers was the class RE^A of all A-r.e. languages, where the partial learnability of RE^A was characterised by the Turing degree of A. Much of this research focussed on the partial learning model imposed with additional learning constraints such as confidence[6] [12], conservativeness[7] [9], reliability[8] [13] and consistency [13]; various blends of these constraints were also considered. Apart from revealing a new hierarchy of learnability notions, it was found that several variants of partial learning admit fairly elegant class structural characterisations. In particular, it was shown that the family of all partially conservatively learnable classes of r.e. languages has a characterisation that is analogous to the "tell-tale" characterisation obtained by Angluin [2] for every

[5] Consistency requires the learner, for any finite segment σ of any text, to output an index of a language that contains all the elements listed in σ [2].

[6] A confident partial learner outputs, on any text of any language, exactly one index infinitely often.

[7] In the context of partial learning, conservativeness requires the learner to output an index of a superset of the target language only if that index is the unique one that is output infinitely often.

[8] A partial learner is reliable if for every language L and every text for L, the learner only outputs indices of L infinitely often.

explanatorily learnable class of r.e. languages. Consistently partially learnable classes of r.e. languages may be described even more concisely: it was shown that a class of r.e. languages is consistently partially learnable if and only if it is a subclass of a uniformly recursive family [13]. This paper reviews much of the work done thus far on partial learning; each of the subsequent sections presents results under a particular learning scenario. Figure 1 summarises most of the paper's results on partial learning of recursively enumerable languages. The learning criteria are abbreviated as follows: BC for behaviourally correct learning [7], Ex for explanatory learning, Vac for vacillatory learning, Fin for finite learning, $Part$ for partial learning, $Conf$ for confidence, $Consv$ for conservativeness, $ClsCons$ for class-consistency (where the learner is required to be consistent only on valid texts for potential targets), $Cons$ for global consistency (where the learner is required to be consistent on all input texts), $EssClsCons$ and $EssCons$ for the "essential" versions of the two consistency models, $Prud$ for prudence, Rel for reliability, and $ClsPresv$ for the class-preserving versions of a model, requiring the learner to return only hypotheses that represent sets from the target class, on any input. For any learning criterion A, $A[\mathbb{K}]$ means learnability under criterion A relative to the oracle $\mathbb{K}$. A directed arc from criterion A to criterion B means that the collection of classes learnable under model A is contained in that learnable under model B. If there is no path from A to B, then the collection of classes learnable under model A is not contained in that learnable under model B.

2 Basic Definitions and a First Result on Partial Learning

Gold's model of *identification in the limit* (also known as *explanatory learning*), as well as the *behaviourally correct learning* and *partial learning* models, are formalised as follows. First, define a text $T = a_0, a_1, a_2, \ldots$ of a language L to be a sequence containing all the members of L in arbitrary order; this sequence may contain repeated elements of L and also pause symbols (denoted by #). The learner, denoted by M, reads a text $a_0, a_1, a_2, \ldots$ of a language L and produces a sequence $e_0, e_1, e_2, \ldots$ of hypotheses; that is to say, if M is modelled as a recursive function mapping $(\mathbb{N} \cup \{\#\})^*$ into $\mathbb{N}$, then for all n, $M(a_0 a_1, \ldots a_n) = e_n$. Each hypothesis e_n describes M's conjecture W_{e_n}, where $W_0, W_1, W_2, \ldots$ is a fixed Gödel numbering of all r.e. languages. Furthermore, $\varphi_0, \varphi_1, \varphi_2, \ldots$ is a Gödel numbering of all partial-recursive functions. The range of a (finite or infinite) sequence σ is denoted by $range(\sigma)$.

Definition 1. Assume that a learner M processes a text of a language L and outputs a sequence $e_0, e_1, \ldots$ of hypotheses.

(I) Gold [15] defined that the learner M *explanatorily* learns L from the text iff there is a k such that $e_k = e_h$ for all $h \geq k$ and $W_{e_k} = L$.

(II) Case and Lynes [7] defined that the learner M *behaviourally correctly* learns L from the text iff $W_{e_k} = L$ for almost all k.

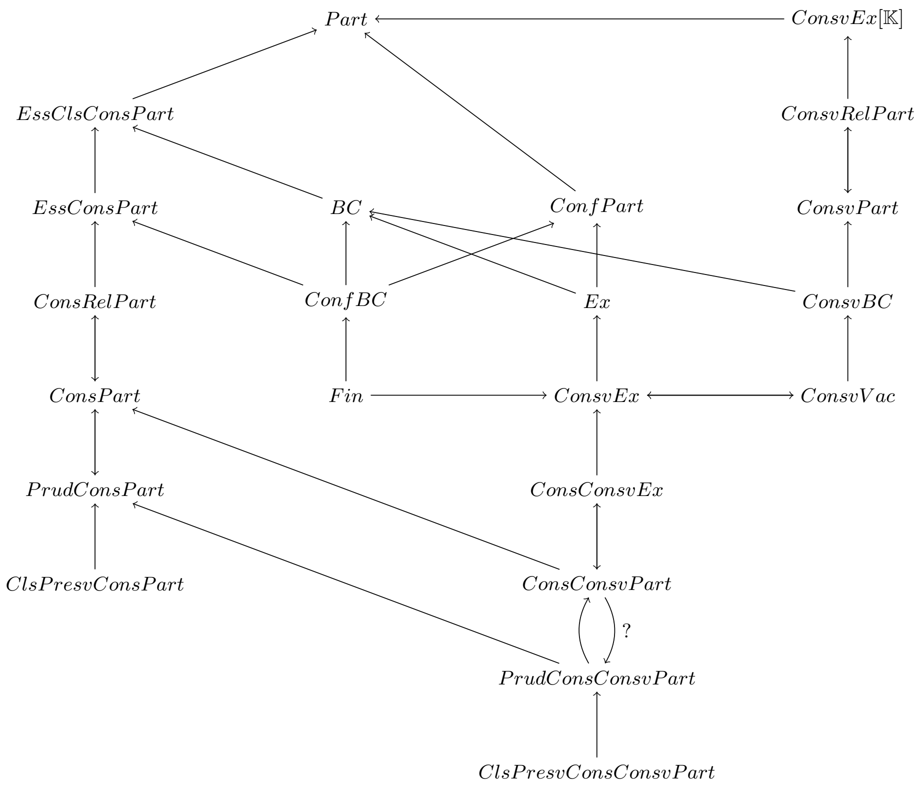

Fig. 1. Learning hierarchy

(III) Case [5] defined that the learner M *vacillatorially* learns L from the text iff M *behaviourally correctly* learns L and the set $\{e_0, e_1, e_2, \ldots\}$ is finite.

