Optimal Language Learning (Expanded Version)

John Case and Samuel E. Moelius III

Department of Computer & Information Sciences
University of Delaware
103 Smith Hall
Newark, DE 19716
{case,moelius}@cis.udel.edu

July 26, 2008

Abstract. Gold's original paper on inductive inference introduced a notion of an *optimal learner*. Intuitively, a learner identifies a class of objects optimally iff there is no *other* learner that: requires *as little* of each presentation of each object in the class in order to identify that object, and, for *some* presentation of *some* object in the class, requires *less* of that presentation in order to identify that object. Wiehagen considered this notion in the context of *function* learning, and *characterized* an optimal function learner as one that is *class-preserving*, *consistent*, and (in a very strong sense) *non-U-shaped*, with respect to the class of functions learned.

Herein, Gold's notion is considered in the context of *language* learning. Intuitively, a language learner identifies a class of languages optimally iff there is no other learner that: requires as little of each *text* for each language in the class in order to identify that language, and, for some text for some language in the class, requires less of that text in order to identify that language.

Many interesting results concerning optimal language learners are presented. First, it is shown that a characterization analogous to Wiehagen's does *not* hold in this setting. Specifically, optimality is *not* sufficient to guarantee Wiehagen's conditions; though, those conditions *are* sufficient to guarantee optimality. Second, it is shown that the failure of this analog is *not* due to a restriction on algorithmic learning power imposed by non-U-shapedness (in the strong form employed by Wiehagen). That is, non-U-shapedness, even in this strong form, does *not* restrict algorithmic learning power. Finally, for an arbitrary optimal learner $\mathbf{F}$ of a class of languages $\mathcal{L}$, it is shown that $\mathbf{F}$ optimally identifies a subclass $\mathcal{K}$ of $\mathcal{L}$ iff $\mathbf{F}$ is class-preserving with respect to $\mathcal{K}$.

1 Introduction

Gold's original paper on inductive inference introduced a notion of an *optimal learner* [Gol67]. Intuitively, a learner identifies a class of objects optimally iff there is no *other* learner that: requires *as little* of each presentation of each object in the class in order to identify that object, and, for *some* presentation of *some* object in the class, requires *less* of that presentation in order to identify that object.

Gold's notion is perhaps most easily exemplified in the context of function learning, where each object (i.e., function) has one (canonical) presentation, namely, the sequence of all finite initial segments of that function ordered by inclusion (i.e., "$\subseteq$"). We briefly recall the relevant definitions.

Let $\mathbb{N}$ be the set of natural numbers, $\{0, 1, 2, ...\}$. Let $\varphi_0, \varphi_1, ...$ be any acceptable numbering of the partial computable functions from $\mathbb{N}$ to $\mathbb{N}$ [Rog67]. For each function $f : \mathbb{N} \to \mathbb{N}$, and each $n \in \mathbb{N}$, let $f[n]$ denote the initial segment of f whose domain is of size n. A learner $\mathbf{F}$ *identifies* a class of functions $\mathcal{F} \subseteq \mathbb{N} \to \mathbb{N} \overset{\text{def}}{\Leftrightarrow}$ for each $f \in \mathcal{F}$, there exists $n \in \mathbb{N}$ such that $\varphi_{\mathbf{F}(f[n])} = f$ and $(\forall i \geq n)\big[\mathbf{F}(f[i]) = \mathbf{F}(f[n])\big]$.

For each function learner $\mathbf{F}$, and each $f : \mathbb{N} \to \mathbb{N}$, let conv be as follows.

$$\mathrm{conv}(\mathbf{F}, f) = \begin{cases} n, & \text{where } n \text{ is \textit{least} such that } \varphi_{\mathbf{F}(f[n])} = f \\ & \quad \text{and } (\forall i \geq n)\big[\mathbf{F}(f[i]) = \mathbf{F}(f[n])\big]\text{, if} \\ & \quad \text{such an } n \text{ exists;} \\ \text{undefined, otherwise.} \end{cases} \tag{1}$$

Intuitively, $\mathrm{conv}(\mathbf{F}, f)$ indicates *how much* of f must be presented to $\mathbf{F}$ in order for $\mathbf{F}$ to identify f. Thus, if $\mathbf{F}$ and $\mathbf{G}$ are two function learners and $\mathrm{conv}(\mathbf{F}, f) \leq \mathrm{conv}(\mathbf{G}, f)$ (which are both defined), then $\mathbf{F}$ requires *as little* of f as $\mathbf{G}$ requires to identify f.

In the context of function learning, Gold's notion can be made precise as follows. A function learner $\mathbf{F}$ *optimally identifies* a class of functions $\mathcal{F} \overset{\text{def}}{\Leftrightarrow} \mathbf{F}$ identifies $\mathcal{F}$, and, for each function learner $\mathbf{G}$,

$$\begin{aligned} (\forall f \in \mathcal{F})[\mathrm{conv}(\mathbf{G}, f) \leq \mathrm{conv}(\mathbf{F}, f)] \;\Rightarrow \\ (\forall f \in \mathcal{F})[\mathrm{conv}(\mathbf{F}, f) \leq \mathrm{conv}(\mathbf{G}, f)]. \end{aligned} \tag{2}$$

Thus, $\mathbf{F}$ optimally identifies $\mathcal{F}$ iff, for every *other* function learner $\mathbf{G}$, if $\mathbf{G}$ requires *as little* of each $f \in \mathcal{F}$ as $\mathbf{F}$ requires to identify f, then (conversely) $\mathbf{F}$ requires *as little* of each $f \in \mathcal{F}$ as $\mathbf{G}$ requires to identify f. Equivalently: there is *no* other learner $\mathbf{G}$ such that, $\mathbf{G}$ requires as little of each $f \in \mathcal{F}$ as $\mathbf{F}$ requires to identify f, and, for *some* $f \in \mathcal{F}$, $\mathbf{G}$ requires *less* of f than $\mathbf{F}$ requires to identify f.

Wiehagen [Wie91] considered optimal learners in the context of function learning, and characterized them as follows.

Theorem 1 (Wiehagen [Wie91]). Suppose that a function learner $\mathbf{F}$ identifies a class of functions $\mathcal{F}$. Then, $\mathbf{F}$ optimally identifies $\mathcal{F} \Leftrightarrow$ (a) through (c) below.

(a) $\mathbf{F}$ is *class-preserving* [Wie91] with respect to $\mathcal{F}$, i.e.,

$$(\forall f \in \mathcal{F})(\forall n \in \mathbb{N})[\varphi_{\mathbf{F}(f[n])} \in \mathcal{F}]. \tag{3}$$

(b) **F** is *consistent* [Bar77,BB75] with respect to $\mathcal{F}$, i.e.,

$$(\forall f \in \mathcal{F})(\forall n \in \mathbb{N})\big[f[n] \subseteq \varphi_{\mathbf{F}(f[n])}\big]. \tag{4}$$

(c) **F** is *strongly non-U-shaped* [Wie91][1] with respect to $\mathcal{F}$, i.e.,

$$(\forall f \in \mathcal{F})(\forall n \in \mathbb{N})\Big[\varphi_{\mathbf{F}(f[n])} = f \ \Rightarrow\ (\forall i \geq n)\big[\mathbf{F}(f[i]) = \mathbf{F}(f[n])\big]\Big]. \tag{5}$$

Herein, we consider optimal learners in the context of *language* learning, as done in [OSW86, Ch. 8]. In this setting, the situation is slightly more complicated, since, for nearly every object (i.e., language), there is *more than one* presentation (i.e., text). We briefly recall the relevant definitions.

For each $p \in \mathbb{N}$, let $W_p = \{x \in \mathbb{N} \mid \varphi_p(x) \text{ converges}\}$. Thus, $W_0, W_1, ...$ is an enumeration of the recursively enumerable (r.e.) sets [Rog67]. A *language* is a subset of $\mathbb{N}$. A *text* for a language L is a function $T : \mathbb{N} \to (\mathbb{N} \cup \{\#\})$ such that L is *exactly* the non-# elements of the range of T, i.e., $L = \{x \in \mathbb{N} \mid (\exists i)[T(i) = x]\}$. (The symbol '#' is pronounced *pause*.) Clearly, a text uniquely determines a language. Furthermore, if L is a *non*-empty language, then there are *un*countably many texts for L. A language learner **F** *identifies* a class of languages $\mathcal{L} \overset{\text{def}}{\Leftrightarrow}$ for each $L \in \mathcal{L}$, and each text T for L, there exists $n \in \mathbb{N}$ such that $W_{\mathbf{F}(T[n])} = L$ and $(\forall i \geq n)\big[\mathbf{F}(T[i]) = \mathbf{F}(T[n])\big]$.

For each language learner **F**, and each text T, let conv be as follows.

$$\text{conv}(\mathbf{F}, T) = \begin{cases} n, & \text{where } n \text{ is } \textit{least} \text{ such that } W_{\mathbf{F}(T[n])} = L, \\ & \quad (\forall i \geq n)\big[\mathbf{F}(T[i]) = \mathbf{F}(T[n])\big], \text{ and } T \text{ is} \\ & \quad \text{a text for } L, \text{ if such } n \text{ and } L \text{ exist;} \\ \text{undefined, otherwise.} \end{cases} \tag{6}$$

In the context of language learning, Gold's notion can be made precise as follows. A language learner **F** *optimally identifies* a class of languages $\mathcal{L} \overset{\text{def}}{\Leftrightarrow}$ **F** identifies $\mathcal{L}$, and, for each language learner **G**,

$$\begin{aligned} &(\forall L \in \mathcal{L})(\forall T \text{ a text for } L)[\text{conv}(\mathbf{G}, T) \leq \text{conv}(\mathbf{F}, T)] \ \Rightarrow \\ &\qquad (\forall L \in \mathcal{L})(\forall T \text{ a text for } L)[\text{conv}(\mathbf{F}, T) \leq \text{conv}(\mathbf{G}, T)]. \end{aligned} \tag{7}$$

This definition has an interpretation similar to that of the function learning setting. Specifically: **F** optimally identifies $\mathcal{L}$ iff, for every *other* language learner **G**, if **G** requires *as little* of each text for each $L \in \mathcal{L}$ as **F** requires to identify L, then (conversely) **F** requires *as little* of each text for each $L \in \mathcal{L}$ as **G** requires to identify L. Equivalently: there is *no* other learner **G** such that, **G** requires as little of each text for each $L \in \mathcal{L}$ as **F** requires to identify L, and, for *some* text for *some* $L \in \mathcal{L}$, **G** requires *less* of that text than **F** requires to identify L.

---

[1] Wiehagen actually used the term *semantically finite* in place of *strongly non-U-shaped*. However, there is a clear connection between this notion and that of *non-U-shapedness* [CCJS07,BCM+08,CCJS08,CM08]. Our choice of terminology is meant to expose this connection.

Many interesting results concerning optimal language learners are presented. First, we show that a characterization analogous to Wiehagen's (Theorem 1 above) does *not* hold in this setting. Specifically, optimality is *not* sufficient to guarantee Wiehagen's conditions; though, those conditions *are* sufficient to guarantee optimality (Theorem 8 in Section 3). Second, we show that the failure of this analog is *not* due to a restriction on algorithmic learning power imposed by strong non-U-shapedness. That is, strong non-U-shapedness does *not* restrict algorithmic learning power (Theorem 12 in Section 3). Finally, for an arbitrary optimal learner $\mathbf{F}$ of a class of languages $\mathcal{L}$, we show that $\mathbf{F}$ optimally identifies a subclass $\mathcal{K}$ of $\mathcal{L}$ iff $\mathbf{F}$ is class-preserving with respect to $\mathcal{K}$ (Theorem 13 in Section 4).

