Vacillatory and BC Learning on Noisy Data

John Case [a], Sanjay Jain [b], Frank Stephan [c]

[a] *Computer and Information Sciences Department*
University of Delaware
Newark, DE 19716, USA
case@cis.udel.edu

[b] *School of Computing*
National University of Singapore
Singapore 119260
Republic of Singapore
sanjay@comp.nus.edu.sg

[c] *Mathematisches Institut*
Im Neuenheimer Feld 294
Ruprecht-Karls-Universität Heidelberg
69120 Heidelberg
fstephan@ira.uka.de

Abstract

The present work employs a model of noise introduced earlier by the third author. In this model noisy data nonetheless uniquely determines the true data: correct information occurs infinitely often while incorrect information occurs only finitely often. The present paper considers the effects of this form of noise on vacillatory and behaviorally correct learning of grammars — both from positive data alone and from informant (positive and negative data). For learning from informant, the noise, in effect, destroys negative data. Various noisy-data hierarchies are exhibited, which, in some cases, are known to collapse when there is no noise. Noisy behaviorally correct learning is shown to obey a very strong "subset principle". It is shown, in many cases, how much power is needed to overcome the effects of noise. For example, the best we can do to simulate, in the presence of noise, the noise-free, no mind change cases takes infinitely many mind changes. One technical result is proved by a priority argument.

1 Introduction

Gold [22] introduced the notion of learning in the limit. In particular he considered a machine, which reads more and more positive information on an r.e.

set and produces in the limit a grammar to generate this set. This is called **Ex** style identification. From then on many variants of this concept have been considered [2,9,17,29].

Barzdin [6] and Case and Smith [18] considered the notion of behaviorally correct inference which is motivated by the fact that no algorithm can check the equivalence of grammars. So, it turns out, the learner can learn more languages if infinitely many guesses are allowed under the condition that almost all of these guessed grammars generate the same correct set. Barzdin and Podnieks [7] introduced the notion of vacillatory inference which is a restriction of behaviorally correct inference in the sense that the learner may change its mind infinitely often, but only between finitely many grammars. They showed that, for learning recursive functions, vacillatory inference is not more powerful than **Ex** style learning in the limit. On the other hand, if one is missing negative information [11–14] or has suitable complexity constraints [15], then vacillatory inference increases learning power.

Many real-world applications of learning or inductive inference have to deal with faulty data, so it is natural to study this phenomenon [5,19,29]. Many of these notions of noise have the disadvantage that noisy data does not specify uniquely the object to be learned. Stephan [33] introduced a notion of noise in order to overcome this difficulty: correct information occurs infinitely often while incorrect information occurs only finitely often.

Many theorems are presented below comparing the learning power for vacillatory and behaviorally correct criteria with or without Stephan's version of noise in the input data.

Stephan [33] showed that the learning power of **Ex** style learning of grammars from noisy positive and negative data (noisy *informant*) is exactly characterized by noise free, one-shot (no mind change) learning (from informant) provided the latter learning machines have access to an oracle for K, the halting problem (see Theorem 12 below). This sort of result provides some insight into the difficulty (as measured by the oracle) of learning with noise. In vacillatory learning, one converges to vacillating between finitely many correct grammars. In $\mathbf{Fex}_b$ style learning, one places a bound of b on the number of different correct grammars one converges to. Theorem 13 implies that, for learning from informant, one can simulate (but not characterize exactly) noisy $\mathbf{Fex}_{b+1}$ style learning with **Ex** style, $\leq b$ mind change learning *provided one has access to the oracle* K. Theorem 13 shows that one can bring the simulation down from *unrestricted* vacillatory learning to *one-shot* **Ex** style learning *using the more complex oracle* K'.

Theorem 14 implies a *very* strong subset principle on noisy behaviorally correct learning from positive information only. It is stronger than that from Angluin's characterization [1] of (uniformly decidable classes) learnable **Ex** style, with no noise, and positive information only. More specifically, Theorem 14 entails that, if $L_1 \subset L_2$, then the class $\{L_1, L_2\}$ can*not* be learned behaviorally correctly from noisy positive data! Even for behaviorally correct learning (from positive data), noise is quite problematic.

It is shown in Theorems 15 and 16 that noise free *two-shot* (one mind change allowed) learning from positive data can*not* be simulated from noisy informant even behaviorally correctly; however, noise free *one-shot* learning from positive data can be simulated (behaviorally correctly) from noisy positive data and from noisy informant!

Theorem 17 while not hard to prove, nicely implies that, in a sense, noise destroys negative information. We indicate how this result may provide the beginnings of a mathematical explanation for some phenomena seen in schizophrenics.

Theorem 27 says that behaviorally correct learning from noisy informant can be simulated by **Ex** style learning from a noise free informant. Hence, for informant data, noise destroys the advantage of behaviorally correct over **Ex** style learning!

If one is missing negative information [11–14] or has suitable complexity constraints [15], then $\mathbf{Fex}_{b+1}$ style learning is more powerful than $\mathbf{Fex}_b$. Theorem 24 implies that, one also gets such a hierarchy result for $\mathbf{Fex}_b$ style learning from noisy informant.

Suppose
a is a natural number or a $*$. Let $\mathrm{Var}^a(L) \stackrel{\mathrm{def}}{=} \{L' \colon L'$ is an a variant of $L\}$, where a $*$ variant is (by definition) a finite variant. In Theorem 32 we show that the classes $(n+1)$-shot **Ex** style learnable from noisy positive data (with final program correct except at up to a arguments) and the classes $(n+1)$-shot **Ex** style learnable from noisy informant (with final program correct except at up to a arguments) are essentially just those of the form $\mathrm{Var}^a(L)$ for some r.e. set L. One can show that $\mathrm{Var}^*(K)$ can be learned **Ex** style from a noise free informant; however, Theorem 35 interestingly *proved by a priority argument*, says that, for any n, for *some* r.e. set L, $\mathrm{Var}^{n+1}(L)$ can*not* be learned **Ex** style from a noise free informant and with final program correct except at up to n arguments.

2 Notations and Identification Criteria

The recursion theoretic notions are from the books of Odifreddi [28] and Soare [32]. $N = \{0, 1, 2, \ldots\}$ is the set of all natural numbers, and this paper considers r.e. subsets L of N. All conventions regarding range of variables apply, with or without decorations (decorations are subscripts, superscripts, primes and the like), unless otherwise specified. We let $c, e, i, j, k, l, m, n, p, s, t, u, v, w, x, y, z,$ range over N. $\emptyset, \in, \subseteq, \supseteq, \subset, \supset$ denote empty set, member of, subset, superset, proper subset, and proper superset respectively. $\max(), \min(), \mathrm{card}()$ denote the maximum, minimum and cardinality of a set respectively, where by convention $\max(\emptyset) = 0$ and $\min(\emptyset) = \infty$. $\mathrm{card}(S) \leq *$ means cardinality of set S is finite. $\langle \cdot, \cdot \rangle$ stands for an arbitrary, one to one, computable encoding of all pairs of natural numbers onto N. $\overline{L}$ denotes the complement of set L. χ_L denotes the characteristic function of set L. $L_1 \mathbf{\Delta} L_2$ denotes the symmetric difference of L_1 and L_2, i.e., $L_1 \mathbf{\Delta} L_2 = (L_1 - L_2) \cup (L_2 - L_1)$. $L_1 =^a L_2$ means that $\mathrm{card}(L_1 \mathbf{\Delta} L_2) \leq a$. Quantifiers $\forall^\infty, \exists^\infty$, and $\exists!$ denote for all but finitely many, there exist infinitely many, and there exists a unique respectively.

$\mathcal{R}$ denotes the set of total recursive functions from N to N. f, g, range over total recursive functions. $\mathcal{E}$ denotes the set of all recursively enumerable sets. L, ranges over $\mathcal{E}$. $\mathcal{L}$, ranges over subsets of $\mathcal{E}$. φ denotes a standard acceptable programming system (acceptable numbering). φ_i denotes the function computed by the i-th program in the programming system φ. We also call i a program or index for φ_i. For a (partial) function η, $\mathrm{domain}(\eta)$ and $\mathrm{range}(\eta)$ respectively denote the domain and range of partial function η. We often write $\eta(x)\downarrow$ ($\eta(x)\uparrow$) to denote that $\eta(x)$ is defined (undefined). W_i denotes the domain of φ_i. W_i is considered as the language enumerated by the i-th program in φ system, and we say that i is a grammar or index for W_i. Φ denotes a standard Blum complexity measure [10] for the programming system φ. $W_{i,s} = \{x < s : \Phi_i(x) < s\}$.

L is called a *single valued total language* iff $(\forall x)(\exists! y)[\langle x, y \rangle \in L]$. $\mathcal{SVT} = \{L : L \text{ is a single valued total language }\}$. If $L \in \mathcal{SVT}$, then we say that L represents the total function f such that $L = \{\langle x, f(x) \rangle : x \in N\}$. K denotes the set $\{x : \varphi_x(x)\downarrow\}$.

A *text* is a mapping from N to $N \cup \{\#\}$. We let T, range over texts. $\mathrm{content}(T)$ is defined to be the set of natural numbers in the range of T (i.e. $\mathrm{content}(T) = \mathrm{range}(T) - \{\#\}$). T is a *text for* L iff $\mathrm{content}(T) = L$. That means a text for L is an infinite sequence whose range, except for a possible #, is just L.

An *information sequence or informant* is a mapping from N to $(N \times N) \cup \{\#\}$. We let I, range over informants. $\mathrm{content}(I)$ is defined to be the set of pairs in the range of I (i.e. content(I) = range(I) − {#}). An *informant for* L is an infinite sequence I such that content$(I) = \{(x,b)\colon \chi_L(x) = b\}$. It is useful to consider canonical information sequence for L. I is a canonical information sequence for L iff $I(x) = (x, \chi_L(x))$. We sometimes abuse notation and refer to the canonical information sequence for L by χ_L.

σ and τ, range over finite initial segments of texts or information sequences, where the context determines which is meant. We denote the set of finite initial segments of texts by SEG and set of finite initial segments of information sequences by SEQ. We use $\sigma \preceq T$ (respectively, $\sigma \preceq I$, $\sigma \preceq \tau$) to denote that σ is an initial segment of T (respectively, I, τ). $|\sigma|$ denotes the length of σ. $T[n]$ denotes the initial segment of T of length n. Similarly, $I[n]$ denotes the initial segment of I of length n. $\sigma \diamond \tau$ (respectively, $\sigma \diamond T$, $\sigma \diamond I$) denotes the concatenation of σ and τ (respectively, concatenation of σ and T, concatenation of σ and I). We sometimes abuse notation and say $\sigma \diamond w$ to denote the concatenation of σ with the sequence of one element w.

A *learning machine* $\mathbf{M}$ is a mapping from initial segments of texts (information sequences) to $(N \cup \{?\})$. The point of using ?'s is to avoid biasing the count of *mind changes* by requiring a learning machine on the empty sequence to output a *program* as its conjecture. For criteria of inference discussed in this paper, we assume, without loss of generality, that $\mathbf{M}(\sigma) \neq {?} \Rightarrow (\forall \tau)[\mathbf{M}(\sigma \diamond \tau) \neq {?}]$.

We say that $\mathbf{M}$ converges on T to i, (written: $\mathbf{M}(T){\downarrow} = i$) iff, for all but finitely many n, $\mathbf{M}(T[n]) = i$. Convergence on information sequences is defined similarly.

Definition 1
(a) Suppose $a, b \in N \cup \{*\}$. Below, for each of several learning criteria $\mathcal{J}$, we define what it means for a machine $\mathbf{M}$ to $\mathcal{J}$*-identify* a language L from a text T or informant I.

- [22,18,9] $\mathbf{M}$ $\mathbf{TxtEx}^a_b$*-identifies* L *from text* T iff $(\exists i\colon W_i =^a L)[\mathbf{M}(T){\downarrow} = i]$ and card$(\{n\colon {?} \neq \mathbf{M}(T[n]) \neq \mathbf{M}(T[n+1])\}) \leq b$.
  We call each instance of ${?} \neq \mathbf{M}(T[n]) \neq \mathbf{M}(T[n+1])$ as a mind change by $\mathbf{M}$ on T.
- [22,18,9] $\mathbf{M}$ $\mathbf{InfEx}^a_b$*-identifies* L *from informant* I iff $(\exists i\colon W_i =^a L)[\mathbf{M}(I){\downarrow} = i]$ and card$(\{n\colon {?} \neq \mathbf{M}(I[n]) \neq \mathbf{M}(I[n+1])\}) \leq b$.
  We call each instance of ${?} \neq \mathbf{M}(I[n]) \neq \mathbf{M}(I[n+1])$ as a mind change by $\mathbf{M}$ on I.
- [6,18]. $\mathbf{M}$ $\mathbf{TxtBc}^a$*-identifies* L *from text* T iff $(\forall^\infty n)[W_{\mathbf{M}(T[n])} =^a L]$.
  $\mathbf{InfBc}^a$-identification is defined similarly.
- [11–13,7]. $\mathbf{M}$ $\mathbf{TxtFex}^a_b$*-identifies* L *from text* T iff $(\exists S\colon \text{card}(S) \leq b \wedge (\forall i \in S)[W_i =^a L])(\forall^\infty n)[\mathbf{M}(T[n]) \in S]$.

$\mathbf{InfFex}_b^a$ is defined similarly.

Based on the definition of $\mathbf{TxtFex}_b^a$ and $\mathbf{InfFex}_b^a$ identification criteria, we sometimes also say that $\mathbf{M}$ on T converges to a set of b grammars iff there exists a set S of cardinality at most b such that $(\forall^\infty n)\,[\mathbf{M}(T[n]) \in S]$. If no such S exists, then we say that $\mathbf{M}$ on T does not converge to a set of b grammars. Similarly we define convergence and divergence on information sequences.

$\mathrm{Last}_b(\mathbf{M},\sigma)$ denotes the set of last b grammars output by $\mathbf{M}$ on σ. Formally, $\mathrm{Last}_b(\mathbf{M},\sigma)$ is defined as follows. Let τ be the smallest initial segment of σ such that $\mathrm{card}(\{\mathbf{M}(\tau') \colon \tau \preceq \tau' \preceq \sigma\} - \{?\}) \leq b$. Then $\mathrm{Last}_b(\mathbf{M},\sigma) = \{\mathbf{M}(\tau') \colon \tau \preceq \tau' \preceq \sigma\} - \{?\}$. Note that $\mathrm{Last}_*(\sigma)$ is just the set of all grammars $\mathbf{M}$ outputs while reading initial segments of σ.

