1 Fundamentals of Learning Theory

1.1 Learning Paradigms

Learning typically involves

1. a learner,
2. a thing to be learned,
3. an environment in which the thing to be learned is exhibited to the learner,
4. the hypotheses that occur to the learner about the thing to be learned on the basis of the environment.

Learning is said to be successful in a given environment if the learner's hypotheses about the thing to be learned eventually become stable and accurate. To fix our subject matter, let us agree to call something "learning" just in case it can be described in roughly these terms.

Language acquisition by children is an example of learning in the intended sense. Children are the learners; a natural language is the thing to be learned; the corpus of sentences available to the child is the relevant environment; grammars serve as hypotheses. Language acquisition is complete when the child's shifting hypotheses about the ambient language stabilize to an accurate grammar.

By a *(learning) paradigm* we mean any precise rendition of the basic concepts of learning just introduced. Thus a paradigm provides definitions corresponding to 1 through 4 and advances a specific criterion of successful learning. The latter requires, at minimum, definition of the notions of stability and accuracy as used earlier. Alternative learning paradigms thus offer alternative conceptions of learners, environments, hypotheses, and so forth. *Learning theory* is the study of learning paradigms.

In 1967 E. M. Gold introduced a paradigm that has proved to be fundamental to learning theory. This paradigm is called *identification.* All the other paradigms to be discussed in this book may be conceived as generalizations of identification. The present chapter defines the identification paradigm, thereby laying the foundation for all subsequent developments. We proceed as follows. Section 1.2 provides essential background concepts and terminology. Section 1.3 is devoted to the construal of items 1 through 4 within the identification paradigm. The relevant criterion of successful learning is given in section 1.4. Section 1.5 discusses an essential feature of identification and related paradigms.

1.2 Background Material

1.2.1 Functions and Recursive Functions

We let N be the set $\{0, 1, 2, \ldots\}$ of natural numbers. The set of all functions (partial or total) from N to N is denoted $\mathscr{F}$. Following standard mathematical practice, members of $\mathscr{F}$ will be construed as sets of ordered pairs of numbers satisfying the "single-valuedness" condition. Single valuedness specifies that no two pairs of numbers with the same first coordinates may occur in the same function. There are nondenumerably many functions in $\mathscr{F}$. We let the symbols $\varphi, \psi, \theta, \varphi', \ldots$, represent possibly partial functions in $\mathscr{F}$. The symbols $f, g, h, f', \ldots$, are reserved for total functions in $\mathscr{F}$. If $\varphi \in \mathscr{F}$ is defined on $x \in N$, we sometimes write $\varphi(x)\downarrow$. Otherwise, we write $\varphi(x)\uparrow$.

It will often be useful to construe individual numbers as "tuples" of numbers. This is achieved as follows. For each $n \in N$ we select some computable isomorphism between N^n (i.e., the n-fold Cartesian product on N) and N. For $x_1, x_2, \ldots, x_n \in N$, $\langle x_1, x_2, \ldots, x_n \rangle$ denotes the image under this function of the (ordered) n-tuple $(x_1, x_2, \ldots, x_n)$. In using this notation, the reader should keep the following facts in mind (illustrating with $n = 2$):

1. For all x, $y \in N$, $\langle x, y \rangle$ is a number, but (x, y) is an ordered pair of numbers.
2. There is an effective procedure for finding $\langle x, y \rangle$ on the basis of x, $y \in N$.
3. There is an effective procedure for finding both x and y on the basis of $\langle x, y \rangle \in N$.

For $A, B \subseteq N$, we let $A \times B = \{\langle x, y \rangle | x \in A \text{ and } y \in B\}$. Note that $A \times B$ is a set of numbers, not a set of ordered pairs as in the usual definition of $A \times B$, the Cartesian product of A and B. We also introduce "projection functions," π_1 and π_2 with the property that for all $x \in N$, $\langle \pi_1(x), \pi_2(x) \rangle = x$. Thus π_1 "picks out" the first coordinate of the pair coded by x; π_2 picks out the second coordinate.

The set of recursive functions (partial or total) from N to N is denoted: $\mathscr{F}^{\text{rec}}$. $\mathscr{F}^{\text{rec}}$ is a denumerable subset of $\mathscr{F}$. The members of $\mathscr{F}^{\text{rec}}$ may be thought of as those functions that are calculable by machine. "Machines" can be understood as Turing machines, LISP programs, or any other canonical means of computation. For concreteness we shall occasionally invoke Turing machines to explain various definitions and results; however, any other programming system would serve equally well.

We wish now to assign code numbers to the partial recursive functions.

This is achieved by listing the members of $\mathscr{F}^{\text{rec}}$ and using ordinal positions in the list as code numbers. To be useful, however, this listing of $\mathscr{F}^{\text{rec}}$ must meet certain conditions, specifically:

DEFINITION 1.2.1A An *acceptable indexing* of $\mathscr{F}^{\text{rec}}$ is an enumeration φ_0, $\varphi_1, \varphi_2, \ldots$, of (all of) $\mathscr{F}^{\text{rec}}$ that meets the following conditions:

i. For some $\psi \in \mathscr{F}^{\text{rec}}$, $\psi(\langle i, x \rangle) = \varphi_i(x)$ for all i, $x \in N$.
ii. For some total $s \in \mathscr{F}^{\text{rec}}$,

$$\varphi_{s(\langle i, m, x_1, \ldots, x_m \rangle)}(\langle y_1, \ldots, y_n \rangle) = \varphi_i(\langle x_1, \ldots, x_m, y_1, \ldots, y_n \rangle)$$

for all i, m, $x_1, \ldots, x_m$, $y_1, \ldots, y_n \in N$.

Part ii of the definition allows us to parameterize the first m arguments with respect to the ith partial recursive function.

Relative to a given acceptable indexing $\varphi_0, \varphi_1, \ldots, \varphi_i, \ldots$, φ_i is referred to as the partial recursive function of *index i*. Intuitively i may be thought of as the code for a program that computes φ_i. Indeed, one acceptable indexing of $\mathscr{F}^{\text{rec}}$ results from enumerating all Turing machines (or other canonical computing agents) in lexicographical order and taking φ_i to be the function computed by the ith Turing machine in the enumeration. This indexing orders Turing machines by their size (measured in number of symbols), resolving ties by recourse to some arbitrary alphabetization of Turing machine notation. We assume that Turing machines are specified by a finite string of symbols drawn from a fixed, finite alphabet. The reader may safely adopt this size interpretation of indexes, since none of the results in this book depend on which acceptable indexing of $\mathscr{F}^{\text{rec}}$ is selected. This invariance is a consequence of the following result.

LEMMA 1.2.1A (Rogers, 1958) Let $\varphi_0, \varphi_1, \ldots$, and $\psi_0, \psi_1, \ldots$, be any two acceptable indexings of $\mathscr{F}^{\text{rec}}$. Then there is a one-one, onto, total $f \in \mathscr{F}^{\text{rec}}$ such that $\varphi_x = \psi_{f(x)}$ for all $x \in N$.

Proof See Machtey and Young (1978, theorem 3.4.7). □

Thus any two acceptable indexings of the partial recursive functions are identical up to a recursive isomorphism. We now fix on some specific acceptable indexing of $\mathscr{F}^{\text{rec}}$ (of the reader's choice). Indexes are henceforth interpreted accordingly.

The following simple result will be useful in subsequent developments.

LEMMA 1.2.1B For all $i \in N$, the set $\{j | \varphi_j = \varphi_i\}$ is infinite.