• • •

A primary motivation for considering optimal language learners is the following. There is no generally accepted notion of *efficient algorithmic* language learning.[2] Optimal learners are, in some sense, *maximally efficient*, in that they use as little of the presentation of an object as possible. Thus, one way to argue that an algorithmic learner is efficient, is to argue that it is *relatively efficient* compared to an optimal learner. We give an example (beginning with (8) below), following some necessary definitions.

Let σ range over finite initial segments of texts. For each text T, and each $n \in \mathbb{N}$, let $T[n]$ denote the initial segment of T of length n. For each σ, let $\mathrm{content}(\sigma) = \{x \in \mathbb{N} \mid (\exists i)[\sigma(i) = x]\}$. Let K be the diagonal halting problem, i.e., $K = \{p \in \mathbb{N} \mid p \in W_p\}$ [Rog67]. For each set $A \subseteq \mathbb{N}$, let $\overline{A} = \mathbb{N} - A$ and $A + 1 = \{x + 1 \mid x \in A\}$.

Let $\mathcal{L}$ be as follows.

$$\mathcal{L} = \big\{\{0\}\big\} \cup \big\{\{p+1\} \mid p \in K\big\} \cup \big\{\{0, p+1\} \mid p \in \overline{K}\big\}. \tag{8}$$

Let f be such that, for each finite $A \subset \mathbb{N}$,

$$W_{f(A)} = A.^3 \tag{9}$$

For each σ, let $\mathbf{M}$ and $\mathbf{F}$ be as follows.

$$\mathbf{M}(\sigma) = f\big(\mathrm{content}(\sigma)\big). \tag{10}$$

$$\mathbf{F}(\sigma) = \begin{cases} f(\{0\}), & \text{if } \mathrm{content}(\sigma) \subseteq \{0\}; \\ f\big(\mathrm{content}(\sigma)\big), & \text{if } \mathrm{content}(\sigma) \cap (K+1) \neq \emptyset; \\ f\big(\{0\} \cup \mathrm{content}(\sigma)\big), & \text{if } \mathrm{content}(\sigma) \cap (\overline{K}+1) \neq \emptyset. \end{cases} \tag{11}$$

The discussion proceeds with the observation of a few facts.

Fact 1. $\mathcal{L}$ is *not* algorithmically, optimally identifiable.[4]

[2] See [Pit89] for a discussion.
[3] Such an f exists by s-m-n [Rog67].
[4] This is shown for a nearly identical class of languages in [OSW86, Proposition 8.2.3A]. The proof of Fact 1 is included here for illustration.

Proof. By way of contradiction, let $\mathbf{M}'$ be an algorithmic learner that optimally identifies $\mathcal{L}$. Then, by (b) $\Rightarrow$ (c) of Theorem 8 (Section 3 below), $\mathbf{M}'$ class-preservingly and consistently identifies $\mathcal{L}$. Note that, for each $p \in \mathbb{N}$, there is exactly *one* $L \in \mathcal{L}$ such that $p+1 \in L$. It follows that

$$\overline{K} = \{p \in \mathbb{N} \mid 0 \in W_{\mathbf{M}'(p+1)}\}. \tag{12}$$

Since the right hand side of (12) is r.e. (by supposition), this is a contradiction. □ (*Fact 1*)

Fact 2. $\mathbf{M}$ algorithmically identifies $\mathcal{L}$, but *not* optimally.

Proof. Clearly, $\mathbf{M}$ identifies $\mathcal{L}$, and $\mathbf{M}$ is algorithmic. Thus, by Fact 1, $\mathbf{M}$ can*not* optimally identify $\mathcal{L}$. □ (*Fact 2*)

Fact 3. $\mathbf{F}$ optimally identifies $\mathcal{L}$, but *not* algorithmically.

Proof. Clearly, $\mathbf{F}$ identifies $\mathcal{L}$. Furthermore, $\mathbf{F}$ is class-preserving, consistent, and strongly non-U-shaped with respect to $\mathcal{L}$. Thus, by (a) $\Rightarrow$ (b) of Theorem 8 (Section 3 below), $\mathbf{F}$ optimally identifies $\mathcal{L}$. Finally, by Fact 1, $\mathbf{F}$ can*not* be algorithmic. □ (*Fact 3*)

Fact 4. On any text T for a language in $\mathcal{L}$, $\mathbf{M}$ requires at most *one more* datapoint than $\mathbf{F}$ requires to converge to a correct hypothesis on T. Formally: for each text T for a language in $\mathcal{L}$,

$$\left|\text{content}\big(T[\text{conv}(\mathbf{M},T)]\big)\right| \leq \left|\text{content}\big(T[\text{conv}(\mathbf{F},T)]\big)\right| + 1. \tag{13}$$

Proof. A straightforward case analysis. □ (*Fact 4*)

Fact 4 gives a sense in which $\mathbf{M}$ is *relatively efficient* compared to $\mathbf{F}$. Generalizations of this notion might allow, e.g., that the size of the set on the right-hand side of (13) be the argument of an arbitrary polynomial.[5]

This notion of relative efficiency seems promising. Of course, this notion is only *meaningful* for those classes of languages for which there exists an optimal learner. Fortunately, however, Proposition 8.2.1A in [OSW86] says that, for *every* identifiable class of languages, there exists an optimal learner.

We hope that the ideas presented here provide for a useful notion of efficient algorithmic language learning.

2 Preliminaries

Computability-theoretic concepts not covered below are treated in [Rog67].

Lowercase math-italic letters (e.g., a, b, c), with or without decorations, range over elements of $\mathbb{N}$, unless stated otherwise. Uppercase math-italic letters (e.g.,

[5] We do not mean to suggest that the content-based measure of (13) represents the *best* possible measure of relative efficiency, just that it is a reasonable one. Alternatives might involve, e.g., *mind-change complexity* [BF74,CS83].

A, B, C), with or without decorations, range over subsets of $\mathbb{N}$, unless stated otherwise. $D_0, D_1, \ldots$ denotes a canonical enumeration of all finite subsets of $\mathbb{N}$. $\mathcal{K}$ and $\mathcal{L}$ range over collections of subsets of $\mathbb{N}$. $\mathcal{E} \stackrel{\text{def}}{=} \{W_p \mid p \in \mathbb{N}\}$. For each A, $|A|$ denotes the cardinality of A. For each finite, *non*-empty A, $\max A$ denotes the maximum element of A. For an arbitrary set $\mathfrak{X}$, $2^{\mathfrak{X}}$ denotes the collection of all subsets of $\mathfrak{X}$.

For each one-argument partial function ψ, and each x, $\psi(x){\downarrow}$ denotes that $\psi(x)$ converges; $\psi(x){\uparrow}$ denotes that $\psi(x)$ diverges.[6] We use $\uparrow$ to denote the value of a divergent computation. Φ denotes a fixed Blum complexity measure for φ. For each i and s, $W_i^s \stackrel{\text{def}}{=} \{x \mid x < s \ \wedge \ \Phi_i(x) \leq s\}$.

$\mathbb{N}_\# \stackrel{\text{def}}{=} \mathbb{N} \cup \{\#\}$. $\textsc{Txt}$ denotes the set of all texts, i.e., functions of type $\mathbb{N} \to \mathbb{N}_\#$. $\textsc{Seq}$ denotes the set of all sequences, i.e., finite initial segments of texts. T, with or without decorations, ranges over elements of $\textsc{Txt}$. Lowercase Greek letters (e.g., ρ, σ, τ), with or without decorations, range over elements of $\textsc{Seq}$, unless stated otherwise. For each L and $\mathcal{L}$, $\textsc{Txt}_L$, $\textsc{Txt}_{\mathcal{L}}$, $\textsc{Seq}_L$, and $\textsc{Seq}_{\mathcal{L}}$ are defined as follows.

$$\textsc{Txt}_L = \{T \mid \text{content}(T) = L\}. \tag{14}$$

$$\textsc{Txt}_{\mathcal{L}} = \{T \mid (\exists L \in \mathcal{L})[\text{content}(T) = L]\}. \tag{15}$$

$$\textsc{Seq}_L = \{\sigma \mid (\exists T \in \textsc{Txt}_L)[\sigma \subset T]\}. \tag{16}$$

$$\textsc{Seq}_{\mathcal{L}} = \{\sigma \mid (\exists T \in \textsc{Txt}_{\mathcal{L}})[\sigma \subset T]\}. \tag{17}$$

In order to disambiguate expressions such as $\textsc{Seq}_\emptyset$, we write $\emptyset$ for the empty *language*, and $\{\}$ for the empty *class of languages*.

For each $A \subseteq \mathbb{N}_\#$, $A^* \stackrel{\text{def}}{=} \{\sigma \mid (\forall i)[\sigma(i){\downarrow} \ \Rightarrow \ \sigma(i) \in A]\}$. Similarly, for each $A \subseteq \mathbb{N}_\#$, $A^\omega \stackrel{\text{def}}{=} \{T \mid (\forall i)[T(i) \in A]\}$. For each $x \in \mathbb{N}_\#$, x^ω denotes the *unique* element of $\{x\}^\omega$. For each $A \subseteq \mathbb{N}_\#$, $A^{\leq\omega} = A^* \cup A^\omega$. In particular, $\mathbb{N}_\#^{\leq\omega} = \textsc{Seq} \cup \textsc{Txt}$.

For each $f \in \mathbb{N}_\#^{\leq\omega}$, $\text{content}(f) \stackrel{\text{def}}{=} \{x \in \mathbb{N} \mid (\exists i)[f(i) = x]\}$. For each $f \in \mathbb{N}_\#^{\leq\omega}$ and n, $f[n]$ denotes the initial segment of f of length n, if it exists; f, otherwise. For each σ, $|\sigma|$ denotes the length of σ (equivalently, $|\{i \mid \sigma(i){\downarrow}\}|$). For each *non*-empty σ, $\sigma^- \stackrel{\text{def}}{=} \sigma[|\sigma| - 1]$. For each σ, and each $f \in \mathbb{N}_\#^{\leq\omega}$, $\sigma \cdot f$ denotes the concatenation of σ and f (in that order). Similarly, for each $A \subseteq \textsc{Seq}$ and $\mathcal{B} \subseteq \mathbb{N}_\#^{\leq\omega}$, $A \cdot \mathcal{B} \stackrel{\text{def}}{=} \{\sigma \cdot f \mid \sigma \in A \ \wedge \ f \in \mathcal{B}\}$. λ denotes the empty sequence (equivalently, the everywhere divergent function).