If $\lim_{n\to\infty} \mathrm{Last}_b(\mathbf{M},T[n])\downarrow$, then we say that $\mathrm{Last}_b(\mathbf{M},T) = \lim_{n\to\infty} \mathrm{Last}_b(\mathbf{M},T[n])$. Otherwise $\mathrm{Last}_b(\mathbf{M},T)$ is undefined. $\mathrm{Last}_b(\mathbf{M},I)$ is defined similarly.

- [18]. $\mathbf{M}$ $\mathbf{TxtOex}_b^a$*-identifies* L *from text* T iff $\mathrm{Last}_b(\mathbf{M},T)$ is defined and $(\exists i \in \mathrm{Last}_b(\mathbf{M},T))[W_i =^a L]$.

  $\mathbf{InfOex}_b^a$ is defined similarly.

(b) Suppose $\mathcal{J} \in \{\mathbf{TxtEx}_b^a, \mathbf{TxtFex}_b^a, \mathbf{TxtOex}_b^a, \mathbf{TxtBc}_b^a\}$.

$\mathbf{M}$ $\mathcal{J}$*-identifies* L iff, for all texts T for L, $\mathbf{M}$ $\mathcal{J}$-identifies L from T. In this case we also write $L \in \mathcal{J}(\mathbf{M})$.

We say that $\mathbf{M}$ $\mathcal{J}$-identifies $\mathcal{L}$ iff $\mathbf{M}$ $\mathcal{J}$-identifies each $L \in \mathcal{L}$.

$\mathcal{J} = \{\mathcal{L} \colon (\exists \mathbf{M})[\mathcal{L} \subseteq \mathcal{J}(\mathbf{M})]\}$.

(c) Suppose $\mathcal{J} \in \{\mathbf{InfEx}_b^a, \mathbf{InfFex}_b^a, \mathbf{InfOex}_b^a, \mathbf{InfBc}_b^a\}$.

$\mathbf{M}$ $\mathcal{J}$*-identifies* L iff, for all information sequences I for L, $\mathbf{M}$ $\mathcal{J}$-identifies L from I. In this case we also write $L \in \mathcal{J}(\mathbf{M})$.

We say that $\mathbf{M}$ $\mathcal{J}$-identifies $\mathcal{L}$ iff $\mathbf{M}$ $\mathcal{J}$-identifies each $L \in \mathcal{L}$.

$\mathcal{J} = \{\mathcal{L} \colon (\exists \mathbf{M})[\mathcal{L} \subseteq \mathcal{J}(\mathbf{M})]\}$.

We often write $\mathbf{TxtEx}_b^0$ as $\mathbf{TxtEx}_b$, $\mathbf{TxtEx}_*^a$ as $\mathbf{TxtEx}^a$, and $\mathbf{TxtEx}_*^0$ as $\mathbf{TxtEx}$. Similar convention applies to **TxtFex**, **TxtOex**, **TxtBc**, **InfEx**, **InfFex**, **InfOex**, **InfBc** criteria. Also, for criteria of inference which do not count mind changes (that is all criteria of inference discussed in this paper except for $\mathbf{TxtEx}_b^a$, $\mathbf{InfEx}_b^a$, for $b \in N$, and corresponding criteria involving noise discussed below), we assume, without loss of generality, that machine never outputs ?.

For the sake of measuring the difficulty of some learning situations, we sometimes consider learning machines with access to (possibly non-computable oracle). Suppose $\mathcal{I}$ is an identification criterion considered in this paper. Then $\mathcal{I}[A]$ denotes the identification criteria formed from $\mathcal{I}$ by allowing the learning machines access to oracle A. Gasarch and Pleszkoch [21], building on earlier work of L. Adleman and M. Blum, were first in print to consider the notion of learning with oracle.

Several proofs in this paper depend on the concept of locking sequence.

Definition 2 (a) σ is said to be a $\mathbf{TxtEx}^a$*-locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{content}(\sigma) \subseteq L$, $W_{\mathbf{M}(\sigma)} =^a L$, and $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) = \mathbf{M}(\sigma)]$.

(b) σ is said to be a $\mathbf{TxtFex}^a_b$*-locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{content}(\sigma) \subseteq L$, and there exists a set S such that

(b.1) $\mathrm{card}(S) \leq b$,
(b.2) $S \subseteq \mathrm{Last}_b(\mathbf{M}, \sigma)$,
(b.3) $(\forall i \in S)[W_i =^a L]$, and
(b.4) $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) \in S]$.

(c) σ is said to be a $\mathbf{TxtOex}^a_b$*-locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{content}(\sigma) \subseteq L$, and there exists a set S such that

(c.1) $\mathrm{card}(S) \leq b$,
(c.2) $S \subseteq \mathrm{Last}_b(\mathbf{M}, \sigma)$,
(c.3) $(\exists i \in S)[W_i =^a L]$, and
(c.4) $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) \in S]$.

(d) σ is said to be a $\mathbf{TxtBc}^a$*-locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{content}(\sigma) \subseteq L$, and $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) =^a L]$.

Lemma 3
(Based on [9]) Suppose $\mathcal{J} \in \{\mathbf{TxtEx}^a, \mathbf{TxtFex}^a_b, \mathbf{TxtOex}^a_b, \mathbf{TxtBc}^a\}$*. If* $\mathbf{M}$ $\mathcal{J}$*-identifies* L *then there exists a* $\mathcal{J}$*-locking sequence for* $\mathbf{M}$ *on* L*.*

Next we prepare to introduce our noisy inference criteria, and, in that interest, we define some ways to calculate the number of occurrences of words in (initial segments of) a text or informant.

For $\sigma \in \mathrm{SEG}$, and text T, let

$$\mathrm{occur}(\sigma, w) \stackrel{\mathrm{def}}{=} \mathrm{card}(\{j : j < |\sigma| \ \wedge \ \sigma(j) = w\}).$$

and

$$\mathrm{occur}(T, w) \stackrel{\mathrm{def}}{=} \mathrm{card}(\{j : j \in N \ \wedge \ T(j) = w\}).$$

For $\sigma \in \mathrm{SEQ}$ and information sequence I, $\mathrm{occur}(\cdot, \cdot)$ is defined similarly except that w is replaced by (v, b).

For any language L,

$$\mathrm{occur}(T, L) \stackrel{\mathrm{def}}{=} \Sigma_{x \in L}\, \mathrm{occur}(T, x).$$

It is useful to introduce the set of positive and negative occurrences in (initial segment of) an informant. Suppose $\sigma \in \mathrm{SEQ}$

$$\mathrm{Pos}(\sigma) \stackrel{\mathrm{def}}{=} \{v \colon \mathrm{occur}(\sigma, (v, 1)) \geq \mathrm{occur}(\sigma, (v, 0)) \wedge \mathrm{occur}(\sigma, (v, 1)) \geq 1\}$$

$$\mathrm{Neg}(\sigma) \stackrel{\mathrm{def}}{=} \{v \colon \mathrm{occur}(\sigma, (v, 1)) < \mathrm{occur}(\sigma, (v, 0)) \wedge \mathrm{occur}(\sigma, (v, 0)) \geq 1\}$$

That means, that $\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma)$ is just the set of all v such that either $(v, 0)$ or $(v, 1)$ occurs in σ. Then $v \in \mathrm{Pos}(\sigma)$ if $(v, 1)$ occurs at least as often as $(v, 0)$ and $v \in \mathrm{Neg}(\sigma)$ otherwise.

Similarly,

$$\mathrm{Pos}(I) = \{v \colon \mathrm{occur}(I, (v, 1)) \geq \mathrm{occur}(I, (v, 0)) \wedge \mathrm{occur}(I, (v, 1)) \geq 1\}$$

$$\mathrm{Neg}(I) = \{v \colon \mathrm{occur}(I, (v, 1)) < \mathrm{occur}(I, (v, 0)) \wedge \mathrm{occur}(I, (v, 0)) \geq 1\}$$

where, if $\mathrm{occur}(I, (v, 0)) = \mathrm{occur}(I, (v, 1)) = \infty$, then we place v in $\mathrm{Pos}(I)$ (this is just to make definition precise; we will not need this for criteria of inference discussed in this paper).

Definition 4 [33] An information sequence I is a *noisy information sequence* (or noisy informant) for L iff $(\forall x)\, [\mathrm{occur}(I, (x, \chi_L(x))) = \infty \wedge \mathrm{occur}(I, (x, \chi_{\overline{L}}(x))) < \infty]$. A text T is a *noisy text* for L iff $(\forall x \in L)[\mathrm{occur}(T, x) = \infty]$ and $\mathrm{occur}(T, \overline{L}) < \infty$.

On one hand, both concepts are similar since $L = \{x \colon \mathrm{occur}(I, (x, 1)) = \infty\} = \{x \colon \mathrm{occur}(T, x) = \infty\}$. On the other hand, the concepts differ in the way they treat errors. In the case of informant every false item $(x, \chi_{\overline{L}}(x))$ may occur a finite number of times. In the case of text, it is mathematically more interesting to require, as we do, that the *total* amount of false information has to be finite. The alternative of allowing each false item in a text to occur finitely often is too restrictive. It would, then, be impossible to learn even the class of all singleton sets.

Definition 5 *Suppose* $a, b \in N \cup \{*\}$.
Suppose $\mathcal{J} \in \{\mathbf{TxtEx}^a_b, \mathbf{TxtFex}^a_b, \mathbf{TxtOex}^a_b, \mathbf{TxtBc}^a\}$. *Then* $\mathbf{M}$ $\mathbf{Noisy}\mathcal{J}$*-identifies* L *iff, for all noisy texts* T *for* L, $\mathbf{M}$ $\mathcal{J}$*-identifies* L *from* T. *In this case we write* $L \in \mathbf{Noisy}\mathcal{J}(\mathbf{M})$.

$\mathbf{M}$ $\mathbf{Noisy}\mathcal{J}$*-identifies a class* $\mathcal{L}$ *iff* $\mathbf{M}$ $\mathbf{Noisy}\mathcal{J}$*-identifies each* $L \in \mathcal{L}$.

$\mathbf{Noisy}\mathcal{J} = \{\mathcal{L} \colon (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{Noisy}\mathcal{J}(\mathbf{M})]\}$.

Inference criteria for learning from noisy informants are defined similarly.

Several proofs use the existence of locking sequences. Definition of locking sequences for learning from noisy texts is similar to that of learning from noise free texts (we just drop the requirement that $\mathrm{content}(\sigma) \subseteq L$). However, definition of locking sequence for learning from noisy informant is more involved.

Definition 6 (a) σ is said to be a $\mathbf{NoisyTxtEx}^a$-*locking sequence for* $\mathbf{M}$ *on* L iff, $W_{\mathbf{M}(\sigma)} =^a L$, and $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) = \mathbf{M}(\sigma)]$.

(b) σ is said to be a $\mathbf{NoisyTxtFex}_b^a$-*locking sequence for* $\mathbf{M}$ *on* L iff there exists a set S such that

(b.1) $\mathrm{card}(S) \leq b$,
(b.2) $S \subseteq \mathrm{Last}_b(\mathbf{M}, \sigma)$,
(b.3) $(\forall i \in S)[W_i =^a L]$, and
(b.4) $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) \in S]$.

(c) σ is said to be a $\mathbf{NoisyTxtOex}_b^a$-*locking sequence for* $\mathbf{M}$ *on* L iff there exists a set S such that

(c.1) $\mathrm{card}(S) \leq b$,
(c.2) $S \subseteq \mathrm{Last}_b(\mathbf{M}, \sigma)$,
(c.3) $(\exists i \in S)[W_i =^a L]$, and
(c.4) $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) \in S]$.

(d) σ is said to be a $\mathbf{NoisyTxtBc}^a$-*locking sequence for* $\mathbf{M}$ *on* L iff $(\forall \tau : \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma \diamond \tau) =^a L]$.

For defining locking sequences for learning from noisy informant, we need the following.

Definition 7 $\mathrm{Inf}[S, L] \stackrel{\mathrm{def}}{=} \{\tau : (\forall x \in S)\,[\mathrm{occur}(\tau, (x, \chi_{\overline{L}}(x))) = 0]\}$.

Definition 8 (a) σ is said to be a $\mathbf{NoisyInfEx}^a$-*locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{Pos}(\sigma) \subseteq L$, $\mathrm{Neg}(\sigma) \subseteq \overline{L}$, $W_{\mathbf{M}(\sigma)} =^a L$, and $(\forall \tau \in \mathrm{Inf}[\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma), L])[\mathbf{M}(\sigma \diamond \tau) = \mathbf{M}(\sigma)]$.

(b) σ is said to be a $\mathbf{NoisyInfFex}_b^a$-*locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{Pos}(\sigma) \subseteq L$, $\mathrm{Neg}(\sigma) \subseteq \overline{L}$, and there exists a set S such that,

(b.1) $\mathrm{card}(S) \leq b$,
(b.2) $S \subseteq \mathrm{Last}_b(\mathbf{M}, \sigma)$,
(b.3) $(\forall i \in S)[W_i =^a L]$, and
(b.4) $(\forall \tau \in \mathrm{Inf}[\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma), L])[\mathbf{M}(\sigma \diamond \tau) \in S]$.

(c) σ is said to be a $\mathbf{NoisyInfOex}_b^a$-*locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{Pos}(\sigma) \subseteq$

L, $\mathrm{Neg}(\sigma) \subseteq \overline{L}$, and there exists a set S such that,

(c.1) $\mathrm{card}(S) \leq b$,
(c.2) $S \subseteq \mathrm{Last}_b(\mathbf{M}, \sigma)$,
(c.3) $(\exists i \in S)[W_i =^a L]$, and
(c.4) $(\forall \tau \in \mathrm{Inf}[\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma), L])[\mathbf{M}(\sigma \diamond \tau) \in S]$.

(d) σ is said to be a $\mathbf{NoisyInfBc}^a$*-locking sequence for* $\mathbf{M}$ *on* L iff, $\mathrm{Pos}(\sigma) \subseteq L$, $\mathrm{Neg}(\sigma) \subseteq \overline{L}$, and $(\forall \tau \in \mathrm{Inf}[\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma), L])[W_{\mathbf{M}(\sigma \diamond \tau)} =^a L]$.