(IV) Osherson, Stob and Weinstein [25] defined that M *partially* learns L from the text iff there is exactly one e with $\exists^\infty k\,[e_k = e]$ and this e satisfies $W_e = L$.

Furthermore, a class $\mathcal{S}$ of r.e. languages is learnable under criterion I iff there is a recursive learner that learns every $L \in \mathcal{S}$ from every text with respect to the criterion I.

The class of all finite languages is an example of an explanatorily learnable class; the learner for it just conjectures at each time a canonical index for the set of elements seen so far; it is obvious that this set grows only finitely often and after the last element of the language to be learnt has been seen the learner stabilises on the index of the elements of this language. However, given a non-recursive r.e. language like the halting problem $\mathbb{K}$, the class of all unions $\mathbb{K} \cup D$ with D being finite is in principle learnable by the same algorithm; however, the learner cannot distinguish between irrelevant data (which are members of

$\mathbb{K}$) and relevant data (which require an update), therefore one would have to update in each step and conjecture the union of $\mathbb{K}$ and all data-items seen so far. This learner is then behaviourally correct and one can show that there is no explanatory learner for this class. Gold [15] showed that the class consisting of $\mathbb{N}$ and all of its finite subsets is not explanatorily learnable; this proof directly transfers to behaviourally correct learning and thus this is the standard example for a class which is neither explanatorily nor behaviourally correctly learnable.

Osherson, Stob and Weinstein [25] noted that partial learning is more powerful than the stricter notions of explanatory and behaviourally correct learning, for it allows to learn the whole class of r.e. languages and therefore also to learn any class of r.e. languages.

Theorem 2 (Osherson, Stob and Weinstein [25], Exercise 7.5A). *The class of all r.e. languages is partially learnable.*

Proof. One should first note that partial learners can often be built by not specifying when a hypothesis is output, but only how often it is output and the fine-tuning of these conjectures can then be realised with maintaining a queue. So the basic principle would be that one makes a learner M which outputs the hypothesis e at least n times iff there is a time $t > n$ such that for each $m < n$, m appears in the first t data items iff m is enumerated into the e-th r.e. language W_e within t time steps. If a hypothesis is correct, then it will eventually qualify for each n and be output infinitely often; if a hypothesis is incorrect, then it will be output only a finite number of times: Either there is a data item m which is not in W_e and if n is so large that $n > m$ and m has appeared within the first n data items, then e will not be output n times or more. Or there is an element $m \in W_e$ which does not occur in the input and for every $n > m$ being so large that $m \in W_e$ within n steps, it holds that e will not be output n or more times. The only remaining problem with this approach is that M outputs all correct indices infinitely often.

One way to cure this would be to work with one-one numberings rather than acceptable numberings. Another way to cure this is to use padding. Padding means that there is a recursive one-one function *pad* such that $W_{pad(e,k)} = W_e$ for all e, k; padding can be done in all acceptable numberings. Now one can make a new learner N such that N outputs $pad(e, k)$ at least n times iff there is a time $t > n$ such that the simulated learner M has so far output e at least n times and furthermore output all the indices $0, 1, \ldots, e-1$ together exactly k times. In the case that the indices below e are all output only finitely often, say altogether k times, and in the case that furthermore M outputs e infinitely often then N will output $pad(e, k)$ infinitely often and all other indices only finitely often. This padding technique is quite useful and has been employed in many results in the field. ∎

3 Confidence

Gao, Stephan, Wu and Yamamoto [12] introduced *confidence* to moderate the full strength of partial learning. Confidence was originally defined by Osherson, Stob and Weinstein [25, Section 4.6.2] as the requirement that the learner must, on every text for any (possibly non-r.e.) language, converge to some fixed hypothesis; in the partial learning model, this was relaxed to the condition that the learner must output on every text exactly one index infinitely often.

Definition 3 (Gao, Stephan, Wu and Yamamoto [12]). A partial learner M is *confident* iff for every text T for any language (including the case where the language is not r.e.) there is exactly one index e that is output infinitely often by M.

A weaker notion of behaviourally correct learning, known as *confident behaviour ally correct* learning, was introduced in [11].

Definition 4. A class $\mathcal{S}$ of r.e. languages is confidently behaviourally correctly learnable iff there is a recursive learner M that behaviourally correctly learns every $L \in \mathcal{S}$ and has the property that it outputs on every language L, even non-r.e. ones, a sequence of conjectures $e_0, e_1, e_2, \ldots$ that stabilises semantically, that is, $W_{e_{n+1}} = W_{e_n}$ holds for almost all n.

That confidence is indeed a proper restriction on partial learning is witnessed by the class of all cofinite languages, which cannot be confidently partially identified [12, Theorem 12]. In addition, behaviourally correct learnability does not in general imply confident partial learnability, as the following example demonstrates. However, confident behaviourally correct learnability is a more restrictive notion which implies confident partial learnability.

Example 5. The class $\{\mathbb{K} \cup D : D \text{ is finite}\}$ is behaviourally correctly learnable but not confidently partially learnable.

Conversely, it is known [13, Theorem 7] that any class consisting of all finite languages and one infinite language is confidently partially learnable but not behaviourally correctly learnable. Somewhat more surprising, however, is the fact that explanatory learning implies confident partial learnability [10]; the same result holds even when the criterion of explanatory learning is weakened to explanatory learning with at most c anomalies (whereby the learner is permitted to converge to a hypothesis that disagrees with the target language on at most c numbers).

Theorem 6 (Gao, Stephan and Zilles [13], Theorem 8). *If a class of r.e. languages is explanatorily learnable with at most c anomalies, then it is also confidently partially learnable.*

Proof. Assume that M is a learner which explanatorily learns a class S with up to c anomalies. Without loss of generality, M chooses at every mind change a hypothesis larger than the previous one, so that M never vacillates between two hypotheses forever.

The new learner N maintains in each stage s a sequence $e_{0,s}, e_{1,s}, \ldots, e_{c+1,s}$ of hypotheses with $e_{0,s} = 0$ for all s and $e_{1,s}$ being the current hypothesis of M.