Proof See Machtey and Young (1978, proposition 3.4.5). □

Lemma 1.2.1B reflects the fact that any computable function can be programmed in infinitely many ways (e.g., by inserting redundant instructions into a given program). If $\varphi_i = \varphi_j$, we often say that i and j are *equivalent*.

1.2.2 Recursively Enumerable Sets

For all $i \in N$ the domain of φ_i is denoted: W_i. As a consequence of lemma 1.2.1B, for all $i_0 \in N$ the set $\{j \mid W_j = W_{i_0}\}$ is infinite. A set $S \subseteq N$ is called *recursively enumerable* (or *r.e.*) just in case there is $i \in N$ such that $S = W_i$; in this case i is said to be an *index for* S (there are infinitely many indexes for each r.e. set). Intuitively a set S is r.e. just in case there is a mechanical procedure P (called a *positive test*) such that for all $x \in N$, P eventually halts on input x if and only if $x \in S$; indeed, the program coded by i serves this purpose for W_i. Construed another way, the r.e. sets are those that can be "generated" mechanically such as by a grammar.

The class of all r.e. sets is denoted: RE. Thus $\text{RE} = \{W_i \mid i \in N\}$.

Three special kinds of r.e. sets will often be of interest. These are presented in definitions 1.2.2A through 1.2.2E. $S \in \text{RE}$ is called *finite* just in case S has only finitely many members; it is called *infinite* otherwise.

Definition 1.2.2A The collection of all finite sets is denoted: RE_{fin}.

$S \in \text{RE}$ is called *recursive* just in case $\overline{S} \in \text{RE}$.

Definition 1.2.2B The collection of all recursive sets is denoted: RE_{rec}.

It can be shown that $\text{RE}_{\text{fin}} \subset \text{RE}_{\text{rec}} \subset \text{RE}$. RE_{rec} may also be characterized as follows.

Definition 1.2.2C $f \in \mathscr{F}$ is said to be the *characteristic function for* $S \subseteq N$ just in case for all $x \in N$,

$$f(x) = \begin{cases} 0, & \text{if } x \in S, \\ 1, & \text{if } x \in \overline{S}. \end{cases}$$

It is not difficult to prove that $S \subseteq N$ is recursive if and only if its characteristic function is recursive. Intuitively $S \in \text{RE}_{\text{rec}}$ just in case there is a mechanical procedure (called a *test*) that eventually responds "yes" to any input drawn from S and eventually responds "no" to any other input (thus a test, unlike a positive test, is required to respond to every input).

Turning to the third special kind of r.e. set, recall from section 1.2.1 that each $n \in N$ represents a unique ordered pair of numbers, namely the pair (i,j) such that $\langle i,j \rangle = n$. Accordingly:

Definition 1.2.2D

i. $S \subseteq N$ is said to *represent* the set $\{(x,y) \mid \langle x,y \rangle \in S\}$ of ordered pairs.
ii. $S \subseteq N$ is called *single valued* just in case S represents a function.
iii. A single-valued set is said to be *total* just in case the function it represents is total.

Equivalently S is single valued just in case for all $x, y, z \in N$, if $\langle x,y \rangle \in S$ and $\langle x,z \rangle \in S$, then $y = z$. Plainly a single-valued set S represents the function φ defined by the condition that for all x, $y \in N$, $\varphi(x) = y$ if and only if $\langle x,y \rangle \in S$. A single-valued set S is total just in case for all $x \in N$ there is $y \in N$ such that $\langle x,y \rangle \in S$.

Definition 1.2.2E The collection of all single-valued, total, r.e. sets is denoted: RE_{svt}.

Exercises

1.2.2A Let $S \subseteq N$ be single valued, and suppose that S represents $\varphi \in \mathscr{F}$. Show that

a. $\varphi \in \mathscr{F}^{\text{rec}}$ if and only if $S \in \text{RE}$.
b. φ is total recursive if and only if S is total and r.e.
c. if $S \in \text{RE}$ and S is total, then S is recursive.

1.2.2B
a. Prove: Let $f \in \mathscr{F}$ be the characteristic function for $S \subseteq N$. Then $S \in \text{RE}_{\text{rec}}$ if and only if some $T \in \text{RE}_{\text{svt}}$ represents f.
b. Show that there is a total recursive function f such that for all $i \in N$, if $W_i \in \text{RE}_{\text{svt}}$, then $\varphi_{f(i)}$ is the characteristic function for W_i.
c. Prove: $\text{RE}_{\text{svt}} \subset \text{RE}_{\text{rec}}$.

1.3 Identification: Basic Concepts

We now consider items 1 through 4 of section 1.1 as they are construed within the paradigm of identification. We begin with 2.

1.3.1 Languages

Identification is intended as a model of language acquisition by children, so languages are the things to be learned. In the model languages are conceived in a manner familiar from the theory of formal grammar (see Hopcroft and Ullman, 1979, ch. 1) where a sentence is taken to be a finite string of symbols drawn from some fixed, finite alphabet. A language is then construed as a subset of the set of all possible sentences. This definition embraces rich conceptions of sentences, for which derivational histories, meanings, and even bits of context are parts of sentences. Since finite derivations of almost any nature can be collapsed into strings of symbols drawn from a suitably extended vocabulary, it is sufficiently general to construe a language as the set of such strings. Simplifying matters even further, it is useful to conceive of the strings of a language (collapsed derivations or otherwise) as single natural numbers; this is appropriate in light of simple coding techniques for mapping strings univocally into natural numbers (for discussion, see Rogers, 1967, sec. 1.10). In this way a language is conceived as a set of natural numbers.

But not just any subset of N counts as a language within the identification paradigm. Since natural languages are considered to have grammars, and since grammars are intertranslatable with Turing machines, we restrict attention to the recursively enumerable subsets of N—that is, to RE. Henceforth in this book the term "language" is reserved for the r.e. sets. We use the symbols $L, L', \ldots$, to denote languages.

In sum, within the identification paradigm what is learned are languages, where languages are taken to be r.e. subsets of N (equivalently, members of RE).

It is interesting in this context to consider the significance of single-valued languages. Some linguistic theories envision the relation between underlying and superficial representations of a sentence as a species of functional dependence, different natural languages implementing different functions of this kind (e.g., Wexler and Culicover, 1980). It is assumed moreover that contextual clues give the child access to the underlying representation of a sentence as well as to its superficial structure. On this view a sentence is understood as an ordered pair of representations, underlying and superficial, and competence for a language consists in knowing the function that maps one representation onto the other (variants of this basic idea are possible). All of this suggests that empirically faithful models of linguistic development construe natural languages as certain kinds of single-valued sets.

Single-valued languages are also the appropriate means of representing various learning situations distinct from language acquisition. For example, a scientist faced with an unknown functional dependency can be conceived as attempting to master a single-valued language selected arbitrarily from a set of theoretical possibilities.

For these reasons we shall often devote special attention to single-valued languages, treating them separately from arbitrary r.e. sets.

1.3.2 Hypotheses

Languages construed as r.e. subsets of N, it is natural to identify the learner's conjectures (item 4) with associated Turing machines. In turn, we may appeal to our acceptable indexing of the partial recursive functions (section 1.2.1) and identify Turing machines with indexes for r.e. sets (i.e., with N itself). Thus within the identification paradigm the number i is the hypothesis associated with the language W_i (and with the language W_j, if $W_j = W_i$).

1.3.3 Environments

We turn now to item 3.

Definition 1.3.3A A *text* is an infinite sequence $i_0, i_1, \ldots$, of natural numbers. The set of numbers appearing in a text t is denoted: $\text{rng}(t)$. A text is said to be *for* a set $S \subseteq N$ just in case $\text{rng}(t) = S$. The set of all possible texts is denoted: $\mathscr{T}$.