For the criteria of noisy inference discussed in this paper one can prove the existence of a locking sequence as was done in [33, Theorem 2, proof for $\mathbf{NoisyEx} \subseteq \mathbf{Ex}_0[K]$].

Proposition 9 *If* $\mathbf{M}$ *learns* L *from noisy text or informant according to one of the criteria* $\mathbf{NoisyTxtEx}^a$, $\mathbf{NoisyTxtFex}^a$, $\mathbf{NoisyTxtOex}^a_b$ *and* $\mathbf{NoisyTxtBc}^a$, $\mathbf{NoisyInfEx}^a$, $\mathbf{NoisyInfFex}^a$, $\mathbf{NoisyInfOex}^a_b$ *and* $\mathbf{NoisyInfBc}^a$, *then there exists a corresponding locking sequence for* $\mathbf{M}$ *on* L.

The following theorem gives some of the results from the literature when there is no noise in the input data.

Theorem 10 *Let* $a \in N \cup \{*\}$ *and* $n \in N$.
(a) [18] $\mathbf{TxtEx}_0^{n+1} - \mathbf{InfFex}_*^n \neq \emptyset$. $\mathbf{TxtEx}_0^* - \bigcup_{n \in N} \mathbf{InfFex}_*^n \neq \emptyset$.
(b) [18] $\mathbf{TxtEx}_{n+1} - \mathbf{InfEx}_n^* \neq \emptyset$. $\mathbf{TxtEx} - \bigcup_{n \in N} \mathbf{InfEx}_n^* \neq \emptyset$.
(c) [7,18] $\mathbf{InfFex}_*^a = \mathbf{InfEx}^a$.
(d) [11–13] $\mathbf{TxtFex}_{n+1} - \mathbf{TxtFex}_n^* \neq \emptyset$. $\mathbf{TxtFex}_* - \bigcup_{n \in N} \mathbf{TxtFex}_n^* \neq \emptyset$.
(e) [18] $\mathbf{InfFex}_*^* \subset \mathbf{InfBc}$.
(f) [18] $\mathbf{TxtBc}^{n+1} - \mathbf{InfBc}^n \neq \emptyset$. $\mathbf{TxtBc}^* - \bigcup_{n \in N} \mathbf{InfBc}^n \neq \emptyset$.
(g) [17,12,13] $\mathbf{TxtFex}_*^{2n} \subseteq \mathbf{TxtBc}^n$.
(h) [17,12,13] $\mathbf{TxtEx}_0^{2n+1} - \mathbf{TxtBc}^n \neq \emptyset$. $\mathbf{TxtEx}_0^* - \bigcup_{n \in N} \mathbf{TxtBc}^n \neq \emptyset$.
(i) $\mathbf{InfEx}_1 - \mathbf{TxtBc}^* \neq \emptyset$.
(j) [31] $\mathbf{InfEx}_0^a \subseteq \mathbf{TxtEx}^a$.
(k) [18] $\mathbf{InfOex}_*^n = \mathbf{InfEx}_*^n$.
(l) [18,17,22] $\mathbf{TxtOex}_2^* - \bigcup_{n \in N} \mathbf{InfBc}^n \neq \emptyset$. $\mathbf{TxtOex}_2^* - \mathbf{TxtBc}^* \neq \emptyset$.

Moreover parts (a) and (b) can be shown using subsets of $\mathcal{SVT}$ as a diagonalizing class. We do not know if part (i) has been explicitly proved in any paper, but it can be proven using the class defined as follows: Let $L_n = \{\langle i, x \rangle : x \neq n\}$. Let $\mathcal{L} = \{N\} \cup \{L_n : n \in N\}$. It is easy to verify that $\mathcal{L} \in \mathbf{InfEx}_1$. However, using a locking sequence argument, one can show that $\mathcal{L} \notin \mathbf{TxtBc}^*$.

Theorem 11 *Suppose* $n \in N$ *and* $b \in N \cup \{*\}$. $\mathbf{TxtFex}_b^n = \mathbf{TxtOex}_b^n$.

PROOF. Fix $n \in N$ and $b \in N \cup \{*\}$. $\mathbf{TxtFex}_b^n \subseteq \mathbf{TxtOex}_b^n$ by definition. We now show that $\mathbf{TxtOex}_b^n \subseteq \mathbf{TxtFex}_b^n$. Suppose $\mathcal{L} \subseteq \mathbf{TxtOex}_b^n(\mathbf{M})$. We give a machine $\mathbf{M}'$ such that $\mathcal{L} \subseteq \mathbf{TxtFex}_b^n(\mathbf{M}')$. Let

$Q(\sigma, e) = \max(\{m : m \leq |\sigma| \wedge \text{content}(\sigma) \cap \{0, 1, \ldots, m\} =^n W_{e,|\sigma|} \cap \{0, 1, \ldots, m\}\})$.

For each σ, $\mathbf{M}'(\sigma)$ outputs $e \in \text{Last}_b(\mathbf{M}, \sigma)$ such that $Q(\sigma, e)$ is maximized.

Suppose T is a text for $L \in \mathbf{TxtOex}_b^n(\mathbf{M})$.

If, $W_e \neq^n L$, then there exists a c such that $W_e \cap \{0, 1, \ldots, c\} \neq^n L \cap \{0, 1, \ldots, c\}$. Thus, for all but finitely many $\sigma \preceq T$, $Q(\sigma, e) \leq c$.

On the other hand, if $W_e =^n L$, then, for all c, for all but finitely many $\sigma \preceq T$, $Q(\sigma, e) \geq c$.

Therefore, if $W_e =^n L$ and $W_{e'} \neq^n L$, then, for all but finitely many $\sigma \preceq T$, $Q(\sigma, e) > Q(\sigma, e')$.

Thus, for all but finitely many $\sigma \preceq T$, $\mathbf{M}'(\sigma) \in \text{Last}_b(\mathbf{M}, T)$, and $W_{\mathbf{M}'(\sigma)} =^n L$. Thus $\mathbf{M}'$ $\mathbf{TxtFex}_b^n$-identifies L. □

3 Simulating Identification from Noisy Data Using Oracles

Stephan [33] showed that $\mathbf{NoisyInfEx} = \mathbf{InfEx}_0[K]$. His proof also shows,

Theorem 12 *Suppose $a \in N \cup \{*\}$. $\mathbf{NoisyInfEx}^a = \mathbf{InfEx}_0^a[K]$.*

One direction of Theorem 12 can be generalized: learning from noisy informant can be simulated by one-shot (finite) learning with suitable oracle. However, the criterion $\mathbf{NoisyInfFex}_{n+1}$ is too strong to get an exact characterization. Nonetheless, we get some insight into the cost of noise from the following theorem.

Theorem 13 *Suppose $m, n \in N$.*
(a) $\mathbf{NoisyInfFex}_{n+1}^m \subseteq \mathbf{InfEx}_n^m[K]$.
(b) $\mathbf{NoisyInfFex}_* \subseteq \mathbf{InfEx}_0[K']$.

PROOF. (a) Note that if $\mathbf{M}$ $\mathbf{NoisyInfFex}_{n+1}^m$-identifies L, then there exists a $\mathbf{NoisyInfFex}_{n+1}^m$ locking sequence for $\mathbf{M}$ on L. This is what our simulation below utilizes.

It is easy to construct $\mathbf{F}^K$, an algorithmic mapping (with oracle K) from finite information sequences to finite sets, such that the following is satisfied.

Suppose I is an information sequence for $L \in \mathbf{NoisyInfFex}^m_{n+1}(\mathbf{M})$. Then there exists a $\mathbf{NoisyInfFex}^m_{n+1}$-locking sequence σ for $\mathbf{M}$ on L and $t \in N$ such that

$$(\forall t' < t)[\mathbf{F}^K(I[t']) = \emptyset] \;\wedge\; (\forall t' \geq t)[\mathbf{F}^K(I[t']) = \mathrm{Last}_{n+1}(\mathbf{M}, \sigma)].$$

Essentially the trick used by Stephan to prove Theorem 12 can be used to construct such an $\mathbf{F}^K$.

Suppose I is an information sequence for L. Let $\mathbf{M}_1^K(I[n])$ be defined as follows.

$$\mathbf{M}_1^K(I[n]) = \begin{cases} e, & \text{if } \mathbf{F}^K(I[n]) = S \neq \emptyset, \text{ and} \\ & e = \min(\{e' \in S \colon \\ & \mathrm{card}(\{x \colon (x, 1 - \chi_{W_{e'}}(x)) \in \mathrm{content}(I[n])\}) \leq m\}); \\ ?, & \text{otherwise.} \end{cases}$$

Suppose I is an information sequence for $L \in \mathbf{NoisyTxtFex}^m_{n+1}(\mathbf{M})$. Let σ be a $\mathbf{NoisyInfFex}^m_{n+1}$-locking sequence for $\mathbf{M}$ on L such that, for $S = \mathrm{Last}_{n+1}(\mathbf{M}, \sigma)$, $(\exists t)[(\forall t' < t)[\mathbf{F}^K(I[n]) = \emptyset] \;\wedge\; (\forall t' \geq t)[\mathbf{F}^K(I[t']) = S]]$. Then it is easy to verify that the conjectures of $\mathbf{M}_1^K$ on I are from S and monotonically increasing. Moreover, $\mathbf{M}_1^K(I)$ converges to the least grammar e in S such that $W_e =^m L$. Thus $L \in \mathbf{InfEx}^m_n(\mathbf{M}_1^K)$. It follows that $\mathbf{NoisyInfFex}^m_{n+1} \subseteq \mathbf{InfEx}^m_n[K]$.

(b) As in the proof for part (a) one can construct a machine $\mathbf{F}^K$ with the following property.

Suppose I is an information sequence for $L \in \mathbf{NoisyTxtFex}_*(\mathbf{M})$. Then there exists a $\mathbf{NoisyInfFex}_*$-locking sequence σ for $\mathbf{M}$ on L and $t \in N$ such that

$$(\forall t' < t)[\mathbf{F}^K(I[t']) = \emptyset] \;\wedge\; (\forall t' \geq t)[\mathbf{F}^K(I[t']) = \mathrm{Last}_*(\mathbf{M}, \sigma)].$$

Let $\mathbf{M}_1^{K'}(I[n])$ be defined as follows.

$$\mathbf{M}_1^{K'}(I[n]) = \begin{cases} e, & \text{if } \mathbf{F}^K(I[n]) = S \neq \emptyset, \\ & \text{and there is a nonempty } S' \subseteq S \text{ such that} \\ & e = \min(S'),\ (\forall e' \in S')[W_e = W_{e'}] \text{ and} \\ & (\forall e' \in S - S')(\exists (x, d) \in \mathrm{content}(I[n])) \\ & \quad [\chi_{W_{e'}}(x) \neq d]; \\ ?, & \text{otherwise.} \end{cases}$$

Suppose I is an information sequence for $L \in \mathbf{NoisyTxtFex}_*(\mathbf{M})$. Let σ be a $\mathbf{NoisyInfFex}_*$-locking sequence for $\mathbf{M}$ on L such that, for $S = \mathrm{Last}_*(\mathbf{M}, \sigma)$, $(\exists t)[(\forall t' < t)[\mathbf{F}^K(I[n]) = \emptyset] \ \wedge \ (\forall t' \geq t)[\mathbf{F}^K(I[t']) = S]]$. Then it is easy to verify that, $\mathbf{M}_1^{K'}$ on I outputs $\min(\{e \in S : W_e = L\})$, as its only grammar.

It follows that $\mathbf{NoisyInfFex}_*(\mathbf{M}) \subseteq \mathbf{InfEx}_0(\mathbf{M}_1^{K'})$. Hence $\mathbf{NoisyInfFex}_* \subseteq \mathbf{InfEx}_0[K']$. □

4 Disadvantages of Having Noise in the Input

We now prove some results that, in some cases, show that noise in the input is quite restrictive.

The following theorem (Theorem 14) provides a *very* strong subset principle on $\mathbf{NoisyTxtBc}^a$, stronger than that from Angluin's characterization [1] of (uniformly decidable classes in) **TxtEx**. (This latter subset principle, for preventing overgeneralization, is further discussed, for example, in [8,12,13,24,34,3,35,23]. Mukouchi [27] and Lange and Zeugmann [25] present a subset principle for one-shot learning.) Even at the **TxtBc** levels, noise is problematic. A similar theorem for **NoisyTxtEx** was proven by [33].

Theorem 14 *Suppose* $L_1 \subseteq L_2$.
(a) If $L_1 \neq^{2n} L_2$ *then* $\{L_1, L_2\} \notin \mathbf{NoisyTxtBc}^n$.
(b) If $L_1 \neq^* L_2$ *then* $\{L_1, L_2\} \notin \mathbf{NoisyTxtBc}^*$.

Proof. Suppose that $\mathbf{M}$ $\mathbf{NoisyTxtBc}^a$-identifies $\{L_1, L_2\}$. Then there exists a $\mathbf{NoisyTxtBc}^a$-locking sequence σ for $\mathbf{M}$ on L_2. Thus,

$$(\forall \tau : \mathrm{content}(\tau) \subseteq L_2)[W_{\mathbf{M}(\sigma \diamond \tau)} =^a L_2].$$

On the other hand, since $\mathbf{M}$ $\mathbf{NoisyTxtBc}^a$-identifies L_1,

$$(\exists \tau : \mathrm{content}(\tau) \subseteq L_1 \subseteq L_2)[W_{\mathbf{M}(\sigma \diamond \tau)} =^a L_1].$$

For such τ, $L_1 =^a W_{\mathbf{M}(\sigma \diamond \tau)} =^a L_2$. If $a \in N$, it follows that $L_1 =^{2a} L_2$; if $a = *$, it follows that $L_1 =^* L_2$. □

The following theorem shows the disadvantages of noisy text.

Theorem 15 *Suppose* $a \in N \cup \{*\}$ *and* $n \in N$.
(a) $\mathbf{TxtEx}_1 - \mathbf{NoisyTxtBc}^* \neq \emptyset$.
(b) $\mathbf{InfEx}_0 - \mathbf{NoisyTxtBc}^* \neq \emptyset$.
(c) $\mathbf{TxtEx}_0^{n+1} - \mathbf{NoisyTxtBc}^n \neq \emptyset$. $\mathbf{TxtEx}_0^* - \bigcup_{n \in N} \mathbf{NoisyTxtBc}^n \neq \emptyset$.
(d) $\mathbf{TxtEx}_0^a \subseteq \mathbf{NoisyTxtBc}^a$.
(e) $\mathbf{TxtEx}_0 - \mathbf{NoisyTxtOex}_*^* \neq \emptyset$.