Furthermore, $e_{k+1,s}$ is obtained from $e_{k,s}$ by fixing the least conjectured anomaly x. In the case that x is observed in the data so far and $W_{e_{k,s},s}$ does not contain x, then $W_{e_{k+1,s}} = W_{e_{k,s}} \cup \{x\}$; in the case that x has not yet been observed in the data and $x \in W_{e_{k,s},s}$, then $W_{e_{k+1,s}} = W_{e_{k,s}} - \{x\}$. The indices computed by this padding are without loss of generality larger than x. In the case that no such x exists then $e_{k,s} = e_{k+1,s} = \ldots = e_{c+1,s}$.

The hypothesis number s of the new learner N is now $e_{k,s}$ for the maximal k such that $e_{0,s} = e_{0,s+1} \wedge e_{1,s} = e_{1,s+1} \wedge \ldots \wedge e_{k,s} = e_{k,s+1}$. The rationale behind this is that if M does not converge then N outputs $e_{0,s}$ infinitely often, that is, N outputs the fixed value 0 infinitely often. If M converges and the number of anomalies is k then these anomalies are found eventually and the sequence of the $e_{k+1,s}$ converges to a value e_{k+1} which is the index of the language to be learnt. This e_{k+1} is furthermore output infinitely often for the following reason: either there are infinitely many candidates for an anomaly $k+1$ which are each eventually replaced at some stage s by another candidate for that anomaly causing $e_{k+2,s}$ to be different from $e_{k+2,s+1}$ and causing $e_{k+1,s} = e_{k+1}$ to be output at that stage s or there are infinitely many s where $e_{k+1} = e_{k+1,s} = e_{c+1,s}$ and in these cases e_{k+1} is output as well. Thus N does also in this case converge to a right index. Furthermore, there is, independent of the choice of the target language, a maximal k such that the sequences of the $e_{0,s}$, $e_{1,s}$, …, $e_{k,s}$ converge and the limit of this last sequence is then the only hypothesis which is output infinitely often. ∎

Interpreted as a model of language learning, the relation between identification in the limit and confident partial learning provides evidence that learners that are powerful enough to explanatorily learn a class of languages are also able, in a quite broad sense, to isolate a single hypothesis infinitely often in the case that they diverge, for example on any text for a nonnatural language. Although confident partial learning may appear at first sight to be a bit restrictive, one can show that it admits the following general characterisation.

Theorem 7 (Gao, Stephan and Zilles [13], Theorem 9). *A class $\mathcal{S}$ of r.e. languages is confidently partially learnable iff there is a recursive learner M that outputs on every text exactly one index infinitely often, and whenever $L \in \mathcal{S}$ and T is a text for L, then M on T does not output infinitely often any index smaller than the smallest r.e. index of L.*

As a corollary of this result, it follows that confident partial learning has the mathematically appealing property of being closed under finite unions – a property that is also satisfied by confident vacillatory learning as well as confident behaviourally correct learning. The intuitive explanation for this deduction is that if M_1 and M_2 are two confident partial learners of the classes $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively, then one can define a further learner N that outputs on any text T the pair $\langle d_1, d_2 \rangle$ at least n times iff M_1 outputs d_1 and M_2 outputs d_2 at least n times. Note that N outputs $\langle d_1, d_2 \rangle$ infinitely often iff M_1 outputs d_1 and M_2 outputs d_2 infinitely often. Now, if $L \in \mathcal{S}_1 \cup \mathcal{S}_2$ then there is an index e of L such that $e \leq d_1$ or $e \leq d_2$; it follows by the properties of Cantor's pairing function that $e \leq \langle d_1, d_2 \rangle$ and therefore N outputs an upper bound of e infinitely often. Hence $\mathcal{S}_1 \cup \mathcal{S}_2$ is confidently partially learnable.

The study of confident partial learning relative to oracles illustrates quite an interesting interplay between inductive inference and recursion theory. For example, consider the class $\mathcal{C}$ of graphs of recursive functions f_e satisfying the following conditions:

- $f_e(x) = 0$ for all $x < e$;
- $f_e(x) = 1$ for $x = e$;
- $f_e(x) \downarrow = \varphi_e(x) \downarrow$ for almost all x.

For a given recursive function f, the sequence $(0, f(0)), (1, f(1)), (2, f(2)), \ldots$ is called the *canonical text* for f. While it is not exceedingly difficult to show that $\mathcal{C}$ is behaviourally correctly learnable, the problem of showing that $\mathcal{C}$ cannot be confidently partially learnt from canonical texts is much more involved. Gao and Stephan [10, Theorem 4] utilise for their proof a relativised version of the Low Basis Theorem of Jockusch and Soare [19]. They eventually show that every oracle which permits to confidently partially learn $\mathcal{C}$ is high$_2$. Here the classification of low, high, low$_2$ and high$_2$ is included for the reader's convenience.

Definition 8. An oracle A is high iff $A' \geq_T \mathbb{K}'$; an oracle A is high$_2$ iff $A'' \geq_T \mathbb{K}''$; an oracle A is low iff $A' \leq_T \mathbb{K}$; an oracle A is low$_2$ iff $A'' \leq_T \mathbb{K}'$.

The proof in [10] that $\mathcal{C}$ is not confidently partially learnable with respect to canonical texts establishes a deeper fact: there is a behaviourally correctly learnable class of graphs of recursive functions that is confidently partially learnable only relative to high$_2$ oracles. Moreover, one can show that there do indeed exist high$_2$ oracles that permit the class *REC* of all graphs of recursive functions to be confidently partially learnt; on the other hand, there are also high$_2$ oracles that are not powerful enough to allow *REC* to be confidently partially learnt. These results complement the characterisation theorem obtained by Adleman and Blum [1] many years earlier for the class of oracles relative to which *REC* is explanatorily learnable.

Theorem 9 (Adleman and Blum [1]). *An oracle A is high iff there is an A-recursive explanatory learner for the class REC of the graphs of all recursive functions.*

As explanatory learners can be transformed into confident partial learners, one direction of the result of Adleman and Blum transfers to confident partial learning; the other direction is more complicated and needs a weakening to high$_2$ though there is no characteristation.

Theorem 10 (Gao and Stephan [10]). *If A permits to confidently partially learn REC then A is high$_2$. Furthermore, the oracles which permit to confidently partially learn REC are all high ones plus some but not all of those which are high$_2$ but not high.*

It should be noted that there is even a subclass $G \subseteq REC$ such that G is behaviourally correctly learnable (without any oracle) and such that whenever G is confidently partially learnable relative to A then A is high$_2$.