Following conventions similar to [JORS99], $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, with our without decorations, range over arbitrary (partial) functions of type $\textsc{Seq} \rightharpoonup \mathbb{N}$; whereas, $\mathbf{M}$, with our without decorations, ranges over *algorithmic* (partial) functions of type $\textsc{Seq} \rightharpoonup \mathbb{N}$.

conv was defined in (6) (Section 1). For each $\mathbf{F}$ and T, we write $\text{conv}(\mathbf{F}, T){\downarrow}$ when $\text{conv}(\mathbf{F}, T)$ is defined, and $\text{conv}(\mathbf{F}, T){\uparrow}$ when $\text{conv}(\mathbf{F}, T)$ is undefined. An

[6] For each one-argument partial function ψ, and each x, $\psi(x)$ *converges* iff there exists y such that $\psi(x) = y$; $\psi(x)$ *diverges* iff there is *no* y such that $\psi(x) = y$. If ψ is partial computable, and x is such that $\psi(x)$ diverges, then one can imagine that a program for ψ *goes into an infinite loop* on input x.

expression that the reader will see frequently is

$$T[\text{conv}(\mathbf{F}, T)], \tag{18}$$

which is the *shortest* initial segment of T causing $\mathbf{F}$ to converge to a correct hypothesis for content(T) (if such an initial segment exists).

Proposition 2. $(\forall \mathbf{F}, T, \sigma)\big[T[\text{conv}(\mathbf{F}, T)] \subseteq \sigma \subset T \;\Rightarrow\; W_{\mathbf{F}(\sigma)} = \text{content}(T)\big]$.

Proof of Proposition. Let $\mathbf{F}$, T, and σ be fixed, and suppose that $T[\text{conv}(\mathbf{F}, T)] \subseteq \sigma \subset T$. Let $n = \text{conv}(\mathbf{F}, T)$. By the definition of conv, $W_{\mathbf{F}(T[n])} = \text{content}(T)$ and $(\forall i \geq n)\big[\mathbf{F}(T[i]) = \mathbf{F}(T[n])\big]$. Let i be such that $T[i] = \sigma$. Clearly, $i \geq n$. Thus, $W_{\mathbf{F}(\sigma)} = W_{\mathbf{F}(T[i])} = W_{\mathbf{F}(T[n])} = \text{content}(T)$. □ (*Proposition 2*)

The following are the Gold-style learning criteria of relevance to this paper.

Definition 3. Let $\mathbf{F}$ and $\mathcal{L}$ be fixed.

(a) **(Gold [Gol67])** $\mathbf{F}$ *identifies* $\mathcal{L}$ $\Leftrightarrow$

$$(\forall \sigma \in \text{SEQ}_{\mathcal{L}})[\mathbf{F}(\sigma)\!\downarrow] \;\wedge\; (\forall T \in \text{TXT}_{\mathcal{L}})[\text{conv}(\mathbf{F}, T)\!\downarrow]. \tag{19}$$

(b) **(Wiehagen [Wie91])** $\mathbf{F}$ *class-preservingly identifies* $\mathcal{L}$ $\Leftrightarrow$ $\mathbf{F}$ identifies $\mathcal{L}$ and

$$(\forall \sigma \in \text{SEQ}_{\mathcal{L}})[W_{\mathbf{F}(\sigma)} \in \mathcal{L}]. \tag{20}$$

(c) **(Angluin [Ang80])** $\mathbf{F}$ *consistently identifies* $\mathcal{L}$ $\Leftrightarrow$ $\mathbf{F}$ identifies $\mathcal{L}$ and

$$(\forall \sigma \in \text{SEQ}_{\mathcal{L}})[\text{content}(\sigma) \subseteq W_{\mathbf{F}(\sigma)}]. \tag{21}$$

(d) **(Baliga, *et al.* [BCM$^+$08], Carlucci, *et al.* [CCJS08])** $\mathbf{F}$ *non-U-shapedly identifies* $\mathcal{L}$ $\Leftrightarrow$ $\mathbf{F}$ identifies $\mathcal{L}$ and

$$(\forall L \in \mathcal{L})(\forall \sigma, \tau \in \text{SEQ}_L)\big[[\sigma \subseteq \tau \;\wedge\; W_{\mathbf{F}(\sigma)} \neq W_{\mathbf{F}(\tau)}] \;\Rightarrow\; W_{\mathbf{F}(\sigma)} \neq L\big]. \tag{22}$$

(e) **(Wiehagen [Wie91])** $\mathbf{F}$ *strongly non-U-shapedly*[7] *identifies* $\mathcal{L}$ $\Leftrightarrow$ $\mathbf{F}$ identifies $\mathcal{L}$ and

$$(\forall L \in \mathcal{L})(\forall \sigma, \tau \in \text{SEQ}_L)\big[[\sigma \subseteq \tau \;\wedge\; \mathbf{F}(\sigma) \neq \mathbf{F}(\tau)] \;\Rightarrow\; W_{\mathbf{F}(\sigma)} \neq L\big]. \tag{23}$$

N.B. Some authors (including ourselves, at times) make allowances outside of those of Definition 3(a), such as: (1) allowing $(\exists \sigma \in \text{SEQ}_{\mathcal{L}})[\mathbf{F}(\sigma)\!\uparrow]$, and (2) allowing $\mathbf{F} : \text{SEQ} \rightharpoonup (\mathbb{N} \cup \{?\})$. However, for the purposes of this paper, insisting that $\mathbf{F}$ satisfy the more stringent requirements of Definition 3(a) greatly simplifies the presentation. Moreover, such insistence does not affect the essential content of our results.

Definition 4. For each $\mathcal{L}$, $\mathcal{L}$ is *identifiable* $\Leftrightarrow$ $(\exists \mathbf{F})[\mathbf{F}$ identifies $\mathcal{L}]$.

[7] See footnote 1 above.

N.B. "$\mathcal{L}$ is identifiable" is *not* equivalent to "$\mathcal{L}$ is *algorithmically* identifiable", the latter of which would mean $(\exists \mathbf{M})[\mathbf{M}$ identifies $\mathcal{L}]$.

Definition 5. Let $\mathcal{L}$ be fixed.

(a) Suppose that $\mathbf{F}$ and $\mathbf{G}$ each identify $\mathcal{L}$. Then, (i) and (ii) below.
  (i) $\mathbf{F} \preceq_{\mathcal{L}} \mathbf{G} \;\Leftrightarrow\; (\forall T \in \text{TXT}_{\mathcal{L}})[\text{conv}(\mathbf{F}, T) \leq \text{conv}(\mathbf{G}, T)]$.
  (ii) $\mathbf{F} \prec_{\mathcal{L}} \mathbf{G} \;\Leftrightarrow\; \big[\mathbf{F} \preceq_{\mathcal{L}} \mathbf{G} \;\wedge\; (\exists T \in \text{TXT}_{\mathcal{L}})[\text{conv}(\mathbf{F}, T) < \text{conv}(\mathbf{G}, T)]\big]$.

(b) For each $\mathbf{F}$, $\mathbf{F}$ *optimally identifies* $\mathcal{L} \Leftrightarrow \big[\mathbf{F}$ identifies $\mathcal{L} \;\wedge\; (\forall \mathbf{G})[\mathbf{G} \not\prec_{\mathcal{L}} \mathbf{F}]\big]$.

3 Properties of Optimal Learners

In this section, we show that a characterization analogous to Wiehagen's (Theorem 1 in Section 1) does *not* hold in the language learning setting. Specifically, optimality is *not* sufficient to guarantee Wiehagen's conditions; though, those conditions *are* sufficient to guarantee optimality (Theorem 8 below). We also show that the failure of this analog is *not* due to a restriction on algorithmic learning power imposed by strong non-U-shapedness. That is, strong non-U-shapedness does *not* restrict algorithmic learning power (Theorem 12 below).

The proof of Theorem 8 relies on the following two lemmas.

Lemma 6. Suppose that $\mathbf{F}$ class-preservingly, consistently, and strongly non-U-shapedly identifies $\mathcal{L}$. Then, for each $\sigma \in \text{SEQ}_{\mathcal{L}}$, there exists $L \in \mathcal{L}$ such that

$$\text{content}(\sigma) \subseteq L \;\wedge\; (\forall T \in \text{TXT}_L)[\sigma \subset T \;\Rightarrow\; \text{conv}(\mathbf{F}, T) \leq |\sigma|]. \tag{24}$$

Proof. Let $\mathbf{F}$, $\mathcal{L}$, and σ be as stated. Since $\mathbf{F}$ class-preservingly identifies $\mathcal{L}$, $W_{\mathbf{F}(\sigma)} \in \mathcal{L}$. Let $L = W_{\mathbf{F}(\sigma)}$. Since $\mathbf{F}$ consistently identifies $\mathcal{L}$, $\text{content}(\sigma) \subseteq L$. Since $\mathbf{F}$ strongly non-U-shapedly identifies $\mathcal{L}$, $(\forall \tau \in \text{SEQ}_L)[\sigma \subseteq \tau \;\Rightarrow\; \mathbf{F}(\sigma) = \mathbf{F}(\tau)]$. Clearly, the lemma follows. □ (*Lemma 6*)

Lemma 7. Suppose that $\mathbf{F}$ and $\mathbf{G}$ each identify $\mathcal{L}$. Further suppose that $A \subseteq \text{SEQ}$ is such that

$$(\forall \sigma \notin A)[\mathbf{F}(\sigma) = \mathbf{G}(\sigma)]. \tag{25}$$

Then,

$$(\forall T \in \text{TXT}_{\mathcal{L}})\Big[\text{conv}(\mathbf{F}, T) < \text{conv}(\mathbf{G}, T) \;\Rightarrow\; (\exists \sigma \in A)\big[T[\text{conv}(\mathbf{F}, T)] \subseteq \sigma \subset T\big]\Big]. \tag{26}$$

Proof. Let $\mathbf{F}$, $\mathbf{G}$, $\mathcal{L}$, and A be as stated. Let $T \in \text{TXT}_{\mathcal{L}}$ be such that $\text{conv}(\mathbf{F}, T) < \text{conv}(\mathbf{G}, T)$. Clearly, there exists σ such that $T[\text{conv}(\mathbf{F}, T)] \subseteq \sigma \subset T$ and $\mathbf{F}(\sigma) \neq \mathbf{G}(\sigma)$. By (25), $\sigma \in A$. □ (*Lemma 7*)

The following is the first main result of this section.

Theorem 8. Let $\mathbf{F}$ and $\mathcal{L}$ be fixed. Then,

$$\text{(a)} \underset{\not\Leftarrow}{\Rightarrow} \text{(b)} \underset{\not\Leftarrow}{\Rightarrow} \text{(c)}, \tag{27}$$

where (a) through (c) are as follows.[8]

(a) $\mathbf{F}$ class-preservingly, consistently, and strongly non-U-shapedly identifies $\mathcal{L}$.
(b) $\mathbf{F}$ optimally identifies $\mathcal{L}$.
(c) $\mathbf{F}$ class-preservingly and consistently identifies $\mathcal{L}$.

Proof. Let $\mathbf{F}$ and $\mathcal{L}$ be as stated.