PROOF. (a), (b) Let L_1 and L_2 be two r.e. languages such that $L_1 \subseteq L_2$ and $L_1 \neq^* L_2$. Clearly, $\{L_1, L_2\} \in \mathbf{TxtEx}_1$ and $\{L_1, L_2\} \in \mathbf{InfEx}_0$. However, $\{L_1, L_2\} \notin \mathbf{NoisyTxtBc}^*$ by Theorem 14.

(c) Let $L_1 \subseteq L \subseteq L_2$ be three r.e. languages such that $\text{card}(L_2 - L) = \text{card}(L - L_1) = n + 1$. Clearly, $\{L_1, L_2\} \in \mathbf{TxtEx}_0^{n+1}$ via guessing a grammar for L independently of the input. However $\{L_1, L_2\} \notin \mathbf{NoisyTxtBc}^n$ by Theorem 14.

Let $\mathcal{L} = \{L \colon \text{card}(L) < \infty\}$. Clearly, $\mathcal{L} \in \mathbf{TxtEx}_0^*$. However, for all n, $\mathcal{L} \notin \mathbf{NoisyTxtBc}^n$ by Theorem 14.

(d) The proof of $\mathbf{TxtEx}_0^a \subseteq \mathbf{NoisyTxtBc}^a$ is identical to that of Theorem 23 in [33].

(e) Let $L_K = \{\langle x, y\rangle : x \in K, y \in N\}$ and $L_x = \{\langle x, y\rangle : y \in N\}$. Let $\mathcal{L} = \{L_K\} \cup \{L_x \colon x \notin K\}$. It is easy to verify that $\mathcal{L} \in \mathbf{TxtEx}_0$.

Suppose by way of contradiction that $\mathcal{L} \subseteq \mathbf{NoisyTxtOex}_*^*(\mathbf{M})$. Then there exists a $\mathbf{NoisyTxtOex}_*^*$-locking sequence σ for $\mathbf{M}$ on L_K. Thus, for all τ such that $\text{content}(\tau) \subseteq L_K$, $\mathbf{M}(\sigma \diamond \tau) \in \text{Last}_*(\mathbf{M}, \sigma)$. Intuitively, after reading σ, $\mathbf{M}$ does not make any *new* guess on inputs from L_K. In particular,

$$(\forall x \in K)(\forall \tau \colon \text{content}(\tau) \subseteq L_x)[\mathbf{M}(\sigma \diamond \tau) \in \text{Last}_*(\mathbf{M}, \sigma)].$$

On the other hand, for all but finitely many $x \notin K$, $\text{Last}_*(\mathbf{M}, \sigma)$ does not contain a grammar for a finite variant of L_x. Thus, since $L_x \in \mathbf{NoisyTxtOex}_*^*(\mathbf{M})$,

$$(\forall^\infty x \notin K)(\exists \tau \colon \text{content}(\tau) \subseteq L_x)[\mathbf{M}(\sigma \diamond \tau) \notin \text{Last}_*(\mathbf{M}, \sigma)].$$

It follows that

$$(\forall^\infty x)[x \notin K \Leftrightarrow (\exists \tau \colon \text{content}(\tau) \subseteq L_x)\,[\mathbf{M}(\sigma \diamond \tau) \notin \text{Last}_*(\mathbf{M}, \sigma)]].$$

But then K is co-r.e., a contradiction. Thus no such machine $\mathbf{M}$ can exist. □

The following theorem shows the disadvantages of noisy informant.

Theorem 16 *Suppose $a \in N \cup \{*\}$ and $n \in N$.*
(a) $\mathbf{TxtEx}_1 - (\mathbf{NoisyInfBc}^* \cup \mathbf{NoisyInfOex}_*^*) \neq \emptyset$.
(b) $\mathbf{InfEx}_0^a \subseteq \mathbf{NoisyInfEx}^a$.
(c) $\mathbf{InfEx}_0^{2n} \subseteq \mathbf{NoisyInfBc}^n$.

(d) $\mathbf{TxtEx}_0^{n+1} - \mathbf{NoisyInfOex}_*^n \neq \emptyset$. $\mathbf{TxtEx}_0^* - \bigcup_{n \in N} \mathbf{NoisyInfOex}_*^n \neq \emptyset$.
(e) $\mathbf{TxtEx}_0^{2n+1} - \mathbf{NoisyInfBc}^n \neq \emptyset$. $\mathbf{TxtEx}_0^* - \bigcup_{n \in N} \mathbf{NoisyInfBc}^n \neq \emptyset$.

PROOF. (a) Let $L_x = \{\langle x, y\rangle : y \in N\}$. Consider $\mathcal{L} = \{L_x : x \in N\} \cup \{\emptyset\}$. Clearly $\mathcal{L} \in \mathbf{TxtEx}_1$.

We first show that $\mathcal{L} \notin \mathbf{NoisyInfBc}^*$. Suppose by way of contradiction $\mathcal{L} \subseteq \mathbf{NoisyInfBc}^*(\mathbf{M})$. Then there exists a $\mathbf{NoisyInfBc}^*$-locking sequence σ for $\mathbf{M}$ on $\emptyset$. But then $\mathbf{M}$ does not $\mathbf{NoisyInfBc}^*$-identify any L_x such that $L_x \cap (\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma)) = \emptyset$. It follows that $\mathcal{L} \not\subseteq \mathbf{NoisyInfBc}^*(\mathbf{M})$.

We now show that $\mathcal{L} \notin \mathbf{NoisyInfOex}_*^*$. Suppose by way of contradiction $\mathcal{L} \subseteq \mathbf{NoisyInfOex}_*^*(\mathbf{M})$. Then there exists a $\mathbf{NoisyInfOex}_*^*$-locking sequence σ for $\mathbf{M}$ on $\emptyset$. But then $\mathbf{M}$ does not $\mathbf{NoisyInfOex}_*^*$-identify all but finitely many L_x such that $L_x \cap (\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma)) = \emptyset$. It follows that $\mathcal{L} \not\subseteq \mathbf{NoisyTxtOex}_*^*(\mathbf{M})$.

(b) Follows from Theorem 12.

(c) Follows from part (b) and Theorem 23.

(d) Follows from Theorem 10.

(e) Follows from Theorem 10 and $\mathbf{NoisyInfBc}^n \subseteq \mathbf{TxtBc}^n$ (see Theorem 17). □

From the above theorems, we see that $\mathbf{TxtEx}_1$-inference cannot be simulated from noisy data (even for $\mathbf{Bc}^*$-identification criteria). This contrasts nicely with the fact that finite learning can be simulated by behaviorally correct learning from noisy data.

5 Advantages of Weaker Inference Criterion Despite the Presence of Noise

The next theorem (Theorem 17), while not hard to prove, is quite interesting since, in part, it means that noise destroys negative information.

We told the mathematician and psychiatrist Tom Nordahl about Theorem 17 after he had contacted us inquiring about [19]. He was interested in the possible relevance to schizophrenia. Tom told us that schizophrenics, compared to normals and in contexts requiring some conscious processing, have trouble ignoring irrelevant data and also do not exhibit a kind of normal inhibitory use of negative information (i.e., they do not exhibit *negative priming*) [30]. Furthermore, schizophrenics' deficit in inhibitory processes may occur at a *later* stage of processing than their difficulties with filtering out "noise" [30]. Theorem 17, then, provides the beginnings of a possible mathematical, causal explanation: schizophrenia, in effect, gives people noisy input and, then, their deficient, net behavior is subsumable by that of a noise-free(r) normal who just ignores negative information. It would be interesting to get some learning theory characterizations extending Theorem 17 and which show a *necessity* for some negative information blindness in the face of noise.

Theorem 17 *Suppose* $a, b \in N \cup \{*\}$.
(a) $\mathbf{NoisyInfFex}_b^a \subseteq \mathbf{TxtFex}_b^a$.
(b) $\mathbf{NoisyInfOex}_b^a \subseteq \mathbf{TxtOex}_b^a$.
(c) $\mathbf{NoisyInfBc}^a \subseteq \mathbf{TxtBc}^a$.

PROOF. An idea similar to that used in this proof was also used by Lange and Zeugmann [26]. The proof is based on the fact that any text T for L can be translated into a noisy informant I_T for L via

$$I_T(\langle i,j\rangle) = \begin{cases} (i,1), & \text{if } i \in \text{content}(T[\langle i,j\rangle]); \\ (i,0), & \text{if } i \notin \text{content}(T[\langle i,j\rangle]). \end{cases}$$

Note that, $I_T[n]$ can be obtained effectively from $T[n]$.

For a given $\mathbf{M}$, let $\mathbf{M}'$ be defined as follows:

$$\mathbf{M}'(T[n]) = \mathbf{M}(I_T[n]).$$

Since T is a text for L iff I_T is a noisy informant for L, we have, $\mathbf{NoisyInfFex}_b^a(\mathbf{M}) \subseteq \mathbf{TxtFex}_b^a(\mathbf{M}')$, $\mathbf{NoisyInfOex}_b^a(\mathbf{M}) \subseteq \mathbf{TxtOex}_b^a(\mathbf{M}')$, and $\mathbf{NoisyInfBc}^a(\mathbf{M}) \subseteq \mathbf{TxtBc}^a(\mathbf{M}')$. □

We next show that learning from noisy texts and noisy informants are incomparable.

Theorem 18 *(a)* $\mathbf{NoisyInfEx} - (\mathbf{NoisyTxtOex}_*^* \cup \mathbf{NoisyTxtBc}^*) \neq \emptyset$.
(b) $\mathbf{NoisyTxtEx} - (\mathbf{NoisyInfOex}_*^* \cup \mathbf{NoisyInfBc}^*) \neq \emptyset$.

PROOF. (a) For $x \in N$, let $L_x = \{\langle y,z\rangle : z \in N \wedge y \geq x\}$. Let $\mathcal{L} = \{L_x : x \in N\}$. Clearly, $\mathcal{L} \in \mathbf{InfEx}_0^0 \subseteq \mathbf{NoisyInfEx}$.

Note that $L_1 \subseteq L_0$ and $L_1 \neq^* L_0$. Thus by Theorem 14, $\mathcal{L} \notin \mathbf{NoisyTxtBc}^*$.

We now show that $\mathcal{L} \notin \mathbf{NoisyTxtOex}_*^*$. Suppose by way of contradiction that $\mathcal{L} \subseteq \mathbf{NoisyTxtOex}_*^*(\mathbf{M})$ via $\mathbf{M}$. Then there exists a $\mathbf{NoisyTxtOex}_*^*$-locking sequence σ for $\mathbf{M}$ on L_0. Let n be large enough so that $\text{Last}_*(\mathbf{M}, \sigma)$

does not contain a grammar for a finite variant of L_n (note that there exists such an n). Now, for any text T for L_n, $\mathrm{Last}_*(\mathbf{M}, \sigma \diamond T)$ $(= \mathrm{Last}_*(\mathbf{M}, \sigma))$ does not contain a grammar for a finite variant of L_n. A contradiction. Thus $\mathcal{L} \notin \mathbf{NoisyTxtOex}_*^*$.

(b) Let $L_0 = \{\langle x, 0\rangle : x \in N\}$. For $i \geq 0$, let $L_i = \{\langle x, 0\rangle : x \leq i\} \cup \{\langle x, i\rangle : x > i\}$.

Let $\mathcal{L} = \{L_i : i \in N\}$.

It is easy to verify that $\mathcal{L} \in \mathbf{NoisyTxtEx}$.

We first show that $\mathcal{L} \notin \mathbf{NoisyInfOex}_*^*$. Suppose by way of contradiction that $\mathcal{L} \subseteq \mathbf{NoisyInfOex}_*^*(\mathbf{M})$. Then there exists a $\mathbf{NoisyInfOex}_*^*$-locking sequence σ for $\mathbf{M}$ on L_0. Thus,

$$(\forall \tau \in \mathrm{Inf}[\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma), L_0])[W_{\mathbf{M}(\sigma \diamond \tau)} \in \mathrm{Last}_*(\mathbf{M}, \sigma)].$$

Let n be such that, $n > \max(\{x \colon (\exists y)[\langle x, y\rangle \in \mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma)]\})$, and $\mathrm{Last}_*(\mathbf{M}, \sigma)$ does not contain a grammar for a finite variant of L_n (note that since L_i's are pairwise infinitely different, there exists such an n). Now, for any informant I for L_n, $\mathrm{Last}_*(\mathbf{M}, \sigma \diamond I)$ $(= \mathrm{Last}_*(\mathbf{M}, \sigma))$ does not contain a grammar for a finite variant of L_n. A contradiction. Thus $\mathcal{L} \notin \mathbf{NoisyInfOex}_*^*$.

We now show that $\mathcal{L} \notin \mathbf{NoisyInfBc}^*$. Suppose by way of contradiction that $\mathcal{L} \subseteq \mathbf{NoisyInfBc}^*(\mathbf{M})$. Then there exists a $\mathbf{NoisyInfBc}^*$-locking sequence σ for $\mathbf{M}$ on L_0. Thus,

$$(\forall \tau \in \mathrm{Inf}[\mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma), L_0])[W_{\mathbf{M}(\sigma \diamond \tau)} =^* L_0].$$

Let $n > \max(\{x \colon (\exists y)[\langle x, y\rangle \in \mathrm{Pos}(\sigma) \cup \mathrm{Neg}(\sigma)]\})$. Now, for any informant I for L_n, for all $\tau \preceq I$, $[W_{\mathbf{M}(\sigma \diamond \tau)} =^* L_0]$. Thus, since $L_0 \neq^* L_n$, $\mathbf{M}$ does not $\mathbf{NoisyInfBc}^*$-identify L_n. A contradiction. Thus $\mathcal{L} \notin \mathbf{NoisyInfBc}^*$. □

The following theorem shows that $\mathbf{InfFex}_*^n$ and $\mathbf{InfOex}_*^n$ are same even in the presence of noise. However this equality breaks down for noisy texts.

Theorem 19 *Suppose $n \in N$ and $b \in N \cup \{*\}$.*
(a) $\mathbf{NoisyInfFex}_b^n = \mathbf{NoisyInfOex}_b^n$.
(b) $\mathbf{NoisyTxtOex}_2 - \mathbf{NoisyTxtBc}^* \neq \emptyset$.
(c) $(\mathbf{NoisyTxtOex}_2^* \cap \mathbf{NoisyInfOex}_2^*) - \mathbf{TxtBc}^* \neq \emptyset$.