4 Reliability

Reliable partial learning was introduced in [12] as a counterpart of confident partial learning — while a confident learner always converges to a hypothesis (according to the given learning criterion), the reliable learner only converges to a hypothesis iff it is correct. Originally defined by Minicozzi [24] as the requirement that the learner identifies all languages on whose texts it converges syntactically, the partial learning analogue stipulates that the learner must, on every text T for any (possibly non-r.e.) language L, output only indices of L infinitely often. For explanatory learning from texts, this notion is very restrictive and only permits to learn classes of finite languages. Only in the framework of partial learning, the notion of reliability becomes interesting for learning from text.

Definition 11 (Gao, Stephan, Wu and Yamamoto [12]). A partial learner M is *reliable* iff for every language L and on every text for L, M only outputs indices of L infinitely often.

The empirical justification for defining such a criterion is based on the observation that children may implement a reliable partial learning function, so that they eventually abandon any initially conjectured hypothesis that is incorrect, while returning infinitely often to exactly one correct hypothesis when presented with a text for a language to be learnt. The usual algorithm to partially learn all r.e. languages is already reliable, that is, an incorrect hypothesis is always output only finitely often. Thus the class of all r.e. languages is already reliably partially learnable and so it makes sense to go into the direction which asks: which classes more comprehensive than the class of all r.e. languages are also partially reliably learnable.

Gao, Stephan, Wu and Yamamoto [12, Theorem 16] investigated this question and showed that every uniformly $\mathbb{K}$-recursive class is reliably partially learnable using a padded version of the given indexing as hypothesis space. Furthermore, they considered for an oracle A the class RE^A of all A-r.e. languages from an acceptable numbering as a hypothesis space; here "acceptable" can be understood in the strict sense that for every numbering of A-r.e. languages there is a recursive function f which translates the indices of this numbering into equivalent indices of the given acceptable numbering.

Theorem 12 (Gao, Stephan, Wu and Yamamoto [12], Corollary 20). *The class of all A-r.e. languages is reliably partially learnable using A-r.e. indices iff A is low.*

The corresponding result for the class of all A-recursive languages reads as follows.

Theorem 13 (Gao, Stephan, Wu and Yamamoto [12], Theorem 21). *The class of all A-recursive languages is reliably partially learnable iff $A \leq_T \mathbb{K}$ and A is low$_2$.*

These results stand in contrast to the case of non-reliable partial learners where there is even an uncountable class of oracles A such that the class of all A-recursive languages can be learnt by a recursive partial learner. So the requirement of reliability is a real restriction in this general setting.

5 Consistency

Angluin [2] proposed *consistency* as a constraint on conjectures for r.e. languages. A learner M is said to be consistent if and only if it always generates at least all the data it has seen so far.

Definition 14 (Angluin [2]). A learner M is *consistent* iff $range(\sigma) \subseteq W_{M(\sigma)}$ for all data sequences σ.

Consistency is a fairly intuitive learning constraint, for it is conceivable that no learner will hold on to a hypothesis that is falsified by the available data. It is also known [25, Proposition 4.3.3A] that only recursive languages can be consistently identified. Going further, one may ask whether the limitations of consistent explanatory learning may be overcome by relaxing the requirement to partial learning. This, however, is not the case: it was proven [13, Theorem 18] that consistently partially learnable classes of r.e. languages must be subclasses of uniformly recursive families. It was shown, moreover, that a consistently partially learnable class of r.e. languages can even be *prudently* consistently partially learnt.

Definition 15 (Osheron, Stob and Weinstein [25]). A learner M is *prudent* iff it learns the class $\{W_{M(\sigma)} : \sigma \in (\mathbb{N} \cup \{\#\})^*\}$. In other words, M learns every language for which it outputs a conjecture, on some input data.

Theorem 16 (Gao, Stephan and Zilles [13], Theorem 18). *The following statements are equivalent for a class $\mathcal{C}$ of r.e. languages:*

(I) *$\mathcal{C}$ is a subclass of a uniformly recursive family;*
(II) *$\mathcal{C}$ is consistently partially learnable;*
(III) *$\mathcal{C}$ is prudently consistently partially learnable;*
(IV) *$\mathcal{C}$ is prudently consistently partially learnable using a uniformly recursive hypothesis space.*

The learnability of classes of recursive functions from canonical text versus arbitrary text found quite a lot of attention. For example, Grieser [16] solved an important open problem by showing that there is a class of functions which has a total recursive consistent learner using canonical text while it does not have a class-consistent learner succeeding on arbitrary texts. Jain and Stephan [18] studied consistent partial learning of recursive functions and in particular which classes of recursive functions can be consistently partially learnt from canonical text using the oracle A. The answer is that there are only two cases: If A has hyperimmune Turing degree then these are all classes of recursive functions and if A has hyperimmune-free Turing degree these are those classes which can be learnt using any oracle (or none).

Theorem 17 (Jain and Stephan [18], Theorem 6 and Corollary 7). *Let A be an oracle. If A is hyperimmune then some A-recursive learner can consistently partially learn the class REC from canonical text. If A is not equivalent to any hyperimmune set and some A-recursive learner can consistently partially learn $\mathcal{S}$ from canonical text, then some recursive learner can do the same.*

Jain and Stephan [18] also compared the inference degrees of learning in the limit to those of consistent partial identification with respect to arbitrary texts (that is, texts for recursive functions in which the order of the elements can be arbitrary) for the case of learning recursive functions. They showed that in the case of learning recursive functions relative to a given oracle A from arbitrary texts, consistent learning in the limit is just as powerful as consistent partial identification. Moreover, they provided a characterisation of all classes of recursive functions that are consistently partially identifiable relative to a given oracle A.

Theorem 18 (Jain and Stephan [18], Theorem 8). *For every oracle A and every class $\mathcal{C}$ of recursive functions, the following conditions are equivalent:*

(I) *$\mathcal{C}$ is consistently partially learnable relative to A from arbitrary texts;*
(II) *$\mathcal{C}$ is consistently explanatorily learnable relative to A from arbitrary texts;*
(III) *There is a strictly increasing A-recursive function g^A such that every $f \in \mathcal{C}$ has an index e which satisfies $\Phi_e(x) \leq g^A(\max(\{x, f(x)\}))$ for almost all x, where Φ denotes a fixed Blum complexity measure such that for all e, $\Phi_e(x)$ is the complexity assigned to the computation of $\varphi_e(x)$.*

It is not immediately obvious whether or not the preceding theorem still holds if one relaxes the criteria of consistent partial learning and explanatory learning to *class consistent* partial learning and *class consistent* explanatory learning respectively; in the class consistent versions of these criteria, the learner is only required to be consistent on texts for recursive functions in the given target class.