(a) $\Rightarrow$ (b): Suppose that $\mathbf{F}$ class-preservingly, consistently, and strongly non-U-shapedly identifies $\mathcal{L}$. Further suppose, by way of contradiction, that there exist $\mathbf{G}$, $L \in \mathcal{L}$, and $T \in \text{TXT}_L$ such that $\mathbf{G} \preceq_{\mathcal{L}} \mathbf{F}$ and $\text{conv}(\mathbf{G}, T) < \text{conv}(\mathbf{F}, T)$. Let $\sigma = T[\text{conv}(\mathbf{G}, T)]$. By Lemma 6, there exists $L' \in \mathcal{L}$ such that $\text{content}(\sigma) \subseteq L'$ and

$$(\forall T' \in \text{TXT}_{L'})[\sigma \subset T' \;\Rightarrow\; \text{conv}(\mathbf{F}, T') \le |\sigma|]. \tag{28}$$

If $L = L'$, then

$$\begin{aligned} \text{conv}(\mathbf{F}, T) &\le |\sigma| && \{\text{by (28)}\} \\ &= \text{conv}(\mathbf{G}, T) && \{\text{by the choice of } \sigma\} \\ &< \text{conv}(\mathbf{F}, T) && \{\text{by the choice of } T\} \end{aligned}$$

— a contradiction. So, it must be the case that $L \neq L'$. Note that, by the choice of σ,

$$W_{\mathbf{G}(\sigma)} = W_{\mathbf{G}(T[\text{conv}(\mathbf{G},T)])} = L. \tag{29}$$

Let $T' \in \text{TXT}_{L'}$ be *any* such that $\sigma \subset T'$. Then,

$$\begin{aligned} \text{conv}(\mathbf{F}, T') &\le |\sigma| && \{\text{by (28)}\} \\ &< \text{conv}(\mathbf{G}, T') && \{\text{by (29) and } L \neq L'\}. \end{aligned}$$

But this contradicts $\mathbf{G} \preceq_{\mathcal{L}} \mathbf{F}$.

(b) $\Rightarrow$ (c): Suppose that $\mathbf{F}$ optimally identifies $\mathcal{L}$. Further suppose, by way of contradiction, that $\mathbf{F}$ does *not* class-preservingly identify $\mathcal{L}$, *or* that $\mathbf{F}$ does *not* consistently identify $\mathcal{L}$. Then, there exists $\rho \in \text{SEQ}_{\mathcal{L}}$ such that *at least one* of (i) or (ii) below holds.

(i) $W_{\mathbf{F}(\rho)} \notin \mathcal{L}$ (in the case that $\mathbf{F}$ does *not* class-preservingly identify $\mathcal{L}$).
(ii) $\text{content}(\rho) \not\subseteq W_{\mathbf{F}(\rho)}$ (in the case that $\mathbf{F}$ does *not* consistently identify $\mathcal{L}$).

Let $T \in \text{TXT}_{\mathcal{L}}$ be such that $\rho \subset T$.

Claim 8.1. Suppose that $\sigma \in \text{SEQ}_{\mathcal{L}}$ is such that $\rho \subseteq \sigma \subseteq T[\text{conv}(\mathbf{F}, T) - 1]$. Then,

$$(\forall T' \in \text{TXT}_{\mathcal{L}})[\sigma \subset T' \;\Rightarrow\; \text{conv}(\mathbf{F}, T') > |\sigma|]. \tag{30}$$

[8] (a) $\Rightarrow$ (b) of Theorem 8 is an improvement on Proposition 8.2.2A in [OSW86].

Proof of Claim. The proof is by induction on the length of σ. The case when $\rho = \sigma$ is straightforward by the choice of ρ and (i) *or* (ii) above. So, let σ be such that $\rho \subset \sigma \subseteq T[\mathrm{conv}(\mathbf{F}, T) - 1]$, and suppose that

$$(\forall T' \in \mathrm{T{\scriptstyle XT}}_{\mathcal{L}})[\sigma^- \subset T' \;\Rightarrow\; \mathrm{conv}(\mathbf{F}, T') > |\sigma^-|]. \tag{31}$$

Further suppose, by way of contradiction, that, for some $T' \in \mathrm{T{\scriptstyle XT}}_{\mathcal{L}}$, $\sigma \subset T'$ and $\mathrm{conv}(\mathbf{F}, T') \leq |\sigma|$. Then, by (31), $\mathrm{conv}(\mathbf{F}, T') = |\sigma|$. Let $\mathbf{G}$ be such that, for each τ,

$$\mathbf{G}(\tau) = \begin{cases} \mathbf{F}(\sigma), \text{ if } \tau = \sigma^-; \\ \mathbf{F}(\tau), \text{ otherwise.} \end{cases} \tag{32}$$

Clearly, $\mathbf{G}$ identifies $\mathcal{L}$ and $\mathrm{conv}(\mathbf{G}, T') \leq |\sigma^-| < |\sigma|$. Thus, if it can be shown that $\mathbf{G} \preceq_{\mathcal{L}} \mathbf{F}$, then this would (as desired) contradict the fact that $\mathbf{F}$ optimally identifies $\mathcal{L}$. So, suppose that $\mathbf{G} \not\preceq_{\mathcal{L}} \mathbf{F}$. Let $T'' \in \mathrm{T{\scriptstyle XT}}_{\mathcal{L}}$ be such that $\mathrm{conv}(\mathbf{F}, T'') < \mathrm{conv}(\mathbf{G}, T'')$. Then, by Lemma 7 (with $A = \{\sigma^-\}$), $T''[\mathrm{conv}(\mathbf{F}, T'')] \subseteq \sigma^- \subset T''$. But this contradicts (31). □ (*Claim 8.1*)

Let $\mathbf{G}$ be such that, for each σ,

$$\mathbf{G}(\sigma) = \begin{cases} \mathbf{F}\big(T[\mathrm{conv}(\mathbf{F}, T)]\big), \text{ if } \rho \subseteq \sigma \subseteq T[\mathrm{conv}(\mathbf{F}, T) - 1]; \\ \mathbf{F}(\sigma), \qquad \text{otherwise.} \end{cases} \tag{33}$$

Clearly, $\mathbf{G}$ identifies $\mathcal{L}$ and $\mathrm{conv}(\mathbf{G}, T) \leq |\rho| < \mathrm{conv}(\mathbf{F}, T)$. Thus, if it can be shown that $\mathbf{G} \preceq_{\mathcal{L}} \mathbf{F}$, then this would (as desired) contradict the fact that $\mathbf{F}$ optimally identifies $\mathcal{L}$. So, suppose that $\mathbf{G} \not\preceq_{\mathcal{L}} \mathbf{F}$. Let $T' \in \mathrm{T{\scriptstyle XT}}_{\mathcal{L}}$ be such that $\mathrm{conv}(\mathbf{F}, T') < \mathrm{conv}(\mathbf{G}, T')$. Then, by Lemma 7, there exists σ such that $\rho \subseteq \sigma \subseteq T[\mathrm{conv}(\mathbf{F}, T) - 1]$ and $T'[\mathrm{conv}(\mathbf{F}, T')] \subseteq \sigma \subset T'$. But this contradicts Claim 8.1.

(a) $\not\Leftarrow$ (b): Let $\mathcal{L} = \big\{\emptyset, \{0\}\big\}$. Let $p_{\emptyset}$ and $p_{\{0\}}$ be grammars for $\emptyset$ and $\{0\}$, respectively. Let $\mathbf{M}$ be as follows.

$$\begin{aligned} &\mathbf{M}(\lambda) &&= p_{\{0\}}. \\ &\mathbf{M}(0 \cdot \{\#, 0\}^*) &&= p_{\{0\}}. \\ &\mathbf{M}(\# \cdot \{\#\}^*) &&= p_{\emptyset}. \\ &\mathbf{M}(\# \cdot \{\#\}^* \cdot 0 \cdot \{\#, 0\}^*) &&= p_{\{0\}}. \end{aligned} \tag{34}$$

Clearly, $\mathbf{M}$ identifies $\mathcal{L}$. Note that $\mathbf{M}$ is U-shaped, e.g., on the text $\# \cdot 0^\omega$. It remains to show that $\mathbf{M}$ optimally identifies $\mathcal{L}$. By way of contradiction, let $\mathbf{F}$ be such that $\mathbf{F} \prec_{\mathcal{L}} \mathbf{M}$. Note that, for each $T \in \mathrm{T{\scriptstyle XT}}_{\mathcal{L}}$, and each n,

$$\begin{aligned} &\mathrm{conv}(\mathbf{M}, \#^\omega) &&= 1; \\ &\mathrm{conv}(\mathbf{M}, 0 \cdot \{\#, 0\}^\omega) &&= 0; \\ &\mathrm{conv}(\mathbf{M}, \# \cdot \#^n \cdot 0 \cdot \{\#, 0\}^\omega) &&= n + 2. \end{aligned} \tag{35}$$

Let $T \in \mathrm{T{\scriptstyle XT}}_{\mathcal{L}}$ be such that $\mathrm{conv}(\mathbf{F}, T) < \mathrm{conv}(\mathbf{M}, T)$. Clearly, $\mathrm{conv}(\mathbf{M}, T) \geq 1$. Thus, by (35), it suffices to consider the following cases.

Case $[T = \#^\omega]$. Then, by (35), it must be the case that $\mathrm{conv}(\mathbf{F}, T) = 0$ and, thus, $W_{\mathbf{F}(\lambda)} = \emptyset$. It follows that $\mathrm{conv}(\mathbf{F}, 0^\omega) > 0$. But then, by (35), $\mathrm{conv}(\mathbf{F}, 0^\omega) > \mathrm{conv}(\mathbf{M}, 0^\omega)$, which contradicts $\mathbf{F} \preceq_{\mathcal{L}} \mathbf{M}$.

Case $[T \in (\# \cdot \#^n \cdot 0 \cdot \{\#, 0\}^\omega)$, for some $n]$. Then, by (35), it must be the case that $\mathrm{conv}(\mathbf{F}, T) \leq n + 1$. Furthermore, by Proposition 2, $W_{\mathbf{F}(\#^{n+1})} = \{0\}$. It follows that $\mathrm{conv}(\mathbf{F}, \#^\omega) > n + 1 \geq 1$. But then, by (35), $\mathrm{conv}(\mathbf{F}, \#^\omega) > \mathrm{conv}(\mathbf{M}, \#^\omega)$, which contradicts $\mathbf{F} \preceq_{\mathcal{L}} \mathbf{M}$.

(b) $\not\Leftarrow$ (c): Let $\mathcal{L} = \{\emptyset\}$. Let $p_\emptyset$ and $p'_\emptyset$ be any two distinct grammars for $\emptyset$. Let $\mathbf{M}$ be as follows.

$$\begin{aligned} \mathbf{M}(\lambda) &= p_\emptyset. \\ \mathbf{M}(\# \cdot \{\#\}^*) &= p'_\emptyset. \end{aligned} \tag{36}$$

Clearly, $\mathbf{M}$ class-preservingly and consistently identifies $\mathcal{L}$. That $\mathbf{M}$ does *not* optimally identify $\mathcal{L}$ is witnessed by, e.g., $\lambda\sigma.p_\emptyset$. □ (*Theorem 8*)

Remark 9. Requiring that $\mathbf{F}$ class-preservingly, consistently, and *decisively* [BCM+08,CCJS08] identify $\mathcal{L}$ is *not* sufficient to guarantee that $\mathbf{F}$ optimally identifies $\mathcal{L}$, as witnessed by the $\mathbf{M}$ constructed in the proof that (b) $\not\Leftarrow$ (c) of Theorem 8. Requiring that $\mathbf{F}$ class-preservingly, consistently, and *non-U-shapedly* identify $\mathcal{L}$ is similarly insufficient.