Example 1.3.3A

$t = 0, 0, 2, 2, 4, 4, 6, 6, \ldots$ is a text. Since $\text{rng}(t) = \{0, 2, 4, 6, \ldots\}$, t is a text for the language consisting of the even numbers.

Let $t \in \mathscr{T}$ be for $L \in \text{RE}$. Then every member of L appears somewhere in t (repetitions allowed), and no member of $\overline{L}$ appears in t. There are nondenumerably many texts for a language with at least two elements. There is only one text for a singleton language (i.e., a language consisting of only one element). There are no texts for the empty language.

Within the identification paradigm an environment for a language L is construed as a text for L. We let the symbols $r, s, t, r', \ldots,$ represent texts.

From the point of view of language acquisition, texts may be understood as follows. We imagine that the sentences of a language are presented to the child in an arbitrary order, repetitions allowed, with no ungrammatical intrusions. Negative information is withheld—that is, ungrammatical strings, so marked, are not presented. Each sentence of the language eventually appears in the available corpus, but no restriction is placed on the order of their arrival. Sentences are presented forever; no sentence ends the series.

The foregoing picture of the child's linguistic environment is motivated by recent studies of language acquisition. Brown and Hanlon (1970), for example, give reason to believe that negative information is not systematically available to the language learner. Studies by Newport, Gleitman, and Gleitman (1977) underline the relative insensitivity of the acquisition process to variations in the order in which language is addressed to children. And Lenneberg (1967) describes clinical cases revealing that a child's own linguistic productions are not essential to his or her mastery of an incoming language.

The following asymmetrical property of texts is worth pointing out. Let $t \in \mathscr{T}$ and $n \in N$ be given. If $n \in \mathrm{rng}(t)$, then examination of some initial segment of t suffices to verify the presence of n in t once and for all. On the other hand, no finite examination of t can definitively verify the absence of n from t (since n may turn up in t after the finite examination).

1.3.4 Learners

We turn, finally, to item 1. Consider a child learning a natural language. At any given moment the child has been exposed to only finitely many sentences. Yet he or she is typically willing to conjecture grammars for infinite languages. Within the identification paradigm the disposition to convert finite evidence into hypotheses about potentially infinite languages is the essential feature of a learner. More generally, the relation between finite evidential states and infinite languages is at the heart of inductive inference and learning theory.

Formally, let $t \in \mathscr{T}$ and $n \in N$ be given. Then the nth member of t is denoted: t_n. The sequence determined by the first n members of t is denoted: $\bar{t}_n$. The sequence $\bar{t}_n$ is called the *finite sequence* of *length n in t*. Note that for any text t, $\bar{t}_0$ is the unique sequence of length zero, namely the empty sequence which we identify with the empty set $\varnothing$. The set of all finite sequences of any length in any text is denoted: SEQ. SEQ may be thought of as the set of all possible evidential states (e.g., the set of all possible finite corpora of sentences that could be available to a child). We let the symbols $\sigma, \tau, \chi, \sigma', \ldots,$ represent finite sequences.

Now let $\sigma \in \mathrm{SEQ}$. The length of σ is denoted: $\mathrm{lh}(\sigma)$. The (unordered) set of sentences that constitute σ is denoted: $\mathrm{rng}(\sigma)$. We do not distinguish numbers from finite sequences of length 1. As a consequence of the foregoing conventions, note that $\sigma \in \mathrm{SEQ}$ is in $t \in \mathscr{T}$ if and only if $\sigma = \bar{t}_{\mathrm{lh}(\sigma)}$.

Example 1.3.4A

a. Let $t = 0, 0, 2, 2, 4, 4, \ldots$. Then $t_0 = t_1 = 0$, $t_7 = 6$, $\bar{t}_2 = (0,0)$ (but $t_2 = 2$), and $\bar{t}_4 = (0,0,2,2)$ (but $t_4 = 4$). Moreover $\mathrm{lh}(\bar{t}_7) = 7$, $\mathrm{lh}(\bar{t}_4) = 4$, $\mathrm{lh}(t_4) = 1$, $\mathrm{rng}(\bar{t}_4) = \{0,2\}$, and $t_0 = \bar{t}_1 = 0$. $\tau = (2,2,4)$ is not in t because t does not begin with τ.
b. Let $\sigma = (5,2,2,6,8)$. Then $\mathrm{lh}(\sigma) = 5$, and $\mathrm{rng}(\sigma) = \{5,2,6,8\}$.

With evidential states now construed as finite sequences and conjectures construed as natural numbers (section 1.3.2), learners may be conceived as functions from one to the other, that is, as functions from SEQ to N. Put differently, learning may be viewed as the process of converting evidence into theories (successful learning has yet to be defined). However, rather than taking learners to be functions from SEQ to N, it will facilitate later developments to code evidential states as natural numbers. Thus we choose some fixed, computable isomorphism between SEQ and N and interpret, as needed, the number n as some unique member of SEQ. None of our results depend on which computable isomorphism between SEQ and N is chosen for this purpose. Officially then a *learning function* is a member of $\mathscr{F}$ (i.e., a function from N to N) where the domain of the function is to be thought of as the set of all possible evidential states and the range as the set of all possible hypotheses. A learning function may be partial or total, recursive or nonrecursive. A "learner" is any system that embodies a learning function. Learning theory thus applies to learners indirectly via the learning functions they implement.

To talk about learning functions, we need a notation for the mapping that codes SEQ as N. It will reduce clutter to allow finite sequences to symbolize their own code numbers. Thus "σ" represents ambiguously a finite sequence of numbers as well as the single number coding it. No harm will come of this equivocation. According to our notational conventions, for $\varphi \in \mathscr{F}$, $t \in \mathscr{T}$, and $n \in N$, the term "$\varphi(\bar{t}_n)$" denotes the result of applying φ to the code number of the finite sequence constituting the first n members of t.

Let $\varphi \in \mathscr{F}$ and $\sigma \in \text{SEQ}$ be given. In conformity with the convention governing $\uparrow$ and $\downarrow$ (section 1.2.1), if φ is defined on σ, we write: $\varphi(\sigma)\downarrow$. Otherwise, we write: $\varphi(\sigma)\uparrow$. Intuitively $\varphi(\sigma)\uparrow$ signifies that the learner implementing φ advances no hypothesis when faced with the evidence σ. This omission might result from the complexity of σ (relative to the learner's cognitive capacity), or it may arise for other reasons. If $\varphi(\sigma)\downarrow$, we often say that φ conjectures $W_{\varphi(\sigma)}$ on σ.

Example 1.3.4B

We provide some sample learning functions f, g, h, φ, $\psi \in \mathscr{F}$ by describing their behavior on SEQ. For all $\sigma \in \text{SEQ}$:

a. $f(\sigma)$ = the least index for the language $\text{rng}(\sigma)$. Informally, f behaves as if its current evidential state includes all the sentences it will ever see. Consequently it conjectures a grammar for the finite language made up of the elements received to date. Being parsimonious, f's conjectured grammars are as small as possible (relative to the acceptable indexing of section 1.2.1). We shall have occasion to refer to this function many times in later chapters.
b. $g(\sigma) = 5$. The function g has fixed ideas about the language it is observing.
c. $h(\sigma)$ = the smallest i such that $\text{rng}(\sigma) \subseteq W_i$. Here h conjectures the first language (relative to our acceptable indexing) that accounts for all the data it has received.
d. Let E be the set of even numbers.

$$\varphi(\sigma) = \begin{cases} \text{least index for } \text{rng}(\sigma), & \text{if } \text{rng}(\sigma) \subseteq E, \\ \uparrow, & \text{otherwise.} \end{cases}$$

φ is partial.
e. $\psi(\sigma)\uparrow$. Although ψ is the empty partial function, it counts as a learning function.