Grieser [16] had already observed that the order of data presentation is very important for consistent explanatory learning of functions; Jain and Stephan [18] investigated how much these order constraints influence the consistent partial learnability of a class.

Theorem 19 (Jain and Stephan [18], Theorem 12).

(I) *If A is not high, then any class of $\{0, 1\}$-valued recursive functions which is consistently partially learnable relative to A from arbitrary texts is consistently partially learnable (unrelativised) from canonical texts.*
(II) *If A is r.e. and not recursive, then there is a class of recursive functions that is consistently partially learnable relative to A from arbitrary texts but not consistently partially learnable (unrelativised) from canonical texts.*

It is well-known that restricting the requirement of consistency only to members of a class makes the criterion more general, for example the class $\{\mathbb{K}\}$ is class-consistently explanatorily learnable but not consistently partially learnable. A further generalisation is linked to the model of partial learning: here the consistency is not required for all hypotheses but only from some time point onwards; note that explanatory and behaviourally correct learners satisfy this constraint of "essential class-consistent learning" by definition. The corresponding criterion is defined for partial learning as below and was studied only recently [10, 13].

Definition 20 (Gao, Stephan and Zilles [13]). A learner M *essentially class consistently* partially learns $\mathcal{S}$ iff M partially learns $\mathcal{S}$ and for every $L \in \mathcal{S}$ and every text T for L and almost all $\sigma \preceq T$, M outputs on σ an index e with $range(\sigma) \subseteq W_e$. A learner M *essentially globally consistently* partially learns $\mathcal{S}$ iff M partially learns $\mathcal{S}$ and for every L (not necessarily in $\mathcal{S}$) and every text T for L and almost all $\sigma \preceq T$, M outputs on σ an index e with $range(\sigma) \subseteq W_e$.

It was found [13, Theorem 15] that essentially class consistent partial learnability is stronger than behaviourally correct learnability, providing yet another link between partial learning and the traditional models of learning in which the learner is required to almost always output correct hypotheses. Moreover, the three constraints of essential class consistency, confidence and conservativeness, are fairly robust when they are compared pairwise; one can show that any two of these notions have incomparable learning strengths. The foregoing results [13, Theorem 15, Corollary 21, Theorem 22, Example 25, Theorem 27] are summed up below.

Theorem 21 (Gao, Stephan and Zilles [13]).

(I) *Every behaviourally correctly learnable class is essentially class consistently partially learnable.*
(II) *Every confidently behaviourally correctly learnable class is essentially global consistently partially learnable.*
(III) *The notions of essentially class consistent partial learning and confident partial learning and conservative partial learning are pairwise incomparable.*

It has been shown [11, Theorem 27] that a confidently behaviourally correctly learnable class of recursive functions is essentially globally consistently partially learnable – a result that is parallel to Theorem 21(I). An analogous result holds for classes of r.e. languages.

Theorem 22. *Every confidently behaviourally correctly learnable class of r.e. languages is essentially globally consistently partially learnable.*

Proof. Assume that M is a learner that confidently behaviourally correctly learns a class $\mathcal{S}$ of r.e. languages. Theorem 21(I) implies that there is a learner N such that N outputs on any text for every language L a sequence of conjectures $d_0, d_1, d_2, \ldots$ satisfying the following conditions.

– There is exactly one index e such that $\exists^\infty m[d_m = e]$ and $\forall^\infty m[W_{d_m} = W_e]$ both hold.

– If $L \in \mathcal{S}$, then $W_e = L$.

One can further define a learner P as follows. P maintains a list I of indices for bookkeeping. I is defined in stages, starting with $I_0 = \mathbb{N}$. On input σ, P checks whether it can compute the least $n \leq |\sigma|$ such that $n \in I_{|\sigma|-1}$ and $\{0, 1, \ldots, n\} \cap range(\sigma) \subseteq W_{d_n, |\sigma|}$ both hold. If such an n exists, then P outputs d_n and sets $I_{|\sigma|} = I_{|\sigma|-1} - \{n\}$. If no such n exists, then P outputs a grammar for range(σ) and sets $I_{|\sigma|} = I_{|\sigma|-1}$. P essentially globally consistently partially learns $\mathcal{S}$ for the following reasons.

– If P is fed with a text for some $L \in \mathcal{S}$, then there is exactly one index e such that $\exists^\infty m[d_m = e]$ and $W_e = L$ both hold. Thus there are infinitely many n such that $d_n = e$ and $\{0, 1, \ldots, n\} \cap range(\sigma) \subseteq W_{d_n}$ hold. On the other hand, for every $f \neq e$, there are only finitely many m for which $d_m = f$, and so P will output f only finitely often. Moreover, since it holds for almost m that $d_m = e$, P must be essentially globally consistent.
– Suppose that P is fed with a text T for a language L such that $\forall^\infty m[W_{d_m} = W_h]$ holds. If $L \not\subseteq W_h$, then there is an n such that $\{0, 1, \ldots, k\} \cap L \not\subseteq W_{d_k}$ holds for all $k \geq n$, and so P will output $range(T[k])$ for almost all k. If $L \subseteq W_h$, then P will almost always output either a superset of L, or the range of the current input. Hence P is essentially globally consistent.

This establishes that P is an essentially globally consistent partial learner of $\mathcal{S}$. ▮

Like confident partial learning, essentially globally consistent partial learning is closed under finite unions – a property that fails to hold for consistent explanatory learning. Thus partial learning may be quite a liberal criterion, even when it is combined with various learning constraints.

Theorem 23 (Gao, Stephan and Zilles [13], Corollary 16 and Theorem 17). *The union of essentially globally consistently partially learnable classes is essentially globally consistently partially learnable. This does not apply to essentially class consistent partial learning.*

6 Conservativeness

Another learning constraint often considered in parallel to consistency is *conservativeness*, which was also introduced by Angluin [2]. She defined a learner M to be conservative if and only if, for all sequences σ, τ with $\sigma \prec \tau$, $M(\tau) \neq M(\sigma)$ holds only if range(σ) $\not\subseteq W_{M(\sigma)}$. Loosely interpreted, this means that a conservative learner never revises its hypothesis until it has seen data not generated by its original conjecture. Mazurkewich and White [23] provide evidence that children do indeed overgeneralise, specifically with respect to the English dative alternation, that is, alternations between a prepositional indirect-object construction and a double-object construction; for example, the pair of sentences "John gave the book to Fred" and "John gave Fred the book." Hence there is some empirical justification for assuming that learners may not always be conservative. Moreover, conservativeness is a proper restriction on the learning potential of explanatory learners: Angluin [2, Theorem 4] established the existence of a uniformly recursive family that is explanatorily learnable but cannot be conservatively explanatorily learnt.