Problem 10. Is there an *intuitive* property which is *less* restrictive than strong non-U-shapedness, and which, when combined with class-preservation and consistency, *characterizes* optimality? More formally: does there exist an intuitive predicate $P \subseteq \big((\mathrm{Seq} \rightharpoonup \mathbb{N}) \times \mathcal{E}\big)$ satisfying (a) through (c) below?

(a) For each $\mathbf{F}$ and $\mathcal{L}$, if $\mathbf{F}$ strongly non-U-shapedly identifies $\mathcal{L}$, then $P(\mathbf{F}, \mathcal{L})$.
(b) For each $\mathbf{F}$ and $\mathcal{L}$, if $\mathbf{F}$ class-preservingly and consistently identifies $\mathcal{L}$ *and* $P(\mathbf{F}, \mathcal{L})$, then $\mathbf{F}$ optimally identifies $\mathcal{L}$.
(c) For each $\mathbf{F}$ and $\mathcal{L}$, if $\mathbf{F}$ optimally identifies $\mathcal{L}$, then $P(\mathbf{F}, \mathcal{L})$.

One might wonder whether strong non-U-shapedness restricts *algorithmic* learning power, and, if so, whether this contributes to the failure of the analog of Wiehagen's characterization (Theorem 1 in Section 1). However, as Theorem 12 below states, strong non-U-shapedness does, in fact, *not* restrict algorithmic learning power.

The proof of Theorem 12 relies on the following lemma.

Lemma 11. For each $\mathbf{M}$, there exists a computable function $f : \mathbb{N} \to \mathbb{N}$ such that, for each i, $D_i \subseteq W_{f(i)}$ and $\mathbf{M}$ does *not* identify $W_{f(i)}$.

Proof. Let $\mathbf{M}$ be fixed. There are two cases.

Case [$\mathbf{M}$ does *not* identify $\mathbb{N}$]. Then, for each i,

$$W_{f(i)} = \mathbb{N}. \tag{37}$$

Case [$\mathbf{M}$ ids $\mathbb{N}$]. Let $\mathrm{Seq}_< \subseteq \mathrm{Seq}$ be such that

$$\mathrm{Seq}_< = \big\{\sigma \in \mathrm{Seq} \mid (\forall \sigma' \subseteq \sigma)(\exists n)[\mathrm{content}(\sigma') = \{0, ..., n-1\}]\big\}. \tag{38}$$

Let $<$ be such that, for each $\sigma, \tau \in \textsc{Seq}_<$,

$$\sigma < \tau \;\Leftrightarrow\; [\sigma \subseteq \tau \;\wedge\; \mathrm{content}(\sigma) \subset \mathrm{content}(\tau)]. \tag{39}$$

For each i, φ-program $f(i)$ works by constructing (possibly finitely many) $\sigma^0, \sigma^1, ... \in \textsc{Seq}_<$ as below, so that $W_{f(i)} = \bigcup_{s \in \mathbb{N}} \mathrm{content}(\sigma^s)$.

Stage $s = -1$. Find *any* $\tau \in \textsc{Seq}_<$ such that $D_i \subseteq \mathrm{content}(\tau)$. Set $\sigma^0 = \tau$, and go to stage 0.

Stage $s \in 2\mathbb{N}$. Find $\tau \in \textsc{Seq}_<$ (if any) such that

$$\sigma^s \subseteq \tau \;\wedge\; \mathrm{content}(\tau) \subset W_{\mathbf{M}(\tau)}. \tag{40}$$

If such a τ is found, then set $\sigma^{s+1} = \tau$ and go to stage $s+1$. If *no* such τ is found, then search forever.

Stage $s \in 2\mathbb{N}+1$. Find $\tau \in \textsc{Seq}_<$ (if any) such that

$$\sigma^s < \tau \;\wedge\; (\exists \tau')[\sigma^s \subseteq \tau' \subseteq \tau \;\wedge\; \mathbf{M}(\sigma^s) \neq \mathbf{M}(\tau')]. \tag{41}$$

If such a τ is found, then set $\sigma^{s+1} = \tau$ and go to stage $s+1$. If *no* such τ is found, then search forever.

If some stage of the form $2j$ is *not* exited, then, clearly, $\mathbf{M}$ does *not* identify $\mathbb{N}$ from any text beginning with σ^{2j}. So, it must be the case that every stage of the form $2j$ *is* exited.

On the other hand, if *every* stage is exited (including stages of the form $2j+1$), then, clearly, $\lim_{s \to \infty} \sigma^s$ is a text for $\mathbb{N}$ on which $\mathbf{M}$ *never* reaches a final conjecture. So, it must be the case that some stage of the form $2j+1$ is *not* exited.

Let $\sigma = \sigma^{2j+1}$. Clearly, $W_{f(i)} = \mathrm{content}(\sigma) \supseteq D_i$. Furthermore, (a) and (b) below.

(a) $\mathrm{content}(\sigma) \subset W_{\mathbf{M}(\sigma)}$ (by (40)).
(b) $(\forall \tau \in \textsc{Seq}_<)\big[\sigma < \tau \;\Rightarrow\; (\forall \tau')[\sigma \subseteq \tau' \subseteq \tau \;\Rightarrow\; \mathbf{M}(\sigma) = \mathbf{M}(\tau')]\big]$ (by (41)).

Finally, by way of contradiction, suppose that $\mathbf{M}$ ids $\mathrm{content}(\sigma)$. By (a) above, $W_{\mathbf{M}(\sigma)} \neq \mathrm{content}(\sigma)$. Thus, there must exist an n such that $\mathbf{M}(\sigma) \neq \mathbf{M}(\sigma \cdot \#^n)$. Let $\tau = \sigma \cdot \#^n \cdot \min\{x \in \mathbb{N} \mid x \notin \mathrm{content}(\sigma)\}$. Clearly, $\tau \in \textsc{Seq}_<$ and $\sigma < \tau$. But then, by (b) above, $\mathbf{M}(\sigma) = \mathbf{M}(\sigma \cdot \#^n)$ — a contradiction. □ (*Lemma 11*)

The following is the second main result of this section.

Theorem 12. For each $\mathcal{L}$, $(\exists \mathbf{M})[\mathbf{M}$ identifies $\mathcal{L}] \Leftrightarrow (\exists \mathbf{M})[\mathbf{M}$ strongly non-U-shapedly identifies $\mathcal{L}]$.[9]

[9] Theorem 20 in [BCM+08] says: for each $\mathcal{L}$, $(\exists \mathbf{M})[\mathbf{M}$ identifies $\mathcal{L}] \Leftrightarrow (\exists \mathbf{M})[\mathbf{M}$ non-U-shapedly identifies $\mathcal{L}]$. Theorem 12 above is an improvement on this result.

Proof. Clearly, $(\exists \mathbf{M})[\mathbf{M}$ strongly non-U-shapedly identifies $\mathcal{L}] \Rightarrow (\exists \mathbf{M})[\mathbf{M}$ identifies $\mathcal{L}]$. Thus, it suffices to show the converse. Let $\mathcal{L}$ be fixed, and let $\mathbf{M}$ be such that $\mathbf{M}$ identifies $\mathcal{L}$. Without loss of generality, suppose that $\mathbf{M}$ is prudent and total ([Ful90, Theorem 15] and [JORS99, proof of Proposition 4.15]).[10] A machine $\mathbf{M}'$ is constructed such that $\mathbf{M}'$ strongly non-U-shapedly identifies $\mathcal{L}$. (Roughly speaking, certain conjectures of $\mathbf{M}'$ will *self-destruct* when conditions are met that would cause $\mathbf{M}'$ to make a mind-change. See (53).)

Let $f : \mathbb{N} \to \mathbb{N}$ be as in Lemma 11 for $\mathbf{M}$. For each ρ, σ, s, and τ, let $P(\rho, \sigma, s, \tau) \Leftrightarrow$ there exists α satisfying (a) through (c) below.

(a) $|\alpha| \leq s$.
(b) $\text{content}(\alpha) \subseteq W^s_{\mathbf{M}(\rho)}$.
(c) $(\exists \alpha' \subseteq \alpha \cdot \tau)[\mathbf{M}(\sigma \cdot \alpha') \neq \mathbf{M}(\sigma)]$.

For each ρ, σ, s, and τ, let $P^*(\rho, \sigma, s, \tau) \Leftrightarrow$ there exists α satisfying (a) through (c) just above, and (d) just below.

(d) $(\exists s' \leq s)\big[\emptyset \neq \big(W^{s'}_{\mathbf{M}(\rho)} - \text{content}(\sigma)\big) \subseteq \text{content}(\alpha)\big]$.

Clearly, P and P^* are computable predicates. Intuitively, P helps to determine when a segment of text may be extended in a way that causes $\mathbf{M}$ to make a mind-change. In this sense, the arguments of P play the following roles.

- ρ is used to determine a conjecture of $\mathbf{M}$ (i.e., $\mathbf{M}(\rho)$); the elements used to extend the segment of text σ are drawn from the conjectured language.
- σ is the segment of text to be extended.
- s is used to *bound* the process of searching for an extension, and helps to keep P computable.
- τ is a segment of text that should appear at the end of the extension.

Let $g : (\textsc{Seq} \times \mathbb{N}) \to \textsc{Seq}$ be such that, for each ρ and s,

$$g(\rho, s) = \begin{cases} \lambda, & \text{if } s = 0; \\ \sigma \cdot \alpha, & \text{if } s \neq 0 \ \wedge \ P^*(\rho, \sigma, s, \lambda), \text{ where } \sigma = g(\rho, s-1) \\ & \quad \text{and } \alpha \text{ is } \textit{any} \text{ as in (a) through (d) above for} \\ & \quad \rho,\ \sigma,\ s, \text{ and } \tau\ (= \lambda); \\ \sigma, & \text{otherwise, where } \sigma = g(\rho, s-1). \end{cases} \tag{42}$$

Clearly, g is computable.

Claim 12.1. For each ρ, (i) and (ii) below.

(i) $(\forall s_0, s_1)[s_0 < s_1 \ \Rightarrow \ g(\rho, s_0) \subseteq g(\rho, s_1)]$.
(ii) $(\forall s)\big[\text{content}\big(g(\rho, s)\big) \subseteq W^s_{\mathbf{M}(\rho)}\big]$.

[10] Strictly speaking, the technique employed in the proof of [JORS99, Proposition 4.15] takes an arbitrary learner into a total, *not necessarily prudent* learner. However, it is easily seen that if this technique is applied to an *already* prudent learner, then the resulting learner is prudent, as well.

Proof of Claim. Straightforward. □ (*Claim 12.1*)

Claim 12.2. For each ρ, if $g(\rho, \cdot)$ changes infinitely often, then (i) and (ii) below.

(i) $\big(\lim_{s\to\infty} g(\rho, s)\big) \in \textsc{Txt}\big(W_{\mathbf{M}(\rho)}\big)$.
(ii) $\mathrm{conv}\big(\mathbf{M}, \lim_{s\to\infty} g(\rho, s)\big)\uparrow$.