Exercises

1.3.4A Let L be a nonempty language, and let t be a text for L. Let f be the learning function of part a of example 1.3.4B.

a. Show that $L \in \text{RE}_{\text{fin}}$ if and only if for all but finitely many $n \in N$, $\text{rng}(\bar{t}_n) = \text{rng}(\bar{t}_{n+1})$.
b. Show that $L \in \text{RE}_{\text{fin}}$ if and only if $f(\bar{t}_n) = f(\bar{t}_{n+1})$ for all but finitely many $n \in N$.
c. Suppose that there is $n \in N$ such that for infinitely many $m \in N$, $f(\bar{t}_n) = f(\bar{t}_m)$. Show that $W_{f(\bar{t}_n)} = L$.

*1.3.4B A text t is called *ascending* if $t_n \leq t_{n+1}$ for all $n \in N$; t is called *strictly ascending* if $t_n < t_{n+1}$ for all $n \in N$.

a. Let L be a finite language of at least two members. How many ascending texts are there for L?
b. Let L be an infinite language. How many strictly ascending texts are there for L?

These kinds of texts are treated again in section 5.5.1.

1.4 Identification: Criterion of Success

Languages, hypotheses, environments, and learners are the *dramatis personae* of learning theory. In section 1.3 we presented their construal within the identification paradigm. We now turn to the associated criterion of successful learning. Within the current paradigm successful learning is said to result in "identification;" its definition proceeds in stages.

1.4.1 Identifying Texts

Definition 1.4.1A Let $\varphi \in \mathscr{F}$ and $t \in \mathscr{T}$ be given.

i. φ is said to be *defined on* t just in case $\varphi(\bar{t}_n)\downarrow$ for all $n \in N$.
ii. Let $i \in N$. φ is said to *converge on* t *to* i just in case (a) φ is defined on t and (b) for all but finitely many $n \in N$, $\varphi(\bar{t}_n) = i$.
iii. φ is said to *identify* t just in case there is $i \in N$ such that (a) φ converges on t to i and (b) $\text{rng}(t) = W_i$.

Clause ii of the definition may also be put this way: φ converges on t to i just in case φ is defined on t, and there is $n \in N$ such that $\varphi(\bar{t}_m) = i$ for all $m > n$.

The intuition behind definition 1.4.1A is as follows. A text t is fed to a learner l one number at a time. With each new input l is faced with a new finite sequence of numbers. l is defined on t if l offers hypotheses on all of these finite sequences. If l is undefined somewhere in t, then l is "stuck" at that point, lost in endless thought about the current evidence, unable to accept more data. l converges on t to an index i just in case l does not get stuck in t, and after some finite number of inputs l conjectures i thereafter. To identify t, l must converge to an index for $\text{rng}(t)$.

Let $\varphi \in \mathscr{F}$ identify $t \in \mathscr{T}$. Note that definition 1.4.1A places no finite bound on the number of times that φ "changes its mind" on t. In other words, the set $\{n \in N \mid \varphi(\bar{t}_n) \neq \varphi(\bar{t}_{n+1})\}$ may be any finite size; it may not, however, be infinite. Similarly the smallest $n_0 \in N$ such that $W_{\varphi(\bar{t}_{n_0})} = \text{rng}(t)$ may be any finite number. It is also permitted that for some n, $W_{\varphi(\bar{t}_n)} = \text{rng}(t)$, but $W_{\varphi(\bar{t}_{n+1})} \neq \text{rng}(t)$. In other words, φ may abandon correct conjectures, although φ must eventually stick with some correct conjecture.

Example 1.4.1A

a. Let t be the text 2, 4, 6, 6, 6, 6, ... Let $f \in \mathscr{F}$ be as described in part a of example 1.3.4B. On $\bar{t}_0$, f conjectures the least index for $\varnothing$; on $\bar{t}_1$, f conjectures the least index for $\{2\}$; on $\bar{t}_2$, f conjectures the least index for $\{2,4\}$; on $\bar{t}_n$ for $n \geq 3$, f conjectures the least index for $\{2,4,6\}$. Thus f converges on t to this latter index. Since $\text{rng}(t) = \{2,4,6\}$, f identifies t.
b. Let t be the text 0, 1, 2, 3, 4, 5, ... Let f and g be as described in parts a and b of example 1.3.4B. f is defined on t but does not converge on t, g converges on t, and g identifies t if and only if $W_5 = N$.
c. Let t be the text 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, ... Let $\varphi \in \mathscr{F}$ be as described in part d of example 1.3.4B. φ is defined on $\bar{t}_n$ for $n \leq 3$; it is undefined thereafter. φ is thus not defined on t.

Exercise

1.4.1A Let t be the text 0, 1, 2, 3, 4, 5, ... Let h be as described in part c of example 1.3.4B. Does h identify t?

1.4.2 Identifying Languages

Children are able to learn their language on the basis of many orderings of its sentences. Since definition 1.4.1A pertains to individual texts it does not represent this feature of language acquisition. The next definition remedies this defect.

Definition 1.4.2A Let $\varphi \in \mathscr{F}$ and $L \in \text{RE}$ be given. φ is said to *identify* L just in case φ identifies every text for L.

As a special case of the definition every learning function identifies the empty language, for which there are no texts.

Let $\varphi \in \mathscr{F}$ identify $L \in \text{RE}$, and let s and t be different texts for L. It is consistent with definition 1.4.2A that φ converge on s and t to different indexes for L. Likewise φ might require more inputs from s than from t before emitting an index for L.

Example 1.4.2A

a. Let $f \in \mathscr{F}$ be as described in part a of example 1.3.4B. Let $L = \{2,4,6\}$. Given any text t for L, there is some $n_0 \in N$ such that $L = \text{rng}(\bar{t}_m)$ for all $m \geq n_0$. Hence, for all $m \geq n_0$, $f(\bar{t}_m) = f(\bar{t}_{m+1})$ and $W_{f(\bar{t}_m)} = \text{rng}(t)$. Hence f identifies any such t. Hence f identifies L.
b. Let $g \in \mathscr{F}$ be as described in part b of example 1.3.4B. g identifies a language L if and only if 5 is an index for L.
c. Let n_0 be an index for $L = \{0,1\}$. Let $h \in \mathscr{F}$ be defined as follows: for all $\sigma \in \text{SEQ}$

$$h(\sigma) = \begin{cases} n_0, & \text{if } \sigma \text{ does not end in 1,} \\ \text{lh}(\sigma), & \text{otherwise.} \end{cases}$$

h identifies every text for L in which 1 occurs only finitely often; no other texts are identified. h does not identify L.

1.4.3 Identifying Collections of Languages

Children are able to learn any arbitrarily selected language drawn from a large class; that is, their acquisition mechanism is not prewired for just a single language. Definiton 1.4.2A does not reflect this fact. We are thus led to extend the notion of identification to collections of languages.

Definition 1.4.3A Let $\varphi \in \mathscr{F}$ be given, and let $\mathscr{L} \subseteq \text{RE}$ be a collection of languages. φ is said to *identify* $\mathscr{L}$ just in case φ identifies every $L \in \mathscr{L}$. $\mathscr{L}$ is said to be *identifiable* just in case some $\varphi \in \mathscr{F}$ identifies $\mathscr{L}$.