Gao, Jain and Stephan [9] studied conservativeness in the context of partial learning. The partial learning variant of conservativeness is defined as follows. Again, suppose that M partially learns a class $\mathcal{S}$ of r.e. languages.

Definition 24 (Gao, Jain and Stephan [9]). A partial learner M is *conservative* iff for every language $L \in \mathcal{S}$ and every text T for L the learner outputs exactly one index of L (which is also output infinitely often) and no index of a proper superset of L.

It is quite a curious fact that unrelativised conservative partial learning is weaker than learning in the limit with an oracle for the diagonal halting problem $\mathbb{K}$. More precisely, one has the following theorem.

Theorem 25 (Gao, Jain and Stephan [9], Theorem 18). *If $\mathcal{S}$ is conservatively partially learnable, then $\mathcal{S}$ is explanatorily learnable relative to $\mathbb{K}$. If $\mathcal{S}$ consists of only infinite languages and is explanatorily learnable relative to $\mathbb{K}$, then $\mathcal{S}$ is also conservatively partially learnable.*

The condition in the second part of the above theorem, that the class $\mathcal{S}$ consists of only infinite languages, cannot be dropped; this is implied by the next result.

Example 26 (Gao, Jain and Stephan [9], Example 11). If A is not recursive, then there is a class $\mathcal{S}$ that is conservatively explanatorily learnable relative to A but not conservatively partially learnable (unrelativised).

The overall picture of conservativeness in the partial learning model, as well as in models where some sort of hypothesis convergence is enforced – that is, explanatory, vacillatory, and behaviourally correct learning, is summed up in the following theorem.

Theorem 27 (Gao, Jain and Stephan [9], Theorem 7). *Let $\mathcal{S}$ be a class of r.e. languages. Then the following statements hold:*

- (I) *$\mathcal{S}$ is conservatively explanatorily learnable iff $\mathcal{S}$ is conservatively vacillatorily learnable;*
- (II) *If $\mathcal{S}$ is conservatively vacillatorily learnable, then $\mathcal{S}$ is conservatively behaviourally correctly learnable;*
- (III) *The converse of* (II) *is not true in general;*
- (IV) *If $\mathcal{S}$ is conservatively behaviourally correctly learnable, then $\mathcal{S}$ is conservatively partially learnable;*
- (V) *The converse of* (IV) *is not true in general.*

Angluin [2, Theorem 1] introduced the notion of a *tell-tale* set of a language L to be learnt: it is a finite subset E_L of L such that no language L' in the class of languages to be learnt satisfies $E_L \subseteq L' \subset L$. Based on this notion, she completely characterised all uniformly recursive families that are explanatorily learnable. Formally, the result reads as follows.

Theorem 28 (Angluin [2], Theorem 1). *A uniformly recursive family $\{L_d : d \in \mathbb{N}\}$ of languages is explanatorily learnable iff there exists a uniformly recursively enumerable family $\{H_d : d \in \mathbb{N}\}$ such that for all e,*

(I) *H_e is finite;*
(II) *$H_e \subseteq L_e$;*
(III) *for all $d \in \mathbb{N}$, if $H_e \subseteq L_d$ then L_d is not a proper subset of L_e.*

Subsequent investigations by Lange, Kapur and Zeugmann in various papers [21, 22, 29] dealt with several variants of explanatory learning and led to the discovery of a similar tell-tale characterisation of *conservative explanatory learning* in the setting of uniformly recursive families. The characterisation needs the notion of a canonical index e of a finite set D:

$$e \text{ is the canonical index of a finite set } D \Leftrightarrow e = \sum_{x \in D} 2^x.$$

The notion D_e denotes the finite set with canonical index e. The differences to Angluin's tell-tale criterion above are that here the finite tell-tales sets are not enumerated but given by their canonical indices and that in some cases it is furthermore necessary to change the hypothesis space; these differences seem to be essential for characterisations of conservative learning.

Theorem 29 (Lange and Zeugmann [21], Theorem 5). *Let $\mathcal{L} = \{L_d : d \in \mathbb{N}\}$ be a uniformly recursive family. Then $\mathcal{L}$ is conservatively learnable iff there is a hypothesis space $\mathcal{G} = \{G_d : d \in \mathbb{N}\}$ of uniformly recursive languages and a recursive function f such that for all e,*

(I) *there is a $d \in \mathbb{N}$ with $G_d = L_e$;*
(II) *$D_{f(e)} \subseteq G_e$;*
(III) *for all $d \in \mathbb{N}$, if $D_{f(e)} \subseteq G_d$, then G_d is not a proper subset of G_e.*

The preceding characterisation of conservative learning was later generalised by de Jongh and Kanazawa [20, Theorem 28] to the case of uniformly r.e. families, although their formulation of the tell-tale condition was somewhat more complicated. A characterisation similar to the one of Lange and Zeugmann was obtained for the notion of conservatively partially learnable classes of r.e. languages [9] – like the criterion of Lange and Zeugmann, it uses canonical indices rather than r.e. indices in Angluin's original characterisation for explanatory learning; furthermore, a hypothesis space is chosen by enumerating the family of the W_{j_e} as explained below.

Theorem 30 (Gao, Jain and Stephan [9], Theorem 21). *A class $\mathcal{S}$ is conservatively partially learnable iff there is a recursive sequence of pairs (i_e, j_e) such that*

(I) *For all $L \in \mathcal{S}$ there is an e with $L = W_{j_e}$;*
(II) *For all e, $D_{i_e} \subseteq W_{j_e}$;*
(III) *For all d, e, if $D_{i_e} \subseteq W_{j_d} \subset W_{j_e}$, then $W_{j_d} \notin \mathcal{S}$.*

7 Combining Reliability with Other Notions

Whilst there are classes of r.e. languages that cannot be partially learnt if both consistency and confidence are enforced, one can nevertheless always impose consistency in tandem with reliability. Central to this observation is the characterisation in Theorem 16 which says that a class $\mathcal{S}$ of r.e. languages is consistently partially learnable iff it is a subclass of a uniformly recursive family. The next theorem shows that for consistent partial learning, one can add in reliablility in the same way as one can do for partial learning.