Proof of (i). Let ρ be fixed, and suppose that $g(\rho, \cdot)$ changes infinitely often. Let $T = \lim_{s\to\infty} g(\rho, s)$. By Claim 12.1(i), T is well-defined, and, since $g(\rho, \cdot)$ changes infinitely often, T is total. By Claim 12.1(ii), $\mathrm{content}(T) \subseteq W_{\mathbf{M}(\rho)}$. Thus, to show that $T \in \textsc{Txt}\big(W_{\mathbf{M}(\rho)}\big)$, it suffices to show that $W_{\mathbf{M}(\rho)} \subseteq \mathrm{content}(T)$. By way of contradiction, let s_0 be *least* such that

$$(W^{s_0+1}_{\mathbf{M}(\rho)} - W^{s_0}_{\mathbf{M}(\rho)}) \not\subseteq \mathrm{content}(T). \tag{43}$$

Let $s_1 \geq s_0$ be such that

$$W^{s_0}_{\mathbf{M}(\rho)} \subseteq \mathrm{content}\big(g(\rho, s_1)\big). \tag{44}$$

Since $g(\rho, \cdot)$ changes infinitely often, there exists a *least* $s_2 > s_1$ such that $g(\rho, s_2) \neq g(\rho, s_1)$. Clearly, $g(\rho, s_2)$ is of the form $g(\rho, s_1) \cdot \alpha$ for some α satisfying

$$(\exists s' \leq s_2)\Big[\emptyset \neq \Big(W^{s'}_{\mathbf{M}(\rho)} - \mathrm{content}\big(g(\rho, s_1)\big)\Big) \subseteq \mathrm{content}(\alpha)\Big]. \tag{45}$$

Since $W^{s_0}_{\mathbf{M}(\rho)} \subseteq \mathrm{content}\big(g(\rho, s_1)\big)$, it must be the case that $s' > s_0$. Thus,

$$\begin{aligned} &\Big(W^{s_0+1}_{\mathbf{M}(\rho)} - \mathrm{content}\big(g(\rho, s_1)\big)\Big) \\ \subseteq\ &\Big(W^{s'}_{\mathbf{M}(\rho)} - \mathrm{content}\big(g(\rho, s_1)\big)\Big) && \{\text{because } s_0 + 1 \leq s'\} \\ \subseteq\ &\mathrm{content}(\alpha) && \{\text{by (45)}\}. \end{aligned} \tag{46}$$

But then

$$W^{s_0+1}_{\mathbf{M}(\rho)} \subseteq \mathrm{content}\big(g(\rho, s_1) \cdot \alpha\big) = \mathrm{content}\big(g(\rho, s_2)\big) \subseteq \mathrm{content}(T), \tag{47}$$

which contradicts (43).

Proof of (ii). To show that $\mathrm{conv}(\mathbf{M}, T)\uparrow$, by way of contradiction, let s_0 be such that $\mathrm{conv}(\mathbf{M}, T) \leq |g(\rho, s_0)|$. Since $g(\rho, \cdot)$ changes infinitely often, there exists a *least* $s_1 > s_0$ such that $g(\rho, s_1) \neq g(\rho, s_0)$. Clearly, $g(\rho, s_1)$ is of the form $g(\rho, s_0) \cdot \alpha$ for some α satisfying

$$(\exists \alpha' \subseteq \alpha)\big[\mathbf{M}\big(g(\rho, s_0) \cdot \alpha'\big) \neq \mathbf{M}\big(g(\rho, s_0)\big)\big]. \tag{48}$$

But since $g(\rho, s_0) \cdot \alpha = g(\rho, s_1) \subset T$, $\mathrm{conv}(\mathbf{M}, T) > |g(\rho, s_0)|$ — a contradiction. □ (*Claim 12.2*)

Claim 12.3. For each ρ, there exists s such that

$$(\forall s' > s)[g(\rho, s') = g(\rho, s)]. \tag{49}$$

Proof of Claim. Follows from Claim 12.2 and the fact the **M** is prudent. □ (*Claim 12.3*)

Let $g_{\text{lim}} : \text{SEQ} \to \text{SEQ}$ be such that, for each ρ,

$$g_{\text{lim}}(\rho) = \lim_{s \to \infty} g(\rho, s). \tag{50}$$

By Claim 12.3, g_{lim} is well-defined.

Claim 12.4. Suppose that ρ, s, and τ are such that

(i) $\text{content}(\tau) \subseteq W_{\mathbf{M}(\rho)}$.
(ii) $\text{content}(\tau) \not\subseteq \text{content}\big(g(\rho, s)\big)$.
(iii) $(\exists \alpha' \subseteq \tau)\big[\mathbf{M}\big(g(\rho, s) \cdot \alpha'\big) \neq \mathbf{M}\big(g(\rho, s)\big)\big]$.

Then, there exists $s'' > s$ such that $g(\rho, s'') \neq g(\rho, s)$.

Proof of Claim. Let ρ, s, and τ be as stated. Since $\text{content}(\tau) \subseteq W_{\mathbf{M}(\rho)}$, there exists $s' \geq s$ such that

$$\text{content}(\tau) \subseteq W^{s'}_{\mathbf{M}(\rho)}. \tag{51}$$

Let s'' be such that

$$s'' = \max\{s', |\tau| + |W^{s'}_{\mathbf{M}(\rho)}|\}. \tag{52}$$

Let α be τ followed by the elements of $W^{s'}_{\mathbf{M}(\rho)}$ in any order. Clearly, α satisfies (a) through (d) in the definition of P^* for $P^*(\rho, g(\rho, s), s'', \lambda)$. Thus, it follows from Claim 12.1(i) that $g(\rho, s'') \neq g(\rho, s)$. □ (*Claim 12.4*)

By 1-1 s-m-n [Rog67], there exists a 1-1, computable function $h : (\text{SEQ} \times \mathbb{N}) \to \mathbb{N}$ such that, for each ρ and s,

$$W_{h(\rho,s)} = \begin{cases} W_{\mathbf{M}(\rho)}, \text{ if } (\forall s' > s)[g(\rho, s') = g(\rho, s)], \\ W_{f(i)}, \quad \text{otherwise, where } D_i = W^{s'}_{\mathbf{M}(\rho)} \text{ for the \textit{least}} \\ \qquad\qquad s' > s \text{ such that } g(\rho, s') \neq g(\rho, s). \end{cases} \tag{53}$$

For each ρ, s, and τ, let $Q(\rho, s, \tau) \Leftrightarrow$ (a) through (c) below.

(a) $g(\rho, s) = g(\rho, s + |\tau|)$.
(b) $\text{content}\big(g(\rho, s)\big) \subseteq \text{content}(\tau)$.
(c) $P\big(\rho, g(\rho, s), |\tau|, \tau\big) \;\Rightarrow\; \text{content}(\tau) \subseteq \text{content}\big(g(\rho, s)\big)$.

Clearly, Q is a computable predicate. Many of the conjectures of $\mathbf{M}'$ are of the form $h(\rho, s)$, for some ρ and s. For such conjectures, Q helps to determine appropriate values of ρ and s. Q also helps to determine when such conjectures should be abandoned.

For each τ, let

$$\mathbf{M}'(\tau) = \begin{cases} \mathbf{M}'(\tau^-), \text{ if } (*)\big[\tau \neq \lambda \\ \qquad\qquad \wedge\ (\exists \rho, s)[\mathbf{M}(\tau^-) = h(\rho, s)\ \wedge\ Q(\rho, s, \tau)]\big]; \\ h(\rho, |\tau|), \text{ where } \rho \subseteq \tau \text{ is \textit{shortest} such that } Q(\rho, |\tau|, \tau), \\ \qquad\qquad \text{if } \neg(*) \text{ and such a } \rho \text{ exists}; \\ f(0), \qquad \text{otherwise.} \end{cases} \tag{54}$$

Clearly, $\mathbf{M}'$ is computable. Let $L \in \mathcal{L}$ and $T \in \textsc{Txt}_L$ be fixed. That $\mathbf{M}'$ is strongly *non*-U-shaped on T follows from Claim 12.5 below. That $\mathbf{M}'$ identifies L from T follows from Claims 12.5 and 12.11 below.

Claim 12.5. For each i, if $\mathbf{M}'(T[i]) \neq \mathbf{M}'(T[i+1])$, then $W_{\mathbf{M}'(T[i])} \neq L$.

Proof of Claim. By way of contradiction, let i be such that

$$\mathbf{M}'(T[i]) \neq \mathbf{M}'(T[i+1]) \ \wedge \ W_{\mathbf{M}'(T[i])} = L. \tag{55}$$

Then, clearly, there exist $\rho \subseteq T[i]$ and $s \leq i$ satisfying (a) through (c) below.

(a) $\mathbf{M}'(T[i]) = h(\rho, s)$.
(b) $Q(\rho, s, T[i])$.
(c) $\neg Q(\rho, s, T[i+1])$.

Note that

$$\begin{aligned} L &= W_{\mathbf{M}'(T[i])} \ \{\text{by (55)}\} \\ &= W_{h(\rho,s)} \quad \{\text{by (a) above}\}. \end{aligned}$$

Clearly then, by (53),

$$W_{\mathbf{M}(\rho)} = W_{h(\rho,s)} \ (= L). \tag{56}$$

Consider the following cases (based on (c) above), each of which leads to a contradiction.

Case $[g(\rho, s) \neq g(\rho, s+i+1)]$. Then, clearly, by (53), $L \neq W_{h(\rho,s)}$ $(= W_{\mathbf{M}'(T[i])})$ — a contradiction.

Case $\big[\text{content}\big(g(\rho, s)\big) \not\subseteq \text{content}(T[i+1])\big]$. From $Q(\rho, s, T[i])$, it follows that

$$\text{content}\big(g(\rho, s)\big) \subseteq \text{content}(T[i]) \subseteq \text{content}(T[i+1]). \tag{57}$$

Thus, assuming this case leads to a contradiction.

Case $\big[P\big(\rho, g(\rho, s), i+1, T[i+1]\big) \wedge \text{content}(T[i+1]) \not\subseteq \text{content}\big(g(\rho, s)\big)\big]$. Let α be as in (a) through (c) in the definition of P for $P\big(\rho, g(\rho, s), i+1, T[i+1]\big)$. Then, in particular,

$$\begin{aligned} &\text{content}(\alpha) \subseteq W^{i+1}_{\mathbf{M}(\rho)} \\ &\quad \wedge \ (\exists \alpha' \subseteq \alpha \cdot T[i+1])\big[\mathbf{M}\big(g(\rho, s) \cdot \alpha'\big) \neq \mathbf{M}\big(g(\rho, s)\big)\big]. \end{aligned} \tag{58}$$

Clearly, if one lets $\tau = \alpha \cdot T[i+1]$, then ρ, s, and τ satisfy the conditions of Claim 12.4. Thus, there exists $s'' > s$ such that $g(\rho, s'') \neq g(\rho, s)$. Clearly, then, by (53), $L \neq W_{h(\rho,s)}$ $(= W_{\mathbf{M}'(T[i])})$ — a contradiction. □ (*Claim 12.5*)

Claim 12.6. Suppose that ρ, s, and k are such that $(\forall k' \geq k)[Q(\rho, s, T[k'])]$. Then, $W_{h(\rho,s)} = W_{\mathbf{M}(\rho)}$.

Proof of Claim. Straightforward. □ (*Claim 12.6*)

Claim 12.7. For each ρ, if $L \subset W_{\mathbf{M}(\rho)}$, then $\text{content}\big(g_{\lim}(\rho)\big) \not\subseteq L$.