Proof. (a) By replacing content(σ) by Pos(σ), in the definition of $Q(\sigma, e)$ in the proof for $\mathbf{TxtFex}_b^n = \mathbf{TxtOex}_b^n$ (Theorem 11), we can show that $\mathbf{NoisyInfFex}_b^n = \mathbf{NoisyInfOex}_b^n$.

(b) Let $\mathcal{L} = \{\emptyset, N\}$. Clearly, $\mathcal{L} \in \mathbf{NoisyTxtOex}_2$. It follows from Theorem 14 that $\mathcal{L} \notin \mathbf{NoisyTxtBc}^*$.

(c) Let $\mathcal{L} = \{L : L \text{ is finite } \} \cup \{N\}$. Clearly, $\mathcal{L} \in \mathbf{NoisyTxtOex}_2^* \cap \mathbf{NoisyInfOex}_2^*$ (by using the grammars for $\emptyset$ and N). However $\mathcal{L} \notin \mathbf{TxtBc}^*$ [22,17]. □

Corollary 20 $\mathbf{NoisyInfFex}_* = \mathbf{NoisyInfOex}_*$.
$\mathbf{NoisyTxtFex}_* \subset \mathbf{NoisyTxtOex}_*$.

In the remainder of this section we prove results which show that if $\mathcal{J}_1 - \mathcal{J}_2 \neq \emptyset$, (where $\mathcal{J}_1, \mathcal{J}_2$ are inference criteria without noise, $\mathcal{J}_1$ being a criteria not involving mind changes), then, in most cases, $\mathbf{Noisy}\mathcal{J}_1 - \mathcal{J}_2 \neq \emptyset$. We also note the exceptions. (The noisy inference criteria involving mind changes, are considered in the next section). Reader should compare the theorems with the corresponding diagonalization results mentioned in Theorem 10.

Theorem 21 *Suppose $a \in N \cup \{*\}$ and $n \in N$. Suppose $\mathcal{L} \subseteq \mathcal{SVT}$. Then*
(a) $\mathcal{L} \in \mathbf{TxtEx}^a \Leftrightarrow \mathcal{L} \in \mathbf{InfEx}^a \Leftrightarrow \mathcal{L} \in \mathbf{NoisyTxtEx}^a \Leftrightarrow$
$\mathcal{L} \in \mathbf{TxtFex}_*^a \Leftrightarrow \mathcal{L} \in \mathbf{InfFex}_*^a \Leftrightarrow \mathcal{L} \in \mathbf{NoisyTxtFex}_*^a$.
(b) $\mathcal{L} \in \mathbf{TxtOex}_b^a \Leftrightarrow \mathcal{L} \in \mathbf{InfOex}_b^a \Leftrightarrow \mathcal{L} \in \mathbf{NoisyTxtOex}_b^a$.
(c) $\mathcal{L} \in \mathbf{TxtBc}^a \Leftrightarrow \mathcal{L} \in \mathbf{InfBc}^a \Leftrightarrow \mathcal{L} \in \mathbf{NoisyTxtBc}^a$.
(d) $\mathcal{L} \in \mathbf{TxtEx}_n^a \Leftrightarrow \mathcal{L} \in \mathbf{InfEx}_n^a$.

Proof. For $\mathcal{L} \subseteq \mathcal{SVT}$, the equivalences,

(i) $\mathcal{L} \in \mathbf{TxtEx}^a \Leftrightarrow \mathcal{L} \in \mathbf{InfEx}^a$,

(ii) $\mathcal{L} \in \mathbf{TxtFex}_*^a \Leftrightarrow \mathcal{L} \in \mathbf{InfFex}_*^a$,

(iii) $\mathcal{L} \in \mathbf{TxtOex}_b^a, \Leftrightarrow \mathcal{L} \in \mathbf{InfOex}_b^a$,

(iv) $\mathcal{L} \in \mathbf{TxtBc}^a \Leftrightarrow \mathcal{L} \in \mathbf{InfBc}^a$, and

(v) $\mathcal{L} \in \mathbf{TxtEx}_n^a \Leftrightarrow \mathcal{L} \in \mathbf{InfEx}_n^a$,

hold since a text for $L \in \mathcal{SVT}$ can be effectively converted to an informant for L. Thus, it is sufficient to show

(i) $\mathcal{L} \in \mathbf{TxtEx}^a \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtEx}^a$,

(ii) $\mathcal{L} \in \mathbf{TxtFex}_*^a \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtFex}_*^a$,

(iii) $\mathcal{L} \in \mathbf{TxtOex}_*^a \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtOex}_*^a$,

(iv) $\mathcal{L} \in \mathbf{TxtBc}^a \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtBc}^a$.

The idea of the proof is to convert a noisy text for $L \in \mathcal{SVT}$, limit effectively, into a text for L (similar technique was also used in [19,20]). This is done as follows.

For a text T, let F_T be defined as follows:

$$F_T(i) = \begin{cases} \langle x, y\rangle, & \text{if } T(i) = \langle x, y\rangle, \text{ and} \\ & (\forall j \geq i)[T(j) = \langle x, z\rangle \Rightarrow y = z]; \\ \#, & \text{otherwise.} \end{cases}$$

Let, $G_{T[n]}$ be a sequence of length n defined as follows. For $i < n$,

$$G_{T[n]}(i) = \begin{cases} \langle x, y\rangle, & \text{if } T(i) = \langle x, y\rangle, \text{ and} \\ & (\forall j : i \leq j < n)[T(j) = \langle x, z\rangle \Rightarrow y = z]; \\ \#, & \text{otherwise.} \end{cases}$$

Suppose $L \in \mathcal{SVT}$, and T is noisy text for L. Then it is easy to verify that,

(i) F_T is a text for L, and

(ii) for all but finitely many n, $G_{T[n]} = F_{T[n]}$.

For a given $\mathbf{M}$, let $\mathbf{M}'$ be defined as follows:

$$\mathbf{M}'(T[n]) = \mathbf{M}(G_{T[n]}).$$

Since, for $L \in \mathcal{SVT}$, T is a noisy text for L iff F_T is a text for L, it follows that

(i) $\mathcal{L} \subseteq \mathbf{TxtEx}^a(\mathbf{M}) \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtEx}^a(\mathbf{M}')$,

(ii) $\mathcal{L} \subseteq \mathbf{TxtFex}^a_*(\mathbf{M}) \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtFex}^a_*(\mathbf{M}')$,

(iii) $\mathcal{L} \subseteq \mathbf{TxtOex}^a_*(\mathbf{M}) \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtOex}^a_*(\mathbf{M}')$, and

(iv) $\mathcal{L} \subseteq \mathbf{TxtBc}^a(\mathbf{M}) \Rightarrow \mathcal{L} \in \mathbf{NoisyTxtBc}^a(\mathbf{M}')$.

The theorem follows. $\square$

Theorem 22 *Suppose $n \in N$.*
(a) $\mathbf{NoisyTxtEx}^{n+1} - \mathbf{InfOex}^n_* \neq \emptyset$. $\mathbf{NoisyTxtEx}^* - \bigcup_{n\in N} \mathbf{InfOex}^n_* \neq \emptyset$.
(b) $\mathbf{NoisyTxtEx} - \bigcup_{n\in N} \mathbf{InfEx}^*_n \neq \emptyset$.
(c) $\mathbf{NoisyTxtFex}^*_* \subseteq \mathbf{InfBc}$.
(d) $\mathbf{NoisyTxtFex}^{2n}_* \subseteq \mathbf{TxtBc}^n$.
(e) $\mathbf{NoisyTxtEx}^{2n+1}_0 - \mathbf{TxtBc}^n \neq \emptyset$. $\mathbf{NoisyTxtEx}^*_0 - \bigcup_{n\in N} \mathbf{TxtBc}^n \neq \emptyset$.

(f) $\mathbf{NoisyTxtEx}_0^{n+1} - \mathbf{NoisyTxtBc}^n \neq \emptyset$.

Proof. (a), (b) Case and Smith [18] showed that there exist $\mathcal{L}, \mathcal{L}', \mathcal{L}'' \subseteq \mathcal{SVT}$ such that $\mathcal{L} \in \mathbf{TxtEx}^{n+1} - \mathbf{InfOex}_*^n$, $\mathcal{L}' \in \mathbf{TxtEx}^* - \bigcup_{n\in N} \mathbf{InfOex}_*^n$, and $\mathcal{L}'' \in \mathbf{TxtEx} - \bigcup_{n\in N} \mathbf{TxtEx}_n^*$. (a), (b), now follows from Theorem 21.

(c) Follows from the fact that $\mathbf{InfFex}_*^* \subseteq \mathbf{InfBc}$ (Theorem 10).

(d) Follows from the fact that $\mathbf{TxtFex}_*^{2n} \subseteq \mathbf{TxtBc}^n$ (Theorem 10).

(e) For $a \in N \cup \{*\}$, let $\mathcal{L}_a = \{L : L =^a N\}$. Clearly, $\mathcal{L}_a \in \mathbf{NoisyTxtEx}_0^a$. It was shown in [17] that $\mathcal{L}_{2n+1} \notin \mathbf{TxtBc}^n$ and $\mathcal{L}_* \notin \bigcup_{n\in N} \mathbf{TxtBc}^n$.

(f) Let L, L_1, L_2 be r.e. sets such that $L_1 \subseteq L \subseteq L_2$ and $\text{card}(L_2 - L) = \text{card}(L - L_1) = n + 1$. It is easy to verify that $\{L_1, L_2\} \in \mathbf{NoisyTxtEx}_0^{n+1}$. $\{L_1, L_2\} \notin \mathbf{NoisyTxtBc}^n$ follows by Theorem 14. □

Theorem 23 *Suppose* $n \in N$.
(a) $\mathbf{NoisyInfEx}^{n+1} - \mathbf{InfFex}_*^n \neq \emptyset$. $\mathbf{NoisyInfEx}^* - \bigcup_{n\in N} \mathbf{InfFex}_*^n \neq \emptyset$.
(b) $\mathbf{NoisyInfEx} - \bigcup_{n\in N} \mathbf{InfEx}_n^* \neq \emptyset$.
(c) $\mathbf{NoisyInfFex}_*^* \subseteq \mathbf{InfBc}$.
(d) $\mathbf{NoisyInfFex}_*^{2n} \subseteq \mathbf{NoisyInfBc}^n \subseteq \mathbf{TxtBc}^n$.
(e) $\mathbf{NoisyInfEx}_0^{2n+1} - \mathbf{TxtBc}^n \neq \emptyset$. $\mathbf{NoisyInfEx}_0^* - \bigcup_{n\in N} \mathbf{TxtBc}^n \neq \emptyset$.

Proof. (a), (b) Follow using Theorem 12 and the facts that $\mathbf{InfEx}_0^{n+1} \not\subseteq \mathbf{InfEx}^n = \mathbf{InfFex}_*^n$, $\mathbf{InfEx}_0^* \not\subseteq \bigcup_{n\in N} \mathbf{InfEx}^n = \bigcup_{n\in N} \mathbf{InfFex}_*^n$ [18] and $\mathbf{InfEx}_0[K] \not\subseteq \bigcup_{n\in N} \mathbf{InfEx}_n^*$. (Gasarch and Pleszkoch [21] showed that $\mathbf{InfEx}_0[K] \not\subseteq \bigcup_{n\in N} \mathbf{InfEx}_n$. Cylindrification of their result gives $\mathbf{InfEx}_0[K] - \bigcup_{n\in N} \mathbf{InfEx}_n^* \neq \emptyset$. Also one can prove $\mathbf{InfEx}_0[K] - \mathbf{InfEx}_n^*$ by considering the following $\mathcal{L}$: let, $\mathcal{C} = \{f : W_{f(0)} \neq N \ \wedge \ \text{card}(\{x : f(x) \neq f(x+1)\}) = \min(\overline{W_{f(0)}})\}$; let $\mathcal{L} = \{L : L$ represents some $f \in \mathcal{C}\}$.)

(c) Follows from the fact that $\mathbf{InfFex}_*^* \subseteq \mathbf{InfBc}$.

(d) The idea is essentially the same as used to prove $\mathbf{TxtFex}_*^{2n} \subseteq \mathbf{TxtBc}^n$ from [17,12,13]. Suppose $\mathbf{M}$ is given. $\mathbf{M}'(\sigma)$ is defined as follows.

Let S_σ be the least n elements in $\text{Pos}(\sigma) \mathbf{\Delta} W_{\mathbf{M}(\sigma),|\sigma|}$ (if $\text{Pos}(\sigma) \mathbf{\Delta} W_{\mathbf{M}(\sigma),|\sigma|}$ contains less than n elements, then $S_\sigma = \text{Pos}(\sigma) \mathbf{\Delta} W_{\mathbf{M}(\sigma),|\sigma|}$).

Now $\mathbf{M}'(\sigma)$ is a grammar for $[W_{\mathbf{M}(\sigma),|\sigma|} - S_\sigma] \cup [S_\sigma \cap \text{Pos}(\sigma)]$ (i.e. we obtain $\mathbf{M}'(\sigma)$ by patching the grammar $\mathbf{M}(\sigma)$, based on the elements in S_σ).

The argument to prove that $\mathbf{M}'$ $\mathbf{NoisyInfBc}^n$-identifies every language

$\mathbf{NoisyInfFex}_*^{2n}$-identified by $\mathbf{M}$ is essentially the same as used by [17,12,13]. We omit the details.

(e) For $a \in N \cup \{*\}$, let $\mathcal{L}_a = \{L\colon L =^a N\}$. Clearly, $\mathcal{L}_a \in \mathbf{NoisyInfEx}_0^a$. It was shown in [17] that $\mathcal{L}_{2n+1} \notin \mathbf{TxtBc}^n$ and $\mathcal{L}_* \notin \bigcup_{n \in N} \mathbf{TxtBc}^n$. □

Interestingly, as we see by Corollary 25 to the following theorem (Theorem 24), the hierarchy $\mathbf{NoisyInfFex}_1 \subset \mathbf{NoisyInfFex}_2 \subset \ldots \subset \mathbf{NoisyInfFex}_*$ is proper. This contrasts sharply with the non-noisy case. **Fex** style criteria, in situations taking into account noise (as here), missing information (as in [11–14]), or complexity constraints (as in [15]), provide a hierarchy; but, unconstrained, do not (as in [7,18]).