Theorem 31. *A class $\mathcal{S}$ of r.e. languages is consistently partially learnable iff it is consistently reliably partially learnable.*

Proof. As was shown in the proof of [13, Theorem 18], a consistently partially learnable class $\mathcal{S}$ of r.e. languages is a subclass of some uniformly indexed family $\mathcal{S}' = \{L_1, L_2, L_3, \ldots\}$; without loss of generality, it may be assumed that $\mathcal{S}'$ is one-one and contains all the cofinite languages. Define a learner M as follows. For a text T, M on $T(0)T(1)\ldots T(n)$ finds the least pair $\langle i, j\rangle$ such that $j \leq n$, $|\{m : m < n \wedge M(T(0)T(1)\ldots T(m)) = i\}| = j$ and

$$L_i \cap \{0, 1, \ldots, j\} \subseteq \{T(0), T(1), \ldots, T(n)\} \subseteq L_i \cup \{\#\}. \quad (*)$$

Having identified the pair $\langle i, j\rangle$, M conjectures L_i. Condition $(*)$ ensures that M is always consistent. Now if $\text{range}(T) \neq L_i$, then either $L_i \cap \{0, 1, \ldots, j\} \not\subseteq \text{range}(T)$ for some j, or $\text{range}(T) \not\subseteq L_i$. Hence there is a least j for which the pair $\langle i, j\rangle$ will never be a candidate at sufficiently large stages; this implies, in particular, that M will not conjecture L_i more than j times, and so it is a reliable learner. Furthermore, if $L_i = \text{range}(T)$, then for every j, M will identify the pair $\langle i, j\rangle$ at some stage, since it holds that for all $\langle i', j'\rangle < \langle i, j\rangle$, $\langle i', j'\rangle$ can only qualify at most once. Consequently, M conjectures L_i infinitely often, and so it reliably consistently partially learns $\mathcal{S}'$. ∎

In contrast to this, it follows from the definition that reliability and confidence cannot be combined; the reason is that on texts for non-r.e. languages, reliability requires the learner not to output any hypothesis infinitely often while confidence requires the learner to output exactly one hypothesis infinitely often. One can, however, make the following combination of confidence and reliability; note that although the learner is no longer confident in the precise sense of the word, it can still be made confident by the usual padding trick.

Remark 32. Call a learner N to be confident and weakly reliable iff N satisfies the following conditions:

- N outputs on each text T either one or two hypotheses infinitely often;
- all hypotheses output infinitely often are for the same language L;
- the number of hypotheses output infinitely often is exactly one iff L is the language to be learnt.

Given now a standard confident learner M, one can obtain such an N from M using the following method: N outputs $pad(e, 0)$ at least n times iff M outputs e at least n times; N outputs $pad(e, k+1)$ at least n times iff (i) N outputs $pad(e, 0)$ at least n times and (ii) there is a time $t > n$ such that for all $m < k$, m occurs within the first t data-items iff m is enumerated into W_e within t steps and the same connection fails for $m = k$. So every confident learner can be made confident and weakly reliable.

For conservative learning, the next proposition shows that it can always be combined with reliable learning unless further constraints are postulated.

Proposition 33. *Every conservative partial learner M can be replaced by a conservative reliable partial learner N learning the same languages.*

Proof. The main idea is that one can make the translation by just filtering the stream of M's hypotheses with some consistency checks between data and hypotheses. So N outputs a hypothesis e on a text T at least n times iff M outputs e on T at least n times and there exists a $t > n$ such that for all $m < n$, m occurs within the first t data-items of T iff m is enumerated into W_e within t steps. One can easily verify the following facts:

- N outputs only hypotheses conjectured by M and therefore the conservativeness of M transfers to the conservativeness of N; furthermore, N does not output any hypothesis more frequently than M;
- If M outputs on a text T for L a hypothesis e for L infinitely often then so does N, as each repetition of the hypothesis e will eventually pass the filter;
- If M outputs on a text T for L an incorrect hypothesis e infinitely often then it will pass the filtering only finitely often and thus N is reliable.

This completes the proof. ∎

Osherson, Stob and Weinstein [25, Proposition 4.6.1A] demonstrated via a locking sequence argument that any class that is reliably explanatorily learnable contains only finite languages. In contrast to this, a consistent, conservative and reliable partial learner can identify infinite languages. One may take, for example, the class $\{\mathbb{N}\}$: On any text, a consistent, conservative and reliable partial learner M outputs a canonical index for $\mathbb{N}$ as many times as a new datum (not previously seen) appears, and conjectures for every initial segment σ of the input text a canonical index for $\text{range}(\sigma)$ whenever $|\sigma| = 1$, or if $|\sigma| > 1$ and $\text{range}(\sigma') = \text{range}(\sigma)$, where σ' is the prefix of σ with length $|\sigma| - 1$. As a further illustration of how partial learning can enrich the classes of languages that are learnable under the stricter criterion of explanatory learning with respect to various learning constraints, the next theorem states that any consistent, conservative, explanatory learner of a class $\mathcal{S}$ can be turned into a consistent, conservative, as well as reliable learner for $\mathcal{S}$.

Theorem 34 (Gao, Stephan and Zilles [14]). *If a class $\mathcal{S}$ of r.e. languages is consistently conservatively (with both constraints fulfilled at once) explanatorily learnable, then it is also consistently conservatively reliably (with all three constraints fulfilled at once) partially learnable.*

8 Conclusion and Open Problems

The partial learning criterion of converging to a hypothesis is to output this hypothesis infinitely often while every other hypothesis is output only finitely often. This learning criterion is quite general and permits to learn even the class of all A-r.e. languages using A-r.e. indices in the case that A is low. Similarly one can even learn for uncountably many oracles A the class of all A-recursive languages using A-recursive indices. In both results, the learner itself is recursive, so the learning power stems from this general convergence criterion. Therefore it is quite natural to ask what happens if one combines this general learning criterion with some restrictive constraint like confidence, conservativeness, consistency and reliability. While, when partially learning classes of r.e. languages, reliability is not an additional restrictive constraint, the other three notions are and the types of restrictions imposed lead to different, incomparable learning notions. The present survey points to a wide variety of results [9, 10, 12–14, 18] about these combined criteria and also provides various characterisations for those classes which can be partially learnt with the additional constraint of confidence, consistency or conservativeness, respectively. However, two important problems remain open. The first open problem relates to the just mentioned field of learning non-recursive objects using a recursive learner.