We let $\mathscr{L}, \mathscr{L}', \ldots$, represent collections of languages. As a special case of the definition, the empty collection of languages is identifiable.

Every singleton collection $\{L\}$ of languages is trivially identifiable. To see this, let n_0 be an index for L, and define $f \in \mathscr{F}$ as follows. For all $\sigma \in \text{SEQ}$, $f(\sigma) = n_0$. Then f identifies L, and hence f identifies $\{L\}$ (compare part b of example 1.4.2A). In contrast, questions about the identifiability of collections of more than one language are often nontrivial, for many such questions receive negative answers (as will be seen in chapter 2). Such is the consequence of requiring a single learning function to determine which of several languages is inscribed in a given text.

The foregoing example also serves to highlight the liberal attitude that we have adopted about learning. The constant function f defined above identifies W_{n_0} but exhibits not the slightest "intelligence" thereby (like the man who announces an imminent earthquake every morning). Within the identification paradigm it may thus be seen that learning presupposes neither rationality nor warranted belief but merely stable and true conjectures in the sense provided by the last three definitions. Does this liberality render identification irrelevant to human learning? The answer depends on both the domain in question, and the specific criterion of rationality to hand. To take a pertinent example, normal linguistic development seems not to culminate in warranted belief in any interesting sense, since natural languages exhibit a variety of syntactic regularities that are profoundly underdetermined by the linguistic evidence available to the child (see Chomsky 1980, 1980a, for discussion). Indeed, one might extend this argument (as does Chomsky 1980) to every nontrivial example of human learning, that is, involving a rich set of deductively interconnected beliefs to be discovered by (and not simply told to) the learner. In any such case of inductive inference, hypothesis selection is subject to drastic underdetermination by available data, and thus selected hypotheses, however true, have little warrant. We admit, however, that all of this is controversial (for an opposing point of view, see Putnam 1980), and even the notion of belief in these contexts stands in need of clarification (see section 3.2.4). In any case we shall soon consider paradigms that incorporate rationality requirements in one or another sense (see in particular sections 4.3.3, 4.3.4, 4.5.1, and 4.6.1).

To return to the identification paradigm, the following propositions provide examples of identifiable collections of languages.

PROPOSITION 1.4.3A $\mathrm{RE}_{\mathrm{fin}}$ is identifiable.

Proof Let $f \in \mathscr{F}$ be the function defined in part a of example 1.3.4B. By consulting part a of example 1.4.2A, it is easy to see that f identifies every finite language. □

PROPOSITION 1.4.3B Let $\mathscr{L} = \{N - \{x\} | x \in N\}$. Then $\mathscr{L}$ is identifiable.

Proof We define $g \in \mathscr{F}$ which identifies $\mathscr{L}$ as follows. Given any $\sigma \in \mathrm{SEQ}$, let x_σ be the least $x \in N$ such that $x \notin \mathrm{rng}(\sigma)$. Now define $g(\sigma)$ = the least index for $N - \{x_\sigma\}$. It is clear that g identifies every $L \in \mathscr{L}$, for given $x_0 \in N$ and any text t for $N - \{x_0\}$, there is an n such that $\mathrm{rng}(\bar{t}_n) \supseteq \{0, 1, \ldots, x_0 - 1\}$. Then for all $m \geq n$, $g(\bar{t}_m)$ = the least index for $N - \{x_0\}$. □

PROPOSITION 1.4.3C $\mathrm{RE}_{\mathrm{svt}}$ is identifiable.

Proof The key property of $\mathrm{RE}_{\mathrm{svt}}$ is this. Suppose that L and L' are members of $\mathrm{RE}_{\mathrm{svt}}$ and that $L \neq L'$. Then there are x, y, $y' \in N$ such that

$\langle x, y \rangle \in L$, $\langle x, y' \rangle \in L'$, and $y \neq y'$. Thus, if t is a text for L, there is an $n \in N$ such that by looking at $\bar{t}_n$, we know that t is not a text for L'.

Now we define $h \in \mathscr{F}$ which identifies $\mathrm{RE}_{\mathrm{svt}}$ as follows. For all $\sigma \in \mathrm{SEQ}$, let

$$h(\sigma) = \begin{cases} \text{least } i \text{ such that } W_i \in \mathrm{RE}_{\mathrm{svt}}, \text{ and } \mathrm{rng}(\sigma) \subseteq W_i, & \text{if such an } i \text{ exists,} \\ 0, & \text{otherwise.} \end{cases}$$

Informally, h guesses the first language in $\mathrm{RE}_{\mathrm{svt}}$ that is consistent with σ. By our preceding remarks, given a text t for $L \in \mathrm{RE}_{\mathrm{svt}}$, h will eventually conjecture the least index for L having verified that t is not a text for any L' with a smaller index. □

Exercises

1.4.3A Let $t \in \mathscr{T}$ and total $f \in \mathscr{F}$ be given.

a. Show that if f converges on t, then $\{f(\bar{t}_n) | n \in N\}$ is finite. Show that the converse is false.
b. Show that if f identifies t, then $W_{f(\bar{t}_n)} = \mathrm{rng}(t)$ for all but finitely many $n \in N$. Show that the converse is false.

1.4.3B Let $\mathscr{L} = \{N\} \cup \{E | E$ is a finite set of even numbers$\}$. Specify a learning function that identifies $\mathscr{L}$.

1.4.3C Prove: Every finite collection of languages is identifiable. (*Hint*: Keep in mind that a finite collection of languages is not the same thing as a collection of finite languages.)

1.4.3D Let $L \in \mathrm{RE}$ be given. Specify $\varphi \in \mathscr{F}$ that identifies $\{L \cup D | D \text{ finite}\}$.

1.4.3E Let $\{S_i | i \in N\}$ be any infinite collection of nonempty, mutually disjoint members of $\mathrm{RE}_{\mathrm{rec}}$. Let $\mathscr{L} = \{N - S_i | i \in N\}$. Specifiy a learning function that identifies $\mathscr{L}$.

1.4.3F Given $\mathscr{L}, \mathscr{L}' \subseteq \mathrm{RE}$, let $\mathscr{L} \times \mathscr{L}'$ be $\{L \times L' | L \in \mathscr{L}$ and $L' \in \mathscr{L}'\}$. Prove: If $\mathscr{L}, \mathscr{L}' \subseteq \mathrm{RE}$ are each identifiable, then $\mathscr{L} \times \mathscr{L}'$ is identifiable.

1.4.3G

a. Prove: $\mathscr{L} \subseteq \mathrm{RE}$ is identifiable if and only if some total $f \in \mathscr{F}$ identifies $\mathscr{L}$.
b. Let $t \in \mathscr{T}$ and $\varphi \in \mathscr{F}$ be given. We say that φ *almost identifies* t just in case there exists an $i \in N$ such that (a) $W_i = \mathrm{rng}(t)$ and (b) $\varphi(\bar{t}_n) = i$ for all but finitely many $n \in N$. (Thus φ can almost identify t without being defined on t.) φ *almost identifies* $\mathscr{L} \subseteq \mathrm{RE}$ just in case φ almost identifies every text for every language in $\mathscr{L}$. $\mathscr{L}$ is said to be *almost identifiable* just in case some $\varphi \in \mathscr{F}$ almost identifies $\mathscr{L}$. Prove: $\mathscr{L} \subseteq \mathrm{RE}$ is almost identifiable if and only if $\mathscr{L}$ is identifiable.