Theorem 24 *Suppose* $n \geq 1$.
(a) $(\mathbf{NoisyInfFex}_{n+1} \cap \mathbf{NoisyTxtFex}_{n+1}) - \mathbf{TxtOex}_n^* \neq \emptyset$.
(b) $(\mathbf{NoisyInfFex}_* \cap \mathbf{NoisyTxtFex}_*) - \bigcup_{n \in N} \mathbf{TxtOex}_n^* \neq \emptyset$.

PROOF. (a) Let $\mathrm{Null}_L = \{y\colon \langle 0, y\rangle \in L\}$. Let

$$
\begin{aligned}
\mathcal{L}_n = \{L\colon & \mathrm{card}(L) = \infty \ \wedge\ (\exists S\colon \mathrm{card}(S) = n+1)[\\
& S = \mathrm{Null}_L \ \wedge \\
& (\forall \langle x, y\rangle \in L)[y \in S] \ \wedge \\
& (\forall^\infty \langle x, y\rangle \in L)[W_y = L] \\
]\}.
\end{aligned}
$$

We will show that $\mathcal{L}_n \in (\mathbf{NoisyInfFex}_{n+1} \cap \mathbf{NoisyTxtFex}_{n+1}) - \mathbf{TxtOex}_n^*$.

Clearly, $\mathcal{L}_n \in \mathbf{NoisyTxtFex}_{n+1}$. We next show that $\mathcal{L}_n \in \mathbf{NoisyInfOex}_{n+1}$ (which by Theorem 19 implies that $\mathcal{L}_n \in \mathbf{NoisyInfFex}_{n+1}$). Let $S_\sigma = \{e\colon \langle 0, e\rangle \in \mathrm{Pos}(\sigma)\}$. Let S'_σ denote the least $n+1$ elements in S_σ (if cardinality of S_σ is smaller than $n+1$, then $S'_\sigma = S_\sigma$). It is easy to verify that, for any noisy information sequence I for $L \in \mathcal{L}_n$, for all but finitely many $\sigma \preceq I$, $S'_\sigma = \mathrm{Null}_L$. Thus, using S'_σ, one can easily construct a machine $\mathbf{M}$ such that, for any noisy information sequence I for $L \in \mathcal{L}_n$, $\mathrm{Last}_{n+1}(\mathbf{M}, I) = \mathrm{Null}_L$. It follows that $\mathcal{L}_n \subseteq \mathbf{NoisyInfOex}_{n+1}(\mathbf{M})$.

Now, suppose by way of contradiction that $\mathbf{M}$ $\mathbf{TxtOex}_n^*$-identifies $\mathcal{L}_n$. By implicit use of $n+1$-ary recursion theorem, there exist $e_1 < e_2 < \cdots < e_{n+1}$ such that, for $i = 1, 2, \ldots, n+1$, W_{e_i} may be defined as follows. Enumerate $\langle 0, e_1\rangle, \langle 0, e_2\rangle, \ldots, \langle 0, e_{n+1}\rangle$ into $W_{e_1}, W_{e_2}, \ldots, W_{e_{n+1}}$. Let σ_0 be such that $\mathrm{content}(\sigma_0) = \{\langle 0, e_1\rangle, \langle 0, e_2\rangle, \ldots, \langle 0, e_{n+1}\rangle\}$. Go to stage 0.

Stage s

1. Dovetail steps 2 and 3, until step 2 succeeds. If and when step 2 succeeds, go to step 4.

2. Search for a τ extending σ_s such that content$(\tau) \subseteq \{\langle x, y\rangle : x \in N \ \wedge \ y \in \{e_1, e_2, \ldots, e_{n+1}\}\}$ and $\text{Last}_n(\mathbf{M}, \tau) \neq \text{Last}_n(\mathbf{M}, \sigma_s)$.
3. Let $x = 0$
   loop
      For $i = 1, 2, \ldots, n+1$, enumerate $\langle x, e_i\rangle$ in W_{e_i}
      Let $x = x + 1$.
   forever
4. Let τ be as found in step 2.
   Let $S = \text{content}(\tau) \cup \bigcup_{1 \leq i \leq n+1} [W_{e_i}$ enumerated until now $] \cup \{\langle s, y\rangle : y \in \{e_1, e_2, \ldots, e_{n+1}\}\}$.
   For $i = 1, 2, \ldots, n+1$, enumerate S into W_{e_i}.
   Let σ_{s+1} be an extension of τ such that content$(\sigma_{s+1}) = S$.
   Go to stage $s + 1$.

End stage s

We now consider two cases.

Case 1: All stages finish.

In this case let $L = W_{e_1}$ $(= W_{e_2} = \cdots = W_{e_{n+1}})$. Clearly $L \in \mathcal{L}_n$ and $T = \bigcup_{s \in N} \sigma_s$ is a text for L. However $\text{Last}_n(\mathbf{M}, T)$ is undefined (due to success of step 2 infinitely often, we have that $\mathbf{M}$ does not converge on T to a set of n grammars).

Case 2: Some stage s starts but does not finish.

In this case, for $i \in \{1, 2, \ldots, n+1\}$, let $L_i = W_{e_i}$. Note that these L_i's are pairwise infinitely different (due to step 3 in stage s). Let i be such that no grammar in $\text{Last}_n(\sigma_s)$ is a grammar for $*$-variant of L_i (by pigeonhole principle, there exists such a i). Let T_i be a text for L_i such that $\sigma_s \preceq T_i$. It follows that $\text{Last}_n(\mathbf{M}, T_i) = \text{Last}_n(\mathbf{M}, \sigma_s)$ does not contain a grammar for $*$-variant of L_i. Thus $\mathbf{M}$ does not $\mathbf{TxtOex}_n^*$ identify L_i.

From the above cases it follows that $\mathbf{M}$ does not $\mathbf{TxtOex}_n^*$ identify $\mathcal{L}_n$.

(b) For any L, n let $X_L^n = \{\langle 0, n\rangle\} \cup \{\langle 1, x\rangle : x \in L\}$. Let $\mathcal{L}_n$ be as in part (a).

Let $\mathcal{L}'_n = \{X_n^L : L \in \mathcal{L}_n\}$ and let $\mathcal{L} = \bigcup_{n \in N} \mathcal{L}'_n$.

An easy modification of the proof of part (a) shows $\mathcal{L}'_n \notin \mathbf{TxtOex}_n^*$. Thus, $\mathcal{L} \notin \mathbf{TxtOex}_n^*$.

In a way similar to that of part (a) one can show that $\mathcal{L} \in (\mathbf{NoisyInfFex}_* \cap \mathbf{NoisyTxtFex}_*)$. $\square$

Corollary 25 *Suppose* $a \in (N \cup \{*\})$.
$\mathbf{NoisyInfOex}_1^a \subset \mathbf{NoisyInfOex}_2^a \subset \ldots \subset \mathbf{NoisyInfOex}_*^a$.
$\mathbf{NoisyInfFex}_1^a \subset \mathbf{NoisyInfFex}_2^a \subset \ldots \subset \mathbf{NoisyInfFex}_*^a$.
$\mathbf{NoisyTxtOex}_1^a \subset \mathbf{NoisyTxtOex}_2^a \subset \ldots \subset \mathbf{NoisyTxtOex}_*^a$.
$\mathbf{NoisyTxtFex}_1^a \subset \mathbf{NoisyTxtFex}_2^a \subset \ldots \subset \mathbf{NoisyTxtFex}_*^a$.

The next theorem establishes the hierarchy

$$\mathbf{NoisyTxtBc} \subset \mathbf{NoisyTxtBc}^1 \subset \ldots \subset \mathbf{NoisyTxtBc}^*.$$

Theorem 26 *Suppose* $n \in N$.
(a) $\mathbf{NoisyTxtBc} - \mathbf{InfOex}_*^* \neq \emptyset$.
(b) $\mathbf{NoisyTxtBc}^{n+1} - \mathbf{InfBc}^n \neq \emptyset$. $\mathbf{NoisyTxtBc}^* - \bigcup_{n \in N} \mathbf{InfBc}^n \neq \emptyset$.

Proof. For $a \in N \cup \{*\}$, let $\mathcal{L}_a = \{L \colon \mathrm{card}(L) = \infty \ \wedge\ (\forall^\infty x \in L)\, [W_x =^a L]\}$. It is easy to verify that $\mathcal{L}_a \in \mathbf{NoisyTxtBc}^a$.

Adopting the techniques used by Case and Smith [18] to show $\mathbf{Bc}^{n+1} \not\subseteq \mathbf{Bc}^n$ and $\mathbf{Bc} \not\subseteq \mathbf{Ex}^*$, one can show that $\mathcal{L}_0 \notin \mathbf{InfOex}_*^*$, $\mathcal{L}_{n+1} \notin \mathbf{InfBc}^n$ and $\mathcal{L}_* \notin \bigcup_{n \in N} \mathbf{InfBc}^n$. We omit the details. □

Theorem 27 *Suppose* $n \in N$.
(a) $\mathbf{NoisyInfBc}^{n+1} - \mathbf{InfBc}^n \neq \emptyset$. $\mathbf{NoisyInfBc}^* - \bigcup_{n \in N} \mathbf{InfBc}^n \neq \emptyset$.
(b) $\mathbf{NoisyInfBc}^1 - \mathbf{InfOex}_*^* \neq \emptyset$.
(c) $\mathbf{NoisyInfBc} - \mathbf{TxtOex}_*^* \neq \emptyset$.
(d) $\mathbf{NoisyInfBc} \subseteq \mathbf{InfEx}$.

Proof. (a), (b) Let $\mathcal{L}_a = \{L \colon (\forall x \in W_{\min(L)})\, [W_x =^a L] \vee (\mathrm{card}(W_{\min(L)}) < \infty \ \wedge\ W_{\max(W_{\min(L)})} =^a L)\}$.

It is easy to verify that $\mathcal{L}_{n+1} \in \mathbf{NoisyInfBc}^{n+1}$.

The proof of $\mathbf{Bc}^{n+1} - \mathbf{Bc}^n \neq \emptyset$ and $\mathbf{Bc} - \mathbf{InfOex}_*^* \neq \emptyset$ from [18] can be easily adopted to show that $\mathcal{L}_{n+1} \notin \mathbf{InfBc}^n$, $\mathcal{L}_* \notin \bigcup_{n \in N} \mathbf{InfBc}^n$, and $\mathcal{L}_1 \notin \mathbf{InfOex}_*^*$. We omit the details.

(c) Let $L_x^1 = \{\langle w, z\rangle \colon w \in N \ \wedge\ z \geq x\}$. Let $L_{x,y}^2 = \{\langle w, z\rangle \colon w \in N \ \wedge\ x \leq z \leq y\}$. Let $\mathcal{L} = \{L_x^1 \colon \mathrm{card}(W_x) = \infty\} \cup \{L_{x,y}^2 \colon \mathrm{card}(W_x) < \infty \ \wedge\ y > x\}$.

We show that $\mathcal{L} \in \mathbf{NoisyInfBc} - \mathbf{TxtOex}_*^*$.

Let $\mathbf{M}(\sigma)$ be defined as follows. Let g be a recursive function such that

$$W_{g(x,y,n)} = \begin{cases} L_{x,y}, & \text{if } \mathrm{card}(W_x) \leq n; \\ L_x, & \text{if } \mathrm{card}(W_x) > n. \end{cases}$$

$$\mathbf{M}(\sigma) = \begin{cases} ?, & \text{if } \mathrm{Pos}(\sigma) = \emptyset; \\ g(x, y, |\sigma|), & \text{if } \mathrm{Pos}(\sigma) \neq \emptyset, \\ & \text{where } x = \min(\{x' : \langle 0, x' \rangle \in \mathrm{Pos}(\sigma)\}) \text{ and} \\ & y = \min(\{z : z > x \ \wedge \ \langle 0, z \rangle \notin \mathrm{Pos}(\sigma)\}). \end{cases}$$

We claim that $\mathcal{L} \subseteq \mathbf{NoisyInfBc}(\mathbf{M})$.

If $\mathrm{card}(W_x) = \infty$ and I is a noisy informant for L^1_x, then, for all but finitely many $\sigma \preceq I$, $x = \min(\{x' : \langle 0, x' \rangle \in \mathrm{Pos}(\sigma)\})$. For such σ, since $\mathrm{card}(W_x) = \infty$, $\mathbf{M}(\sigma) = g(x, y, |\sigma|)$ is a grammar for L^1_x. Thus $L^1_x \in \mathbf{NoisyInfBc}(\mathbf{M})$.

If $\mathrm{card}(W_x) < \infty$, $y > x$, and I is a noisy informant for $L^2_{x,y}$, then, for all but finitely many $\sigma \preceq I$, $x = \min(\{x' : \langle 0, x' \rangle \in \mathrm{Pos}(\sigma)\})$, $y = \min(\{z : z > x \ \wedge \langle 0, z \rangle \notin \mathrm{Pos}(\sigma)\})$, and $|\sigma| > \mathrm{card}(W_x)$. Thus, for all but finitely many $\sigma \preceq I$, $\mathbf{M}(\sigma) = g(x, y, |\sigma|)$ is a grammar for $L^2_{x,y}$. Thus $L^2_{x,y} \in \mathbf{NoisyInfBc}(\mathbf{M})$.

It follows that $\mathcal{L} \in \mathbf{NoisyInfBc}$.

Now suppose by way of contradiction, that $\mathbf{M}$ $\mathbf{TxtOex}^*_*$-identifies $\mathcal{L}$.

If $\mathrm{card}(W_x) = \infty$, then there exists a $\mathbf{TxtOex}^*_*$-locking sequence for $\mathbf{M}$ on L^1_x. Thus,

$$(\exists \sigma : \mathrm{content}(\sigma) \subseteq L^1_x)(\forall \tau : \mathrm{content}(\tau) \subseteq L^1_x)[\mathbf{M}(\sigma \diamond \tau) \in \mathrm{Last}_*(\mathbf{M}, \sigma)].$$

If $\mathrm{card}(W_x) < \infty$, then there is no such sequence:

$$(\forall \sigma : \mathrm{content}(\sigma) \subseteq L^1_x)(\exists y)(\exists \tau : \mathrm{content}(\tau) \subseteq L^1_{x,y} \subseteq L^1_x)$$

$$[\mathbf{M}(\sigma \diamond \tau) \notin \mathrm{Last}_*(\mathbf{M}, \sigma)].$$

This is so, since $L^2_{x,y}$ are pairwise infinitely different, and thus $\mathrm{Last}_*(\mathbf{M}, \sigma)$, can contain a grammar for a finite variant of only finitely many $L^2_{x,y}$.