Open Problem 35. *Are there uncountably many oracles A such that the class of all A-r.e. languages has a recursive partial learner using A-r.e. indices?*

The second question deals with function learning and has already been open for 5 years [18]. It asks whether an equivalence of consistent function learning carries over to class-consistent function learning.

Open Problem 36. *Consider the model of learning recursive functions from texts in arbitrary order: Is every class-consistently partially learnable class of recursive functions also class-consistently explanatorily learnable?*

The study of partial learnability, whilst intrinsically interesting, is not an end in itself: some of the proof techniques illustrate applications of deep recursion-theoretic notions, especially when considering learnability with respect to oracles; moreover, the various connections between partial learning, Gold's model of learning in the limit, and Case and Lynes' model of behaviourally correct learning, show that the traditional learning models, in which the learner is required to almost always output correct hypotheses, are probably not as restrictive as they may seem at first sight. Perhaps the philosophical import of these relations is no less significant.

References

1. Leonard M. Adleman and Manuel Blum. Inductive inference and unsolvability. *The Journal of Symbolic Logic* 56(3) (1991):891–900.
2. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control* 45(2) (1980):117–135.
3. Ganesh Baliga, John Case and Sanjay Jain. The synthesis of language learners. *Information and Computation* 152 (1999):16–43.
4. Janis Bārzdiņš. Two theorems on the limiting synthesis of functions. In *Theory of Algorithms and Programs, vol. 1*, pages 82–88. Latvian State University, 1974. In Russian.
5. John Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28 (1999):1941–1969.
6. John Case and Timo Kötzing. Difficulties in forcing fairness of polynomial time inductive inference. *Algorithmic Learning Theory*, Twentieth International Conference, ALT 2009, Porto, Portugal, October 3-5, 2009. Proceedings. Springer LNAI 5809 (2009):263–277.
7. John Case and Chris Lynes. Machine inductive inference and language identification. *Proceedings of the Ninth International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science 140 (1982):107–115.
8. John Case and Carl Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science* 25 (1983):193–220.
9. Ziyuan Gao, Sanjay Jain and Frank Stephan. On conservative learning of recursively enumerable languages. *Ninth Conference on Computability in Europe,* CiE 2013, Milan, Italy, July 1–5, 2013. Proceedings. Springer LNCS 7921 (2013):181–190.
10. Ziyuan Gao and Frank Stephan. Confident and consistent partial learning of recursive functions. *Algorithmic Learning Theory,* Twenty-third International Conference, ALT 2012, Lyon, France, October 2012, Proceedings. Springer LNAI 7568 (2012):51–65.
11. Ziyuan Gao and Frank Stephan. Confident and consistent partial learning of recursive functions. Technical Report TRB7/14, School of Computing, National University of Singapore, 2014. Long version of [10].
12. Ziyuan Gao, Frank Stephan, Guohua Wu and Akihiro Yamamoto. Learning families of closed sets in matroids. *Computation, Physics and Beyond; International Workshop on Theoretical Computer Science*, WTCS 2012, Springer LNCS 7160 (2012):120–139.
13. Ziyuan Gao, Frank Stephan and Sandra Zilles. Partial Learning of Recursively Enumerable Languages. *Algorithmic Learning Theory,* Twenty-fourth International Conference, ALT 2013, Singapore, October 2013, Proceedings. Springer LNAI 8139 (2013):101–115.
14. Ziyuan Gao, Frank Stephan and Sandra Zilles. Partial Learning of Recursively Enumerable Languages. Manuscript, long version of [13].

15. E. Mark Gold. Language identification in the limit. *Information and Control* 10 (1967):447–474.
16. Gunter Grieser. Reflective inductive inference of recursive functions. *Theoretical Computer Science A* 397 (2008):57–69.
17. Sanjay Jain, Daniel Osherson, James S. Royer and Arun Sharma. 1999. *Systems that learn: an introduction to learning theory.* MIT Press, Cambridge, Massachusetts.
18. Sanjay Jain and Frank Stephan. Consistent partial identification. *Proceedings of the Twenty-second Annual Conference on Computational Learning Theory,* COLT 2009:135–145.
19. Carl G. Jockusch, Jr and Robert I. Soare. Π^0_1 classes and degrees of theories. *Transactions of the American Mathematical Society* 173 (1972):33–56.
20. Dick de Jongh and Makoto Kanazawa. Angluin's theorem for indexed families of r.e. sets and applications. *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, pages 193–204, ACM Press, 1996.
21. Steffen Lange and Thomas Zeugmann. A guided tour across the boundaries of learning recursive languages. In Klaus P. Jantke and Steffen Lange, eds., *Algorithmic Learning for Knowledge Based Systems.* Lecture Notes in Artificial Intelligence 961 (1995):190–258.
22. Steffen Lange, Thomas Zeugmann and Shyam Kapur. Monotonic and dual monotonic language learning. *Theoretical Computer Science* 155 (1996):365–410.
23. Irene Mazurkewich and Lydia White. The acquisition of dative-alternation: unlearning overgeneralizations. *Cognition* 16 (1984):261–283.
24. Eliana Minicozzi. Some natural properties of strong-identification in inductive inference. *Theoretical Computer Science* 2 (1976):345–360.
25. Daniel N. Osherson, Michael Stob and Scott Weinstein. 1986. *Systems that learn: an introduction to learning theory for cognitive and computer scientists.* Cambridge, Massachusetts.: MIT Press.
26. Leonard Pitt. Inductive inference, DFAs, and computational complexity. *Analogical and Inductive Inference, Proceedings of the Second International Workshop*, AII 1989. Springer LNAI 397 (1989):18–44.
27. Hartley Rogers, Jr. 1987. *Theory of recursive functions and effective computability.* Cambridge, Massachusetts: MIT Press.
28. Rolf Wiehagen and Thomas Zeugmann. Learning and consistency. *Algorithmic Learning for Knowledge-Based Systems*, GOSLER Final Report, Springer LNAI 961 (1995):1–24.
29. Thomas Zeugmann, Steffen Lange and Shyam Kapur. Characterizations of monotonic and dual monotonic language learning. *Information and Computation* 120(2) (1995):155–173.