Proof of Claim. By way of contradiction, let ρ be such that

$$L \subset W_{\mathbf{M}(\rho)} \;\wedge\; \mathrm{content}\big(g_{\lim}(\rho)\big) \subseteq L. \tag{59}$$

Since $\mathbf{M}$ ids L and $\mathrm{content}\big(g_{\lim}(\rho)\big) \subseteq L$, there must exist an α_0 (possibly empty) such that

$$\mathrm{content}(\alpha_0) \subseteq L \;\wedge\; W_{\mathbf{M}(g_{\lim}(\rho)\cdot\alpha_0)} = L. \tag{60}$$

Furthermore, since $\mathbf{M}$ is prudent, $\mathbf{M}$ ids $W_{\mathbf{M}(\rho)}$. Thus, since $\mathrm{content}\big(g_{\lim}(\rho)\cdot\alpha_0\big) \subseteq L \subset W_{\mathbf{M}(\rho)}$, there must exist an α_1 such that

$$\mathrm{content}(\alpha_1) \subseteq W_{\mathbf{M}(\rho)} \;\wedge\; \mathrm{content}(\alpha_1) \not\subseteq L \;\wedge\; W_{\mathbf{M}(g_{\lim}(\rho)\cdot\alpha_0\cdot\alpha_1)} = W_{\mathbf{M}(\rho)}. \tag{61}$$

Let s be such that $g(\rho,s) = g_{\lim}(\rho)$, and let $\tau = \alpha_0\cdot\alpha_1$. Then, clearly, ρ, s, and τ satisfy the conditions of Claim 12.4. Thus, there exists $s'' > s$ such that $g(\rho,s'') \neq g(\rho,s)$ — a contradiction. □ (*Claim 12.7*)

Claim 12.8. For each ρ, if $L \not\subseteq W_{\mathbf{M}(\rho)}$, then (i) and (ii) below.

(i) $\mathrm{content}\big(g_{\lim}(\rho)\big) \not\subseteq L \;\vee\; (\exists k)\big[P\big(\rho, g_{\lim}(\rho), k, T[k]\big)\big]$.
(ii) $(\exists k)(\forall s)\big[\mathrm{content}(T[k]) \not\subseteq \mathrm{content}\big(g(\rho,s)\big)\big]$.

Proof of (i). Suppose that $\mathrm{content}\big(g_{\lim}(\rho)\big) \subseteq L \not\subseteq W_{\mathbf{M}(\rho)}$. Then, by Claim 12.1(ii),

$$\mathrm{content}\big(g_{\lim}(\rho)\big) \subseteq L \cap W_{\mathbf{M}(\rho)}. \tag{62}$$

Consider the following cases.

Case $[W_{\mathbf{M}(g_{\lim}(\rho))} = L]$. Consider the elements of the following set.

$$\begin{aligned} A = \{\alpha \in \textsc{Seq} \mid\ & \mathrm{content}(\alpha) \subseteq W_{\mathbf{M}(\rho)} \\ & \wedge (\exists \alpha' \subseteq \alpha)[W_{\mathbf{M}(g_{\lim}(\rho)\cdot\alpha')} = W_{\mathbf{M}(\rho)}]\}. \end{aligned} \tag{63}$$

Since $\mathbf{M}$ is prudent, $\mathbf{M}$ ids $W_{\mathbf{M}(\rho)}$. Thus, since $\mathrm{content}\big(g_{\lim}(\rho)\big) \subseteq L \cap W_{\mathbf{M}(\rho)} \subseteq W_{\mathbf{M}(\rho)}$, A is *non*-empty. Furthermore, for each $\alpha \in A$, there clearly exists an s such that α satisfies (a) through (c) in the definition of P for $P(\rho, g_{\lim}(\rho), s, \lambda)$. Thus, since $g(\rho,\cdot)$ does not grow beyond $g_{\lim}(\rho)$, it must be the case that each $\alpha \in A$ does not satisfy (d) in the definition of P^*, i.e.,

$$\begin{aligned} (\forall \alpha \in A)(\forall s')\Big[&\Big(W^{s'}_{\mathbf{M}(\rho)} - \mathrm{content}\big(g_{\lim}(\rho)\big)\Big) \neq \emptyset \;\Rightarrow \\ &\Big(W^{s'}_{\mathbf{M}(\rho)} - \mathrm{content}\big(g_{\lim}(\rho)\big)\Big) \not\subseteq \mathrm{content}(\alpha)\Big]. \end{aligned} \tag{64}$$

Subclaim. $W_{\mathbf{M}(\rho)} \subseteq \mathrm{content}\big(g_{\lim}(\rho)\big)$.

Proof of Subclaim. By way of contradiction, suppose otherwise. Let s' be such that $W^{s'}_{\mathbf{M}(\rho)} \not\subseteq \mathrm{content}\big(g_{\lim}(\rho)\big)$. Since A is *non*-empty, there exists $\alpha \in A$. Let $\bar\alpha$ be α followed by the elements of $W^{s'}_{\mathbf{M}(\rho)}$ in any order. Clearly,

$$\bar\alpha \in A \;\wedge\; \Big(W^{s'}_{\mathbf{M}(\rho)} - \mathrm{content}\big(g_{\lim}(\rho)\big)\Big) \neq \emptyset \;\wedge\; W^{s'}_{\mathbf{M}(\rho)} \subseteq \mathrm{content}(\bar\alpha). \tag{65}$$

But this contradicts (64). □ (*Subclaim*)

Thus, for each $\alpha \in A$,

$$\begin{aligned} \text{content}(\alpha) &\subseteq W_{\mathbf{M}(\rho)} && \{\text{by the def. of } A\} \\ &\subseteq \text{content}\big(g_{\lim}(\rho)\big) && \{\text{by the subclaim}\} \\ &\subseteq L \cap W_{\mathbf{M}(\rho)} && \{\text{by (62)}\} \\ &\subseteq L && \{\text{immediate}\}. \end{aligned}$$

Furthermore, for each $\alpha \in A$, there must exist a k such that

$$|\alpha| \leq k \ \wedge\ \text{content}(\alpha) \subseteq W^k_{\mathbf{M}(\rho)} \ \wedge\ W_{\mathbf{M}(g_{\lim}(\rho)\cdot\alpha\cdot T[k])} = L. \tag{66}$$

Clearly, then, $P\big(\rho, g_{\lim}(\rho), k, T[k]\big)$.

Case $[W_{\mathbf{M}(g_{\lim}(\rho))} \neq L]$. Since $\mathbf{M}$ ids L and $\text{content}\big(g_{\lim}(\rho)\big) \subseteq L \cap W_{\mathbf{M}(\rho)} \subseteq L$, there must exist a k such that

$$W_{\mathbf{M}(g_{\lim}(\rho)\cdot T[k])} = L. \tag{67}$$

Clearly, then, $P\big(\rho, g_{\lim}(\rho), k, T[k]\big)$ (with $\alpha = \lambda$).

Proof of (ii). Let k be such that $\text{content}(T[k]) \not\subseteq W_{\mathbf{M}(\rho)}$. By Claim 12.1(ii), for each s, $\text{content}\big(g(\rho, s)\big) \subseteq W_{\mathbf{M}(\rho)}$. Clearly, part (ii) of the present claim follows. □ (*Claim 12.8*)

Claim 12.9. For each i, if $W_{\mathbf{M}(T[i])} \neq L$, then (i) and (ii) below.

(i) $(\exists s, k)(\forall s' \geq s, k' \geq k)\big[\neg Q(T[i], s', T[k'])\big]$.
(ii) $(\forall s)(\exists k)(\forall k' \geq k)\big[\neg Q(T[i], s, T[k'])\big]$.

Proof of (i). Let i be such that $W_{\mathbf{M}(T[i])} \neq L$. Let s be such that $g(T[i], s) = g_{\lim}(T[i])$. By Claim 12.8(i), it suffices to consider the following cases.

Case $[L \subset W_{\mathbf{M}(T[i])}]$. Then, by Claim 12.7, $\text{content}\big(g(T[i], s)\big) \not\subseteq L$. Clearly, then, for each $s' \geq s$ and *all* k, $\text{content}\big(g(T[i], s')\big) \not\subseteq \text{content}(T[k])$, and, thus, $\neg Q(T[i], s', T[k])$.

Case $\big[L \not\subseteq W_{\mathbf{M}(T[i])} \ \wedge\ \text{content}\big(g_{\lim}(T[i])\big) \not\subseteq L\big]$. Similar to the previous case.

Case $\Big[L \not\subseteq W_{\mathbf{M}(T[i])} \ \wedge\ (\exists k_0)\big[P\big(T[i], g_{\lim}(T[i]), k_0, T[k_0]\big)\big]\Big]$. By Claim 12.8(ii), there exists k_1 such that $(\forall s)\big[\text{content}(T[k_1]) \not\subseteq \text{content}\big(g(T[i], s)\big)\big]$. Let $k = \max\{k_0, k_1\}$. Clearly, for each $s' \geq s$ and $k' \geq k$,

$$P\big(T[i], g(T[i], s'), k', T[k']\big) \ \wedge\ \text{content}(T[k']) \not\subseteq \text{content}\big(g(T[i], s')\big), \tag{68}$$

and, thus, $\neg Q(T[i], s', T[k'])$.

Proof of (ii). Let i be such that $W_{\mathbf{M}(T[i])} \neq L$, and let s be fixed. The proof is straightforward for the case when $g(T[i], s) \neq g_{\lim}(T[i])$. For the case when

$g(T[i], s) = g_{\lim}(T[i])$, the proof of the present part follows from the proof of part (i). □ (*Claim 12.9*)

Claim 12.10. For each i, if $W_{\mathbf{M}(T[i])} = L$, then

$$(\exists s, k)(\forall s' \geq s, k' \geq k)\big[Q(T[i], s', T[k'])\big]. \tag{69}$$

Proof of Claim. Let i be such that $W_{\mathbf{M}(T[i])} = L$. Let s be such that $g(T[i], s) = g_{\lim}(T[i])$. By Claim 12.1(ii), $\mathrm{content}\big(g_{\lim}(T[i])\big) \subseteq W_{\mathbf{M}(T[i])} = L$. So, let k be such that $\mathrm{content}\big(g_{\lim}(T[i])\big) \subseteq \mathrm{content}(T[k])$. To complete the proof of the claim, it suffices to show that, for each $k' \geq k$,

$$P\big(T[i], g_{\lim}(T[i]), k', T[k']\big) \ \Rightarrow\ \mathrm{content}(T[k']) \subseteq \mathrm{content}\big(g_{\lim}(T[i])\big). \tag{70}$$

So, by way of contradiction, suppose otherwise, i.e., there exists $k' \geq k$ such that

$$P\big(T[i], g_{\lim}(T[i]), k', T[k']\big) \ \wedge\ \mathrm{content}(T[k']) \not\subseteq \mathrm{content}\big(g_{\lim}(T[i])\big). \tag{71}$$

Let α be any as in (a) through (c) in the definition of P for $P\big(T[i], g_{\lim}(T[i]), k', T[k']\big)$. Then, in particular,

$$\begin{aligned} &\mathrm{content}(\alpha) \subseteq W^{k'}_{\mathbf{M}(T[i])} \\ &\quad \wedge\ (\exists \alpha' \subseteq \alpha \cdot T[k'])\big[\mathbf{M}\big(g_{\lim}(T[i]) \cdot \alpha'\big) \neq \mathbf{M}\big(g_{\lim}(T[i])\big)\big]. \end{aligned} \tag{72}$$

Clearly, if one lets $\rho = T[i]$ and $\tau = \alpha \cdot T[k']$, then ρ, s, and τ, satisfy the conditions of Claim 12.4. Thus, there exists $s'' > s$ such that $g(\rho, s'') \neq g(\rho, s)$ — a contradiction. □ (*Claim 12.10*)

Claim 12.11. There exists j such that $W_{\mathbf{M}'(T[j])} = L$.