It follows that

$$\mathrm{card}(W_x) = \infty \Leftrightarrow$$
$$(\exists \sigma : \mathrm{content}(\sigma) \subseteq L^1_x)(\forall \tau : \mathrm{content}(\tau) \subseteq L^1_x)[\mathbf{M}(\sigma \diamond \tau) \in \mathrm{Last}_*(\mathbf{M}, \sigma)].$$

But, this would mean that $\{x : \mathrm{card}(W_x) = \infty\}$ is r.e. in K. A contradiction. Thus no such $\mathbf{M}$ can exists.

(d) Suppose $\mathbf{M}$ is given. We generalize the notion of a locking sequence from Proposition 9 to that of a good pair $\langle \sigma, l \rangle$:

> $\langle \sigma, l \rangle$ is *good for* $\mathbf{M}$ *on* L iff, for all $\tau \in \mathrm{Inf}[\{0, 1, \cdots, l-1\}, L]$, $W_{\mathbf{M}(\sigma \diamond \tau)} \subseteq L$.

Note that, for every $L \in \mathbf{NoisyTxtBc}(\mathbf{M})$, there exists a good pair for $\mathbf{M}$ on L. Let

$$X^L_{\sigma,l} = \bigcup_{\tau \in \mathrm{Inf}[\{0,1,\cdots,l-1\},L]} W_{\mathbf{M}(\sigma\diamond\tau)}.$$

Let g be a recursive function such that $g(\sigma, l, \chi_L[l])$ is a grammar for $X^L_{\sigma,l}$. Note that there exists such a recursive g.

Claim 28 *If* $L \in \mathbf{NoisyTxtBc}(\mathbf{M})$*, and* $\langle\sigma, l\rangle$ *is good for* $\mathbf{M}$ *on* L*, then* $X^L_{\sigma,l} = L$*.*

Proof. Suppose $L \in \mathbf{NoisyTxtBc}(\mathbf{M})$, and $\langle\sigma, l\rangle$ is good for $\mathbf{M}$ on L. Clearly, $X^L_{\sigma,l} \subseteq L$, since otherwise $\langle\sigma, l\rangle$ would not be good for $\mathbf{M}$ on L. We now show that $L \subseteq X^L_{\sigma,l}$. Let I be an informant for L such that, for each x, $\langle x, \chi_L(x)\rangle$ appears infinitely often in I. Then, $\sigma \diamond I$ is a noisy informant for L. Thus there exists a $\tau \preceq I$ such that $W_{\mathbf{M}(\sigma\diamond\tau)} = L$. Since $\tau \in \mathrm{Inf}[\{0, 1, \cdots, l-1\}, L]$, we have $L \subseteq X^L_{\sigma,l}$. □

We now give a machine $\mathbf{M}'$ such that $\mathbf{NoisyInfBc}(\mathbf{M}) \subseteq \mathbf{InfEx}(\mathbf{M}')$. Suppose I is an information sequence for $L \in \mathbf{NoisyInfBc}(\mathbf{M})$.

We say that $\langle\sigma', l'\rangle$ *seems good* with respect to $I[m]$ iff

$$(\forall x < l')[(x, \chi_L(x)) \in \mathrm{content}(I[m])], \text{ and}$$

$$(\forall \tau' \in \mathrm{Inf}[\{0, 1, \cdots, l'-1\}, L] : \tau' \leq m)[W_{\mathbf{M}(\sigma'\diamond\tau'),m} \cap \mathrm{Neg}(I[m]) = \emptyset].$$

$\mathbf{M}'(I[m]) = g(\sigma, l, \chi_L[l])$, where $\langle\sigma, l\rangle = \min(\{\langle\sigma', l'\rangle : \langle\sigma', l'\rangle \leq m \ \wedge \ \langle\sigma', l'\rangle$ seems good with respect to $I[m]\})$.

Intuitively, $\mathbf{M}'$ on I searches for the minimum pair $\langle\sigma, l\rangle$ such that $\langle\sigma, l\rangle$ is good for $\mathbf{M}$ on L. It then outputs $g(\sigma, l, \chi_L[l])$, in the limit, on I. By Claim 28, $g(\sigma, l, \chi_L[l])$ is a grammar for $X^L_{\sigma,l} = L$. It is now easy to verify that $L \in \mathbf{InfEx}(\mathbf{M}')$.

It follows that $\mathbf{NoisyInfBc} \subseteq \mathbf{InfEx}$. □

If one considers the definition of $\mathbf{GenEx}^a_b$ from [4] [1], then one can show that $\mathbf{GenInfEx}^a_0 \subseteq \mathbf{NoisyGenInfEx}^a \subseteq \mathbf{NoisyInfBc}^a$. Parts (a), (b) of Theorem 27 can then also be proved using the fact that $\mathbf{GenInfEx}^{n+1}_0 - \mathbf{InfBc}^n \neq \emptyset$ and $\mathbf{GenInfEx}^1_0 - \mathbf{InfOex}^*_* \neq \emptyset$. Part (d) of Theorem 27 is reminiscent of the fact that $\mathbf{GenInfEx} \subseteq \mathbf{InfEx}$.

[1] We say that p is an a-generator for L iff φ_p is total and, for all but finitely many i, $\varphi_p(i)$ is a grammar for a-variant of L (that is $W_{\varphi_p(i)} =^a L$). $\mathbf{M}$ $\mathbf{GenTxtEx}^a_b$-identifies a language L iff, on every text T for L, $\mathbf{M}$ makes at most b mind changes and converges to an a-generator for L. $\mathbf{GenInfEx}^a_b$ and the corresponding noisy inference criteria can be defined similarly.

Combining Theorem 17, which states that $\mathbf{NoisyInfFex}_* \subseteq \mathbf{TxtFex}_* \subseteq \mathbf{TxtOex}_*^*$, with $\mathbf{NoisyInfBc} \not\subseteq \mathbf{TxtOex}_*^*$, one obtains that **NoisyInfBc** is more powerful than $\mathbf{NoisyInfFex}_*$.

Corollary 29 $\mathbf{NoisyInfFex}_* \subset \mathbf{NoisyInfBc}$.

As a corollary to Theorem 19, Theorem 22, Theorem 24, and Theorem 26 we have

Corollary 30 *Suppose* $a, b \in N \cup \{*\}$ *and* $m, n \in N$.
(a) $\mathbf{NoisyTxtEx}^{n+1} - \mathbf{NoisyTxtOex}_*^n \neq \emptyset$.
(b) $\mathbf{NoisyTxtEx}^* - \bigcup_{n\in N} \mathbf{NoisyTxtOex}_*^n \neq \emptyset$.
(c) $\mathbf{NoisyTxtFex}_{n+1}^0 - \mathbf{NoisyTxtOex}_n^* \neq \emptyset$.
(d) $\mathbf{NoisyTxtFex}_*^0 - \bigcup_{n\in N} \mathbf{NoisyTxtOex}_n^* \neq \emptyset$.
(e) $\mathbf{NoisyTxtBc}^0 - \mathbf{NoisyTxtOex}_*^* \neq \emptyset$.
(f) $\mathbf{NoisyTxtBc}^{n+1} - \mathbf{NoisyTxtBc}^n \neq \emptyset$.
(g) $\mathbf{NoisyTxtBc}^* - \bigcup_{n\in N} \mathbf{NoisyTxtBc}^n \neq \emptyset$.
(h) $\mathbf{NoisyTxtOex}_2 - \mathbf{NoisyTxtBc}^* \neq \emptyset$.

As a corollary to Theorem 17, Theorem 19, Theorem 23, Theorem 24, and Theorem 27, we have

Corollary 31 *Suppose* $a, b \in N \cup \{*\}$ *and* $m, n \in N$.
(a) $\mathbf{NoisyInfEx}^{n+1} - \mathbf{NoisyInfOex}_*^n \neq \emptyset$.
(b) $\mathbf{NoisyInfEx}^* - \bigcup_{n\in N} \mathbf{NoisyInfOex}_*^n \neq \emptyset$.
(c) $\mathbf{NoisyInfFex}_{n+1}^0 - \mathbf{NoisyInfOex}_n^* \neq \emptyset$.
(d) $\mathbf{NoisyInfFex}_*^0 - \bigcup_{n\in N} \mathbf{NoisyInfOex}_n^* \neq \emptyset$.
(e) $\mathbf{NoisyInfBc}^0 - \mathbf{NoisyInfOex}_*^* \neq \emptyset$.
(f) $\mathbf{NoisyInfBc}^{n+1} - \mathbf{NoisyInfBc}^n \neq \emptyset$.
(g) $\mathbf{NoisyInfBc}^* - \bigcup_{n\in N} \mathbf{NoisyInfBc}^n \neq \emptyset$.
(h) $\mathbf{NoisyInfOex}_2^* - \mathbf{NoisyInfBc}^* \neq \emptyset$.

6 Mind Changes and Finite Variants of One Fixed R. E. Language

Theorem 32 *Suppose* $a \in N \cup \{*\}$ *and* $n \in N$.
(a) If $\mathcal{L} \in \mathbf{NoisyInfEx}_n^a$, *then there exists a grammar* i *such that* $(\forall L \in \mathcal{L})[W_i =^a L]$.
(b) If $\mathcal{L} \in \mathbf{NoisyTxtEx}_n^a$, *then there exists a grammar* i *such that* $(\forall L \in \mathcal{L})[W_i =^a L]$.

Proof. We only show part (a). Part (b) can proved similarly. Suppose $\mathcal{L} \subseteq \mathbf{NoisyInfEx}_n^a(\mathbf{M})$. Without loss of generality, assume that **M** does not make more than n mind changes on any input (noisy) information sequence. Let

σ be such that, for all τ extending σ, $\mathbf{M}(\sigma) = \mathbf{M}(\tau)$. Note that there exists such a σ, since the number of mind changes by $\mathbf{M}$ on any text is bounded. We claim that $W_{\mathbf{M}(\sigma)} =^a L$, for all $L \in \mathcal{L}$. Consider any $L \in \mathcal{L}$, and noisy information sequence I for L such that $\sigma \preceq I$ (note that there exists such an information sequence I). Now $\mathbf{M}(I) = \mathbf{M}(\sigma)$. Thus $W_{\mathbf{M}(\sigma)} =^a L$. $\square$

Corollary 33 *For all n, $\mathcal{L} \subseteq \mathcal{E}$,*

$\mathcal{L} \in \mathbf{NoisyInfEx}_n^a \Leftrightarrow \mathcal{L} \in \mathbf{NoisyTxtEx}_n^a \Leftrightarrow (\exists L \in \mathcal{E})[\mathcal{L} = \mathrm{Var}^a(L)]$.

Theorem 34 *Suppose $n \in N$.*

(a) $(\forall L \in \mathcal{E})[\mathrm{Var}^{2n}(L) \in \mathbf{NoisyInfBc}^n]$.

(b) $(\forall L \in \mathcal{E})[\mathrm{Var}^{2n}(L) \in \mathbf{TxtBc}^n]$.

(c) $(\forall L \in \mathcal{E})[\mathrm{Var}^{*}(L) \in \mathbf{InfBc}]$.

(d) $(\forall L\colon \mathrm{card}(L) = \infty)[\mathrm{Var}^{2n+1}(L) \notin \mathbf{TxtBc}^n]$.

(e) $(\forall L\colon \mathrm{card}(L) = \infty)[\mathrm{Var}^{2n+1}(N) \notin \mathbf{NoisyInfBc}^n]$.

(f) $(\forall L\colon \mathrm{card}(L) = \infty)[\mathrm{Var}^{n+1}(N) \notin \mathbf{TxtOex}_*^n]$.

(g) $(\forall L\colon \mathrm{card}(L) = \infty \ \wedge\ \mathrm{card}(\overline{L}) = \infty)[\mathrm{Var}^{n+1}(L) \notin \mathbf{NoisyTxtBc}^n]$.

PROOF. Clearly, $\mathrm{Var}^{2n}(L) \in \mathbf{NoisyInfEx}^{2n} \subseteq \mathbf{NoisyInfBc}^n \subseteq \mathbf{TxtBc}^n$. This proves part (a) and (b). Also, $\mathrm{Var}^{*}(L) \in \mathbf{NoisyInfEx}^{*} \subseteq \mathbf{InfEx}^{*} \subseteq \mathbf{InfBc}$. This proves part (c).

Case and Lynes [17] showed that $\mathrm{Var}^{2n+1}(N) \notin \mathbf{TxtBc}^n \supseteq \mathbf{NoisyInfBc}^n$, and $\mathrm{Var}^{n+1}(N) \notin \mathbf{TxtFex}_*^n = \mathbf{TxtOex}_*^n$. Their proof generalizes to any infinite L. This proves (d), (e) and (f).

For part (g), L_1, L_2 be such that $L_1 \subseteq L \subseteq L_2$ and $\mathrm{card}(L_2 - L) = \mathrm{card}(L - L_1) = n + 1$. It follows from Theorem 14 that $\{L_1, L_2\} \notin \mathbf{NoisyTxtBc}^n$. Thus, $\mathrm{Var}^{n+1}(L) \notin \mathbf{NoisyTxtBc}^n$. $\square$

Clearly, $\mathrm{Var}^{*}(N) \in \mathbf{InfEx}$. Since, for inferring a finite variant of a cylinder from informant, every *difference* from the cylinder can be detected in the limit, we have $\mathrm{Var}^{*}(K) \in \mathbf{InfEx}$. However, as the next theorem shows, this does not hold if K is replaced by a suitable (non-cylinder) r.e. set L.

Theorem 35 *Suppose $n \in N$. $\mathcal{L} = \{L'\colon L' =^{n+1} L\} \notin \mathbf{InfEx}^n$ for some r.e. L.*

PROOF. For any language L', let $I_{L'}$ denote a canonical information sequence for L'. Thus I_L is the canonical information sequence for L constructed below. Let X_i^m denote the set $\{\langle i, x\rangle : x \geq m\}$. Let $\mathbf{M}_0, \mathbf{M}_1, \ldots$ be a recursive enumeration of total learning machines such that, for all $\mathcal{L} \in \mathbf{InfEx}^n$, there exists an i, such that $\mathcal{L} \subseteq \mathbf{InfEx}^n(\mathbf{M}_i)$. (There exists such an enumeration. For example see [29].)