1.4.3H $\varphi \in \mathscr{F}$ is said to *P percent identify* $\mathscr{L} \subseteq$ RE just in case for every $L \in \mathscr{L}$ and every text t for L, φ is defined on t, and there is $i \in N$ such that (a) $W_i = L$ and (b) there is $n \in N$ such that for all $m > n$, $\varphi(\bar{t}_j) = i$ for P percent of $\{j \mid m \leq j \leq m + 99\}$. $\mathscr{L} \subseteq$ RE is said to be *P percent identifiable* just in case some $\varphi \in \mathscr{F}$ P percent identifies $\mathscr{L}$. Prove

a. if $P > 50$, then $\mathscr{L} \subseteq$ RE is P percent identifiable if and only if $\mathscr{L}$ is identifiable.
*b. if $P \leq 50$, then there is $\mathscr{L} \subseteq$ RE such that $\mathscr{L}$ is P percent identifiable but $\mathscr{L}$ is not identifiable.

1.4.3I $\varphi \in \mathscr{F}$ is said to *identify* $\mathscr{L} \subseteq$ RE *laconically* just in case for every $L \in \mathscr{L}$ and every text t for L there is $n \in N$ such that (a) $W_{\varphi(\bar{t}_n)} = L$ and (b) for all $m > n$, $\varphi(\bar{t}_m)\uparrow$. Prove: $\mathscr{L} \subseteq$ RE is identifiable if and only if $\mathscr{L}$ is identifiable laconically.

1.4.3J The property of RE_{svt} used in the proof of proposition 1.4.3C is that if L, $L' \in \text{RE}_{\text{svt}}$, $L \neq L'$, and t is a text for L, then there is an $n \in N$ such that $\bar{t}_n$ is enough to determine that t is not a text for L'. Show that there are identifiable infinite collections of languages without this property.

1.4.3K Let $\varphi \in \mathscr{F}$ be given. We define $\mathscr{L}(\varphi)$ to be $\{L \in \text{RE} \mid \varphi \text{ identifies } L\}$.

a. Let $\psi \in \mathscr{F}$ be such that for all $\sigma \in$ SEQ, $\psi(\sigma) =$ the least $n \in \text{rng}(\sigma)$. Characterize $\mathscr{L}(\psi)$.
b. Show by example that for $\varphi, \psi \in \mathscr{F}$, $\mathscr{L}(\varphi) = \mathscr{L}(\psi)$ does not imply $\varphi = \psi$.

1.5 Identification as a Limiting Process

1.5.1 Epistemology of Convergence

Let $\varphi \in \mathscr{F}$ identify $t \in \mathscr{T}$, and let $n \in N$ be given. We say that φ *begins to converge on t at moment n* just in case n is the least integer such that (1) $W_{\varphi(\bar{t}_n)} = \text{rng}(t)$ and (2) for all $m > n$, $\varphi(\bar{t}_m) = \varphi(\bar{t}_n)$. Now let $f \in \mathscr{F}$ be as defined in part a of example 1.3.4B, and let t be a text for a finite language. Then f identifies t. What information about t is required in order to determine the moment at which f begins to converge on t? It is easy to see that no finite initial segment $\bar{t}_n$ of t provides sufficient information to guarantee that f's conjectures on t have stabilized once and for all. Simply no such $\bar{t}_n$ excludes the possibility that $t_{n+10} \notin \text{rng}(\bar{t}_n)$, in which case $f(\bar{t}_{n+11}) \neq f(\bar{t}_n)$. Thus, although f in fact begins to converge on t at some definite moment n, no finite examination of t provides indefeasible grounds for determining n. (Compare the last paragraph of section 1.3.3.)

More generally, identification is said to be a "limiting process" in the sense that it concerns the behavior of a learning function on an infinite

subset of its domain. For this reason Gold (1967) refers to identification as "identification in the limit." Because of the limiting nature of identification, the behavior of a given learning function φ on a given text t cannot in general be predicted from φ's behavior on any finite portion of t. The underdetermination at issue here does not arise from the disadvantages connected with the "external" observation of a learning function at work. To make this clear, the next subsection discusses learning functions that announce their own convergence and may thus be considered to observe their own operation.

Exercise

1.5.1A Let $\mathscr{L} \subseteq$ RE be identifiable, let $\{\sigma^0, \ldots, \sigma^m\}$ be a finite subset of SEQ, and let $\{i_0, \ldots, i_n\}$ be a finite subset of N. Show that there is $\varphi \in \mathscr{F}$ such that (a) φ identifies $\mathscr{L}$ and (b) $\varphi(\sigma^0) = i_0, \ldots, \varphi(\sigma^n) = i_n$.

*1.5.2 Self-Monitoring Learning Functions

Definition 1.5.2A (after Freivald and Wiehagen 1979) Let e_0 be an index for the empty set. A function $\varphi \in \mathscr{F}$ is called *self-monitoring* just in case for all texts t, if φ identifies t, then (a) there exists a unique $n \in N$ such that $\varphi(\bar{t}_n) = e_0$, and (b) for $i > n$, $\varphi(\bar{t}_i) = \varphi(\bar{t}_{n+1})$.

Intuitively, a learner l is self-monitoring just in case it signals its own successful convergence, where the (otherwise useless) index e_0 serves as the signal. Note that once l announces e_0. l's next conjecture is definitive for t. l might be pictured as examining its own conjectures prior to emitting them. If and when l realizes that it has successfully determined the contents of t, l signals this fact by announcing e_0 on the current input, reverting thereafter to the correct hypothesis. The following proposition is suggested by our earlier remarks.

Proposition 1.5.2A No self-monitoring learning function identifies RE_{fin}.

Proof Let $\varphi \in \mathscr{F}$ be self-monitoring, and let t be any text for some $L \in \text{RE}_{\text{fin}}$. Then there is $n \in N$ such that $\varphi(\bar{t}_n) = e_0$, and for all $m > n$, $\varphi(\bar{t}_{n+1}) = \varphi(\bar{t}_m)$. Let x_0 be a fixed integer such that $x_0 \notin L$. Let t' be the text such that (a) $\bar{t}'_{n+1} = \bar{t}_{n+1}$ and (b) for all $m > n$, $t'_m = x_0$. Since $\varphi(\bar{t}'_n) = \varphi(\bar{t}_n) =$

e_0, we must have $\varphi(\bar{t}'_m) = \varphi(\bar{t}'_{n+1}) = \varphi(\bar{t}_{n+1})$ for all $m \geq n$. But $\varphi(\bar{t}_{n+1})$ is an index for L, whereas t' is a text for $L \cup \{x_0\}$. Hence φ does not identify $L \cup \{x_0\}$, and so φ does not identify $\mathrm{RE}_{\mathrm{fin}}$. □

Proposition 1.5.2A shows that identifiability does not entail identifiability by self-monitoring learning function. Informally, a learner may identify a text without it being possible for her to ever know that she has done so.

Exercises

1.5.2A Let $L, L' \in \mathrm{RE}$ be such that $L \subset L'$. Show that no self-monitoring learning function identifies $\{L, L'\}$.

1.5.2B Let $\mathscr{L}$ be the collection of languages of Proposition 1.4.3B. Show that no self-monitoring learning function identifies $\mathscr{L}$.

1.5.2C Call a collection $\mathscr{L}$ of languages *easily distinguishable* just in case for all $L \in \mathscr{L}$ there exists a finite subset S of L such that for all $L' \in \mathscr{L}$, if $L' \neq L$, then $S \nsubseteq L'$.

a. Specify an identifiable collection of languages that is not easily distinguishable.
b. Prove: Let $\mathscr{L} \subseteq \mathrm{RE}$ be given. Then some self-monitoring $\varphi \in \mathscr{F}$ identifies $\mathscr{L}$ if and only if $\mathscr{L}$ is easily distinguishable.