Proof of Claim. Let i_0 be *least* such that $W_{\mathbf{M}(T[i_0])} = L$. By Claim 12.10, there exist s_0 and k_0 such that, for each $j \geq s_0$,

$$(\forall k \geq k_0)\big[Q(T[i_0], j, T[k])\big]. \tag{73}$$

By Claim 12.9(ii), $\mathbf{M}'$ will eventually abandon any conjecture of the form $h(T[i], \cdot)$, where $i < i_0$. Furthermore, by Claim 12.9(i), for sufficiently large j, $T[i_0]$ will be the *shortest* $\rho \subseteq T[j]$ such that $Q(\rho, j, T[j])$. If, for some such j, $\mathbf{M}'$ outputs $h(T[i_0], j)$ *and* j satisfies (73), then, by Claim 12.6, $W_{h(T[i_0], j)} = W_{\mathbf{M}(T[i_0])}$ $(= L)$. On the other hand, if, for each such j, $\mathbf{M}'$ *never* outputs $h(T[i_0], j)$ *or* j does *not* satisfy (73), then, by the construction of $\mathbf{M}'$, there must exist i_1, s_1, and k_1 such that

- $i_1 > i_0$,
- $\mathbf{M}'$ outputs $h(T[i_1], s_1)$, and
- $(\forall k \geq k_1)\big[Q(T[i_1], s_1, k)\big]$.

In such a case, by Claim 12.6, $W_{h(T[i_1], s_1)} = W_{\mathbf{M}(T[i_1])}$, and, by (the contrapositive of) Claim 12.9(ii), $W_{\mathbf{M}(T[i_1])} = L$. □ (*Claim 12.11*)

□ (*Theorem 12*)

4 Optimal Identification of Subclasses

In this section, we show that, for an arbitrary optimal learner **F** of a class of languages $\mathcal{L}$, **F** optimally identifies a subclass $\mathcal{K}$ of $\mathcal{L}$ iff **F** is class-preserving with respect to $\mathcal{K}$ (Theorem 13 below). The reader may wonder: if **F** optimally identifies $\mathcal{L}$, how can there exist a subclass $\mathcal{K}$ of $\mathcal{L}$ such that **F** does *not* optimally identify $\mathcal{K}$? Intuitively, this can occur as follows. A learner **G**, knowing that a language L satisfies $L \in \mathcal{L} - \mathcal{K}$, never outputs a grammar for L. This, in turn, can allow **G** to converge to a correct hypothesis on *less* of some $T \in \textsc{Txt}_{\mathcal{K}}$ for which $(\exists \sigma \in \textsc{Seq}_L)[\sigma \subset T]$.[11]

The following is the main result of this section.

Theorem 13. Suppose that **F** optimally identifies $\mathcal{L}$. Then,

$$(\forall \mathcal{K} \subseteq \mathcal{L})[\mathbf{F} \text{ optimally identifies } \mathcal{K} \;\Leftrightarrow\; \mathbf{F} \text{ class-preservingly identifies } \mathcal{K}]. \tag{74}$$

Proof. Let **F** and $\mathcal{L}$ be as stated, and let $\mathcal{K} \subseteq \mathcal{L}$ be fixed.

($\Rightarrow$): Immediate by (b) $\Rightarrow$ (c) of Theorem 8.

($\Leftarrow$): By way of contradiction, suppose that **F** class-preservingly identifies $\mathcal{K}$, but *not* optimally. Let **G** be such that $\mathbf{G} \prec_{\mathcal{K}} \mathbf{F}$. Let **H** be such that, for each σ,

$$\mathbf{H}(\sigma) = \begin{cases} \mathbf{G}(\sigma), \text{ if } \sigma \in \textsc{Seq}_{\mathcal{K}}; \\ \mathbf{F}(\sigma), \text{ otherwise.} \end{cases} \tag{75}$$

Clearly, **H** identifies $\mathcal{L}$ and $\mathbf{H} \prec_{\mathcal{K}} \mathbf{F}$. Thus, if it can be shown that $\mathbf{H} \preceq_{(\mathcal{L}-\mathcal{K})} \mathbf{F}$, then this would (as desired) contradict the fact that **F** optimally identifies $\mathcal{L}$. So, suppose $\mathbf{H} \not\preceq_{(\mathcal{L}-\mathcal{K})} \mathbf{F}$. Let $L \in \mathcal{L} - \mathcal{K}$ and $T \in \textsc{Txt}_L$ be such that $\mathrm{conv}(\mathbf{F}, T) < \mathrm{conv}(\mathbf{H}, T)$. By Lemma 7 (with $A = \textsc{Seq} - \textsc{Seq}_{\mathcal{K}}$), there exists σ such that $\sigma \in \textsc{Seq}_{\mathcal{K}}$ and $T[\mathrm{conv}(\mathbf{F}, T)] \subseteq \sigma \subset T$. By the latter and Proposition 2, $W_{\mathbf{F}(\sigma)} = L$ ($\notin \mathcal{K}$), which contradicts the supposition that **F** class-preservingly identifies $\mathcal{K}$. □ (*Theorem 13*)

Remark 14. One might hope for a characterization similar to Theorem 13, but involving only $\mathcal{K}$ and $\mathcal{L}$ (and *not* **F**) on the right-hand side of the "$\Leftrightarrow$", i.e., **F** optimally identifies $\mathcal{K} \Leftrightarrow P(\mathcal{K}, \mathcal{L})$, for some predicate $P \subseteq (2^{2^{\mathbb{N}}} \times 2^{2^{\mathbb{N}}})$. However, such a characterization is not possible, as the following example demonstrates. Let $\mathcal{L} = \big\{\{0\}, \{1\}\big\}$. Let $p_{\{0\}}$ and $p_{\{1\}}$ be grammars for $\{0\}$ and $\{1\}$, respectively. For each σ, let $\mathbf{M}_0$ and $\mathbf{M}_1$ be as follows.

$$\mathbf{M}_0(\sigma) = \begin{cases} p_{\{0\}}, \text{ if content}(\sigma) \subseteq \{0\}; \\ p_{\{1\}}, \text{ otherwise.} \end{cases} \tag{76}$$

$$\mathbf{M}_1(\sigma) = \begin{cases} p_{\{1\}}, \text{ if content}(\sigma) \subseteq \{1\}; \\ p_{\{0\}}, \text{ otherwise.} \end{cases} \tag{77}$$

It is easy to verify that both $\mathbf{M}_0$ and $\mathbf{M}_1$ optimally identify $\mathcal{L}$. However, if one lets $\mathcal{K} = \big\{\{0\}\big\}$, then $\mathbf{M}_0$ optimally identifies $\mathcal{K}$; whereas, $\mathbf{M}_1$ does *not*.

[11] This can be observed in the learners **M** and **F** given near the end of Section 1.

Despite Remark 14, Corollary 15 below gives a useful *necessary* condition similar to Theorem 13. Moreover, this condition involves only $\mathcal{K}$ and $\mathcal{L}$ (and *not* $\mathbf{F}$) on the right-hand side of the "$\Rightarrow$".

Corollary 15. Suppose that $\mathbf{F}$ optimally identifies $\mathcal{L}$. Then,

$$(\forall \mathcal{K} \subseteq \mathcal{L})\Big[\mathbf{F} \text{ optimally identifies } \mathcal{K} \ \Rightarrow \\ (\forall L, L' \in \mathcal{L})\big[[L \notin \mathcal{K} \ \wedge \ L \subseteq L'] \ \Rightarrow \ L' \notin \mathcal{K}\big]\Big]. \tag{78}$$

Proof of Corollary. Let $\mathbf{F}$, $\mathcal{L}$, and $\mathcal{K}$ be as stated, and suppose that $\mathbf{F}$ optimally identifies $\mathcal{L}$. Further suppose, by way of contradiction, that $L, L' \in \mathcal{L}$ are such that $[L \notin \mathcal{K} \ \wedge \ L \subseteq L' \ \wedge \ L' \in \mathcal{K}]$. Let $T \in \textsc{Txt}_L$ be fixed, and let $\sigma = T[\text{conv}(\mathbf{F}, T)]$. Clearly, $W_{\mathbf{F}(\sigma)} = L$ $(\notin \mathcal{K})$. Furthermore, since $L \subseteq L'$, $\sigma \in \textsc{Seq}_{L'}$. Thus, $\mathbf{F}$ does *not* class-preservingly identify $\mathcal{K}$, which contradicts ($\Rightarrow$) of Theorem 13. □ (*Corollary 15*)

References

[Ang80] D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135, 1980.

[Bar77] J. M. Barzdin. Inductive inference of automata, functions, and programs. *American Mathematical Society Translations: Series 2*, 109:107–112, 1977. Appeared originally (in Russian) in *Proceeedings of the 20th International Conference of Mathematicians*, 1974.

[BB75] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28(2):125–155, 1975.

[BCM+08] G. Baliga, J. Case, W. Merkle, F. Stephan, and W. Wiehagen. When unlearning helps. *Information and Computation*, 206(5):694–709, 2008.

[BF74] J. Bārzdiņš and R. Freivalds. Prediction and limiting synthesis of recursively enumerable classes of functions. *Latvijas Valsts Univ. Zinatn. Raksti*, 210:101–111, 1974. In Russian.

[CCJS07] L. Carlucci, J. Case, S. Jain, and F. Stephan. Results on memory-limited U-shaped learning. *Information and Computation*, 205(10):1551–1573, 2007.

[CCJS08] L. Carlucci, J. Case, S. Jain, and F. Stephan. Non-U-shaped vacillatory and team learning. *Journal of Computer and Systems Sciences*, 74(4):409–430, 2008.

[CM08] J. Case and S. E. Moelius. U-shaped, iterative, and iterative-with-counter learning. *Machine Learning*, 72(1-2):63–88, 2008.

[CS83] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25(2):193–220, 1983.

[Ful90] M. Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85(1):1–11, 1990.

[Gol67] E. M. Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

[JORS99] S. Jain, D. Osherson, J. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory*. MIT Press, second edition, 1999.

[OSW86] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, first edition, 1986.

[Pit89] L. Pitt. Inductive inference, DFAs, and computational complexity. In *Proceedings of the 2nd International Workshop on Analogical and Inductive Inference (AII'89)*, volume 397 of *Lecture Notes in Computer Science*, pages 18–44. Springer Berlin/Heidelberg, 1989.

[Rog67] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw Hill, New York, 1967. Reprinted, MIT Press, 1987.

[Wie91] R. Wiehagen. A thesis in inductive inference. In *Proceedings of the 1st International Workshop on Nonmonotonic and Inductive Logic (1990)*, volume 543 of *Lecture Notes in Computer Science*, pages 184–207. Springer Berlin/Heidelberg, 1991.