Then one of the following two properties will be satisfied for each i.

(A) $\mathbf{M}_i(I_L)$ diverges.
(B) There is an m such that $X_i^m - L$ is infinite and $(\forall L' : L \subseteq L' \subseteq X_i^m \cup L)[\mathbf{M}_i(I_{L'}) = \mathbf{M}(I_L)]$.

Note that this implies $\mathcal{L} \not\subseteq \mathbf{InfEx}^n(\mathbf{M}_i)$. To see this, suppose $\mathbf{M}$ $\mathbf{InfEx}^n$ identifies L. Then (B) must hold. Let m be as in (B). Thus, $X_i^m - W_{\mathbf{M}_i(I_L)}$ is infinite. Let S be a set of cardinality $n+1$ such that $S \subseteq X_i^m - W_{\mathbf{M}_i(I_L)}$. Then $\mathbf{M}_i$ does not $\mathbf{InfEx}^n$ identify $L \cup S$.

The aim of the construction below is to try to satisfy (A) above for each i (which will not always be successful). For this we place requirements,

$R_{\langle i,j\rangle}$: $\mathbf{M}_i$ on I_L makes at least j mind changes.

Fix i. In case all $R_{\langle i,j\rangle}$ are satisfied, (A) would hold. In case we cannot satisfy all $R_{\langle i,j\rangle}$ (i.e. only finitely many of them are satisfied), we will make sure that (B) holds.

In the process of trying to satisfy a requirement we need to enumerate some elements in L and constrain some elements to be out of L. Due to this, satisfying a requirement may spoil some other requirements already satisfied. To get around such problems, we order the requirements using priority. Lower numbered requirements have higher priority. We assume, without loss of generality, that, for all i, j, $\langle i, j\rangle < \langle i, j+1\rangle$. We will make sure in the construction that satisfying any requirement does not spoil any higher priority requirement.

Furthermore, in order to satisfy requirement $R_{\langle i,j\rangle}$, we will add only elements of the form $\langle i, x\rangle$ to L. This would allow us to argue that if (A) is not satisfied for some i, then (B) would be satisfied.

We let $Z_{\langle i,j\rangle}$ denote the set of elements constrained to be out of L by requirement $R_{\langle i,j\rangle}$. Initially, for all i, j, requirements $R_{\langle i,j\rangle}$ is unsatisfied and $Z_{\langle i,j\rangle}$ is empty. Let L_s denote the set of those elements which are enumerated into L before stage s. In each stage we try to satisfy the least unsatisfied requirement, which is "seen" to be satisfiable in that stage.

Definition of L

Begin stage s

1. If there exists an $\langle i,j\rangle \leq s$, such that
   (a) requirement $R_{\langle i,j\rangle}$ is currently unsatisfied and
   (b) there are σ and $S \subseteq \{\langle i,x\rangle : x \leq s \ \wedge \ (\forall \langle i',j'\rangle < \langle i,j\rangle)[\langle i,x\rangle \notin Z_{\langle i',j'\rangle}]\}$, such that $|\sigma| \leq s$, $\sigma \preceq I_{L_s \cup S}$ and $\mathbf{M}_i$ makes on σ at least j mind changes.

   Then choose the least such $\langle i_l, j_l\rangle$ and a corresponding S_l (which satisfies (b)), and proceed to step 2. Otherwise go to stage $s+1$.
2. Enumerate S_l into L.
3. Let $Z_{\langle i_l,j_l\rangle} = \{x : x \leq s\} - [L$ enumerated until now$]$.
4. (Spoil lower priority requirement)
   For $\langle i',j'\rangle > \langle i_l,j_l\rangle$, let requirement $R_{\langle i',j'\rangle}$ become unsatisfied, and let $Z_{\langle i',j'\rangle} = \emptyset$.
5. Go to stage $s+1$.

End stage s

Each stage above halts (due to finiteness of search). It is easy to verify that satisfying any requirement does not spoil a higher priority requirement. Thus any requirement can be spoiled (and thus satisfied) only finitely many times. Thus we claim

For all i,j, there exists a s such that exactly one of the following holds

(a) $R_{\langle i,j\rangle}$ remains satisfied for all stages beyond stage s;
(b) In all stages beyond stage s, in step 1 of the construction, (a) holds but (b) does not hold for $\langle i,j\rangle$.

The above claim can be proved as in any standard priority argument proofs. Note that if a requirement $R_{\langle i,j\rangle}$ remains unsatisfied in the limit, then so does $R_{\langle i,j+1\rangle}$.

Now fix i. We will show that either (A) or (B) holds. If for all j, $R_{i,j}$ eventually remains satisfied, then clearly, (A) holds. Thus, if (A) does not hold for i, (i.e. $\mathbf{M}_i(I_L)$ converges), then there exists a requirement $R_{i,j}$ such that $R_{i,j}$ remains unsatisfied in the limit. Thus, by the claim, beyond some stage s, in step 1 of the construction, (a) holds but (b) does not hold for $\langle i,j\rangle$. Note that this implies that L contains only finitely many elements in X_i^0 (elements in X_i^0 are introduced in L only when some $R_{\langle i,j\rangle}$ is satisfied).

Now since L contains only finitely much of X_i^0, and for all but finitely many stages, (a) and (b) in step 1 of the construction do not hold for $\langle i,j\rangle$, we have

> There is an m such that $X_i^m \cap L = \emptyset$ and $\mathbf{M}_i(I_L) = \mathbf{M}(I_{L'})$, for all L' such that $L \subseteq L' \subseteq X_i^m \cup L$.

Thus (B) holds for i. $\square$

7 Concluding Remarks

If one considers the non-parameterized versions of the identification criteria considered in this paper ($\mathbf{Ex} = \mathbf{Ex}^0, \mathbf{Fex} = \mathbf{Fex}^0_*, \mathbf{Oex} = \mathbf{Oex}^0_*, \mathbf{Bc} = \mathbf{Bc}^0$), then as a corollary to the results in this paper we have

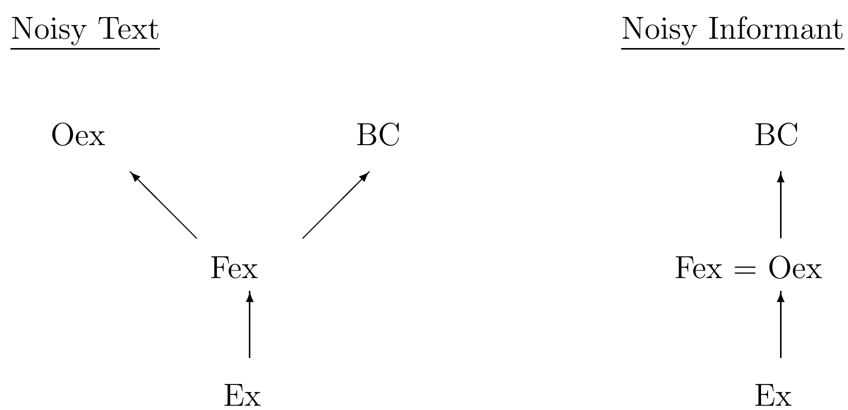
Corollary 36 $\mathbf{NoisyTxtEx} \subset \mathbf{NoisyTxtFex} \subset \mathbf{NoisyTxtBc}$.

$\mathbf{NoisyTxtEx} \subset \mathbf{NoisyTxtFex} \subset \mathbf{NoisyTxtOex}$.

$\mathbf{NoisyTxtOex}$ *and* $\mathbf{NoisyTxtBc}$ *are incomparable by* $\subseteq$.

Corollary 37 $\mathbf{NoisyInfEx} \subset \mathbf{NoisyInfFex} = \mathbf{NoisyInfOex} \subset \mathbf{NoisyTxtBc}$.

The following figure summarizes the above corollaries. In the figure, an arrow from $\mathcal{I}_1$ to $\mathcal{I}_2$ indicates that $\mathcal{I}_1 \subseteq \mathcal{I}_2$. (Also $\mathcal{I}_1 \subseteq \mathcal{I}_2$ iff it follows from the subset relations shown in the figure.)



Note that for learning without noise, we have $\mathbf{TxtEx} \subset \mathbf{TxtFex} = \mathbf{TxtOex} \subset \mathbf{TxtBc}$ and $\mathbf{InfEx} = \mathbf{InfFex} = \mathbf{InfOex} \subset \mathbf{InfBc}$. Thus presence of noise changes the hierarchy structure of common identification criteria.

As we have seen in this paper, the introduction of noise (as defined in this paper and from [33]), in many cases, increases the difficulty of learning, sometimes in interesting ways. It would be good to assuage the difficulty of learning from noisy data, in the future, by finding natural forms of "innate knowledge" or additional information (as, for example, was done for noise free function learning in [16]).

Acknowledgements

Frank Stephan was supported by the Deutsche Forschungsgemeinschaft (DFG) Grant No. Am 60/9-1. We are grateful to Tom Nordahl for his information on schizophrenia. We thank the referees for several helpful comments on the presentation of the paper.

References

[1] D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.

[2] D. Angluin and C. Smith. A survey of inductive inference: Theory and methods. *Computing Surveys*, 15:237–289, 1983.

[3] G. Baliga, J. Case, and S. Jain. Synthesizing enumeration techniques for language learning. In *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, pages 169–180. ACM Press, 1996.

[4] G. Baliga, J. Case, S. Jain, and M. Suraj. Machine learning of higher order programs. *Journal of Symbolic Logic*, 59(2):486–500, 1994.

[5] G. Baliga, S. Jain, and A. Sharma. Learning from multiple sources of inaccurate data. *SIAM Journal of Computing*, 26:961–990, 1997.

[6] J. 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.

[7] J. Bārzdiņš and K. Podnieks. The theory of inductive inference. In *Second Symposium on Mathematical Foundations of Computer Science*, pages 9–15. Math. Inst. of the Slovak Academy of Sciences, 1973.

[8] R. Berwick. *The Acquisition of Syntactic Knowledge*. MIT Press, 1985.

[9] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

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

[11] J. Case. The power of vacillation. In D. Haussler and L. Pitt, editors, *Proceedings of the Workshop on Computational Learning Theory*, pages 133–142. Morgan Kaufmann, 1988.

[12] J. Case. The power of vacillation in language learning. Technical Report LP-96-08, Logic, Philosophy and Linguistics Series of the Institute for Logic, Language and Computation, University of Amsterdam, 1996.

[13] J. Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 1998. To Appear (Preliminary Version Appeared in COLT 88).

[14] J. Case, S. Jain, and A. Sharma. Vacillatory learning of nearly minimal size grammers. *Journal of Computer and System Sciences*, 49:189–207, 1994.

[15] J. Case, S. Jain, and A. Sharma. Complexity issues for vacillatory function identification. *Information and Computation*, 116:174–192, 1995.

[16] J. Case, S. Kaufmann, E. Kinber, and M. Kummer. Learning recursive functions from approximations. In Paul Vitányi, editor, *Second European Conference on Computational Learning Theory*, volume 904 of *Lecture Notes in Artificial Intelligence*, pages 140–153. Springer-Verlag, 1995.

[17] J. Case and C. Lynes. Machine inductive inference and language identification. In M. Nielsen and E. M. Schmidt, editors, *Proceedings of the 9th International Colloquium on Automata, Languages and Programming*, volume 140 of *Lecture Notes in Computer Science*, pages 107–115. Springer-Verlag, 1982.

[18] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.

[19] M. Fulk and S. Jain. Learning in the presence of inaccurate information. *Theoretical Computer Science A*, 161:235–261, 1996.

[20] M. Fulk, S. Jain, and D. Osherson. Open problems in systems that learn. *Journal of Computer and System Sciences*, 49(3):589–604, 1994.

[21] W. Gasarch and M. Pleszkoch. Learning via queries to an oracle. In R. Rivest, D. Haussler, and M. Warmuth, editors, *Proceedings of the Second Annual Workshop on Computational Learning Theory*, pages 214–229. Morgan Kaufmann, 1989.

[22] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[23] S. Kapur and G. Bilardi. Language learning without overgeneralization. In A. Finkel and M. Jantzen, editors, *Proceedings of the Ninth Annual Symposium on Theoretical Aspects of Computer Science*, volume 577 of *Lecture Notes in Computer Science*, pages 245–256. Springer-Verlag, 1992.

[24] S. Kapur, B. Lust, W. Harbert, and G. Martohardjono. Universal grammar and learnability theory: The case of binding domains and the 'subset principle'. In E. Reuland and W Abraham, editors, *Knowledge and Language, Volume I*, pages 185–216. Kluwer, 1993.

[25] S. Lange and T. Zeugmann. Types of monotonic language learning and their characterization. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 377–390. ACM Press, 1992.

[26] S. Lange and T. Zeugmann. Monotonic versus non-monotonic language learning. In *Proceedings of the Second International Workshop on Nonmonotonic and Inductive Logic*, volume 659 of *Lecture Notes in Artificial Intelligence*, pages 254–269. Springer-Verlag, 1993.

[27] Y. Mukouchi. Characterization of finite identification. In K. Jantke, editor, *Analogical and Inductive Inference, Proceedings of the Third International Workshop*, pages 260–267, 1992.

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

[29] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists.* MIT Press, 1986.

[30] R. Salo, L. Robertson, and T. Nordahl. Normal sustained effects of selective attention are absent in schizophrenic patients withdrawn from medication. *Psychiatry Research*, 62:121–130, 1996.

[31] A. Sharma. A note on batch and incremental learnability. *Journal of Computer and System Sciences*, 1998. to appear.

[32] R. Soare. *Recursively Enumerable Sets and Degrees.* Springer-Verlag, 1987.

[33] F. Stephan. Noisy inference and oracles. In K. Jantke, T. Shinohara, and T. Zeugmann, editors, *Algorithmic Learning Theory: Sixth International Workshop (ALT '95)*, volume 997 of *Lecture Notes in Artificial Intelligence*, pages 185–200. Springer-Verlag, 1995.

[34] K Wexler. The subset principle is an intensional principle. In E. Reuland and W Abraham, editors, *Knowledge and Language, Volume I*, pages 217–239. Kluwer, 1993.

[35] T. Zeugmann, S. Lange, and S. Kapur. Characterizations of monotonic and dual monotonic language learning. *Information and Computation*, 120:155–173, 1995.