1.5.2D $\varphi \in \mathscr{F}$ is said to be a 1-*learner* just in case for all $t \in \mathscr{T}$ there exists no more than one $m \in N$ such that $\varphi(\bar{t}_m) \neq \varphi(\bar{t}_{m+1})$. That is, a 1-learner is limited to no more than one "mind change" per text.

a. Prove: If $\mathscr{L} \subseteq \mathrm{RE}$ is identifiable by a self-monitoring learning function, then $\mathscr{L}$ is identifiable by a 1-learner.
b. Show that the converse to part a is false.

1.5.2E Let $i \in N$. $\varphi \in \mathscr{F}$ is said to be an *i-learner* just in case for all texts t there exist no more than i numbers m such that $\varphi(\bar{t}_m) \neq \varphi(\bar{t}_{m+1})$. That is, an i-learner is limited to no more than i "mind changes" per text.

a. For $j \in N$ define $\mathscr{L}_j = \{\{0\}, \{0, 1\}, \{0, 1, 2\}, \ldots, \{0, 1, 2, \ldots, j\}\}$. Prove: For all $j \in N$, $\mathscr{L}_j$ is identifiable by an i-learner if and only if $i > j$. (*Hint:* Suppose that $\varphi \in \mathscr{F}$ is an i-learner and that $i \leq j$. Consider texts of the form $0, 0, \ldots, 0, 1, 1, \ldots, 1, \ldots, j, j, \ldots, j, j, \ldots$. What happens as the repetitions get longer and longer?)
b. For $i \in N$, let F_i be the class of i-learners. Let $F = \bigcup_i F_i$. Show that no $\varphi \in F$ identifies $\mathrm{RE}_{\mathrm{fin}}$. Show that no $\varphi \in F$ identifies $\{N - \{x\} \mid x \in N\}$.

1.5.2F Let e be an index for $\varnothing$. $\varphi \in \mathscr{F}$ is said to be a *one-shot learner* just in case for every text t the cardinality of $\{\varphi(\bar{t}_n) \mid \varphi(\bar{t}_n) \neq e\} \leq 1$. Let $\mathscr{L} \subseteq \mathrm{RE}$ be given. Show that some one-shot learner identifies $\mathscr{L}$ if and only if some self-monitoring learning function identifies $\mathscr{L}$.

---

2 Central Theorems on Identification

Within the paradigm of identification the learnability of a collection of languages amounts to its identifiability. Propositions 1.4.3A through 1.4.3C provide examples of learnable collections. In this chapter we give examples of unlearnable collections.

2.1 Locking Sequences

Many of the theorems in this book rest on the next result. To state and prove it, we introduce some more notation. For $\sigma, \tau \in \mathrm{SEQ}$, let $\sigma \wedge \tau$ be the result of concatenating τ onto the end of σ—thus $(2, 8, 2) \wedge (4, 1, 9, 3) = (2, 8, 2, 4, 1, 9, 3)$. Next, for $\sigma, \tau \in \mathrm{SEQ}$ we write "$\sigma \subseteq \tau$," if σ is an initial segment of τ, and "$\sigma \subset \tau$," if σ is a proper initial segment of τ—thus $(8, 8, 5) \subset (8, 8, 5, 3, 9)$.

Finally, let finite sequences $\sigma^0, \sigma^1, \sigma^2, \ldots$, be given such that (1) for every $i, j \in N$ either $\sigma^i \subseteq \sigma^j$ or $\sigma^j \subseteq \sigma^i$ and (2) for every $n \in N$, there is $m \in N$ such that $\mathrm{lh}(\sigma^m) \geq n$. Then there is a unique text t such that for all $n \in N$, $\sigma^n = \bar{t}_{\mathrm{lh}(\sigma^n)}$; this text is denoted: $\bigcup_n \sigma^n$.

PROPOSITION 2.1A (Blum and Blum 1975) Let $\varphi \in \mathscr{F}$ identify $L \in \mathrm{RE}$. Then there is $\sigma \in \mathrm{SEQ}$ such that (i) $\mathrm{rng}(\sigma) \subseteq L$. (ii) $W_{\varphi(\sigma)} = L$, and (iii) for all $\tau \in \mathrm{SEQ}$, if $\mathrm{rng}(\tau) \subseteq L$, then $\varphi(\sigma \wedge \tau) = \varphi(\sigma)$.

Proof (We follow Blum and Blum.) Assume that the proposition is false; that is, that there is no $\sigma \in \mathrm{SEQ}$ satisfying (i), (ii), and (iii). This implies the following condition:

(*) For every $\chi \in \mathrm{SEQ}$ such that $\mathrm{rng}(\chi) \subseteq L$ and $W_{\varphi(\chi)} = L$, there is some $\tau \in \mathrm{SEQ}$ such that $\mathrm{rng}(\tau) \subseteq L$ and $\varphi(\chi \wedge \tau) \neq \varphi(\chi)$.

We show that (*) implies the existence of a text t for L which φ does not identify, contrary to the hypothesis that φ identifies L. Let $s = s_0, s_1, s_2, \ldots$, be a text for L. We construct t in stages $0, 1, 2, \ldots$, at each stage n specifying a sequence σ^n which is in t.

Stage 0 Let $\sigma^0 \in \mathrm{SEQ}$ be such that $\mathrm{rng}(\sigma^0) \subseteq L$ and $W_{\varphi(\sigma^0)} = L$; σ^0 must exist since φ identifies L.
Stage $n + 1$ Given σ^n, there are two cases. If $W_{\varphi(\sigma^n)} \neq L$, let $\sigma^{n+1} = \sigma^n \wedge s_n$. Otherwise, by (*), let $\tau \in \mathrm{SEQ}$ be such that $\mathrm{rng}(\tau) \subseteq L$ and $\varphi(\sigma^n \wedge \tau) \neq \varphi(\sigma^n)$. Let $\sigma^{n+1} = \sigma^n \wedge \tau \wedge s_n$.

We observe that $\sigma^i \subset \sigma^{i+1}$ for all $i \in N$. Let $t = \bigcup_n \sigma^n$. t is a text for L since s_n is added to t at stage $n+1$ and no nonmembers of L are ever added to t. Finally, φ does not converge on t to an index for L since for every n either $W_{\varphi(\sigma^n)} \neq L$ or $\varphi(\sigma^n \wedge \tau) \neq \varphi(\sigma^n)$. □

Intuitively, if $\varphi \in \mathscr{F}$ identifies $L \in \mathrm{RE}$, then proposition 2.1.A guarantees the existence of a finite sequence σ that "locks" φ onto a conjecture for L in the following sense: no presentation from L can dislodge φ from $\varphi(\sigma)$. This suggests the following definiton.

Definition 2.1A Let $L \in \mathrm{RE}$, $\varphi \in \mathscr{F}$, and $\sigma \in \mathrm{SEQ}$ be given. σ is called a *locking sequence for L and φ* just in case (i) $\mathrm{rng}(\sigma) \subseteq L$, (ii) $W_{\varphi(\sigma)} = L$, and (iii) for all $\tau \in \mathrm{SEQ}$, if $\mathrm{rng}(\tau) \subseteq L$, then $\varphi(\sigma \wedge \tau) = \varphi(\sigma)$.

Thus proposition 2.1A can be put this way: if $\varphi \in \mathscr{F}$ identifies $L \in \mathrm{RE}$, then there is a locking sequence for φ and L.

As a corollary to the proof of proposition 2.1A, we have the following.

Corollary 2.1A Let $\varphi \in \mathscr{F}$ identify $L \in \mathrm{RE}$. Let $\sigma \in \mathrm{SEQ}$ be such that $\mathrm{rng}(\sigma) \subseteq L$. Then there is $\tau \in \mathrm{SEQ}$ such that $\sigma \wedge \tau$ is a locking sequence for φ and L.

Proof Just as in the proof of proposition 2.1A, if the corollary fails, we could construct a text t for L which φ fails to identify. Central to this construction is the following condition, which is analogous to (*), that holds if the corollary fails:

(**) For every $\chi \supseteq \sigma$, $\chi \in \mathrm{SEQ}$ such that $\mathrm{rng}(\chi) \subseteq L$ and $W_{\varphi(\chi)} = L$, there is some $\tau \in \mathrm{SEQ}$ such that $\mathrm{rng}(\tau) \subseteq L$ and $\varphi(\chi \wedge \tau) \neq \varphi(\chi)$.

The construction of t proceeds exactly as in the proof of proposition 2.1A, except that at stage 0 we also require $\sigma^0 \supseteq \sigma$. □

Note that proposition 2.1A does not characterize φ's behavior on elements drawn from $\bar{L}$. In particular, if $\tau \in \mathrm{SEQ}$ is such that $\mathrm{rng}(\tau) \nsubseteq L \in \mathrm{RE}$, then even if $\sigma \in \mathrm{SEQ}$ is a locking sequence for $\varphi \in \mathscr{F}$ and L, $\varphi(\sigma \wedge \tau)$ may well differ from $\varphi(\sigma)$.

Example 2.1A

a. Let $f \in \mathscr{F}$ be as described in part a of example 1.3.4B. Let $L = \{2, 4, 6\}$. Then one locking sequence for f and L is $(2, 4, 6)$; another is $(6, 4, 2, 6)$. Indeed, it is easy to see that for all $\sigma \in \mathrm{SEQ}$, σ is a locking sequence for f and $\mathrm{rng}(\sigma)$.

b. Let $g \in \mathscr{F}$ be as described in the proof of proposition 1.4.3B. Let $L = \{0, 2, 3, 4, \ldots\}$. Then $(22, 8, 4, 0)$ is a locking sequence for g and L. Indeed, any $\sigma \in \mathrm{SEQ}$ such that $0 \in \mathrm{rng}(\sigma)$ and $1 \notin \mathrm{rng}(\sigma)$ is a locking sequence for g and L.

Exercises

2.1A Let σ be a locking sequence for $\varphi \in \mathscr{F}$ and $L \in \mathrm{RE}$. Let $\tau \in \mathrm{SEQ}$ be such that $\mathrm{rng}(\tau) \subseteq L$. Show that $\sigma \wedge \tau$ is a locking sequence for φ and L. Distinguish this result from corollary 2.1A.

2.1B Refute the converse to proposition 2.1A. In other words, exhibit $\varphi \in \mathscr{F}$, $L \in \mathrm{RE}$, and $\sigma \in \mathrm{SEQ}$ such that σ is a locking sequence for φ and L, but φ does not identify L.

2.1C Let $\varphi \in \mathscr{F}$ identify $L \in \mathrm{RE}$. Let t be a text for L. t is called a *locking text* for φ and L just in case there exists $n \in N$ such that $\bar{t}_n$ is a locking sequence for φ and L. Provide a counterexample to the following conjecture: If $\varphi \in \mathscr{F}$ identifies $L \in \mathrm{RE}$, then every text for L is a locking text for φ and L.

2.2 Some Unidentifiable Collections of Languages

Proposition 2.1A may now be used to show that certain simple collections of languages are unidentifiable.

Proposition 2.2A

i. (Gold 1967). $\mathrm{RE}_{\mathrm{fin}} \cup \{N\}$ is not identifiable.

ii. Let $\mathscr{L} = \{N - \{x\} \mid x \in N\}$. Then $\mathscr{L} \cup \{N\}$ is not identifiable.

Proof

i. Suppose for a contradiction that $\varphi \in \mathscr{F}$ identifies $\mathrm{RE}_{\mathrm{fin}} \cup \{N\}$, and let σ be a locking sequence for φ and N. Note that $\mathrm{rng}(\sigma) \in \mathrm{RE}_{\mathrm{fin}}$. Clearly there is a text t for $\mathrm{rng}(\sigma)$ such that $\bar{t}_{\mathrm{lh}(\sigma)} = \sigma$. But then φ does not identify $\mathrm{rng}(\sigma)$ since φ converges on t to an index for N.

ii. Again, suppose that σ is a locking sequence for $\varphi \in \mathscr{F}$ and N, where φ identifies $\mathscr{L} \cup \{N\}$. Choose $x \notin \mathrm{rng}(\sigma)$. Then, on any text t for $N - \{x\}$ such that $\bar{t}_{\mathrm{lh}(\sigma)} = \sigma$, φ converges to an index for N and not one for $N - \{x\}$. □

Proposition 2.2A should be compared with propositions 1.4.3A and 1.4.3B.

The following fact is evident and often very useful.

LEMMA 2.2A Suppose that $\mathscr{L} \subseteq$ RE is not identifiable. Then no superset of $\mathscr{L}$ is identifiable.

From lemma 2.2A and proposition 2.2A we have corollary 2.2A.

COROLLARY 2.2A RE is not identifiable.

Corollary 2.2A should be compared with proposition 1.4.3C.

Since the collections of languages invoked in proposition 2.2A consist entirely of recursive languages, we also have corollary 2.2B.

COROLLARY 2.2B RE_{rec} is not identifiable.

Exercises

2.2A (Gold 1967) Let L be an arbitrary infinite language. Show that $\text{RE}_{\text{fin}} \cup \{L\}$ is not identifiable.

2.2B Let $\mathscr{L} \subseteq$ RE be such that for every $\sigma \in$ SEQ there is $L \in \mathscr{L}$ such that $\text{rng}(\sigma) \subseteq L$ and $L \neq N$. Show that $\mathscr{L} \cup \{N\}$ is not identifiable. (This abstracts the content of proposition 2.2A)

2.2C

a. Let $i_0 \in N$ be given. Define $\mathscr{L} = \{N - D \mid D \subseteq N$ has exactly i_0 members$\}$. Show that $\mathscr{L}$ is identifiable.
b. Let $i_0, j_0 \in N$ be such that $i_0 \neq j_0$. Define $\mathscr{L} = \{N - D \mid D \subseteq N$ has either exactly i_0 members or exactly j_0 members$\}$. Prove that $\mathscr{L}$ is not identifiable.

2.2D Exhibit $\varphi, \psi \in \mathscr{F}$ such that $\mathscr{L}(\varphi) \cup \mathscr{L}(\psi)$ is not identifiable. (For notation, see exercise 1.4.3K.) This shows that the identifiable subsets of RE are not closed under union.

2.2E

a. Let $\mathscr{L} \subseteq$ RE be an identifiable collection of infinite languages. Show that there is some infinite $L \notin \mathscr{L}$ such that $\mathscr{L} \cup \{L\}$ is identifiable. (*Hint:* First use proposition 2.1A to argue that if $L_0 \in \mathscr{L}$, then there is an $x_0 \in L_0$ such that if $L = L_0 - \{x_0\}$, then L is not a member of $\mathscr{L}$. Next define a function $\psi \in \mathscr{F}$ that identifies $\mathscr{L} \cup \{L\}$ by modifying the output of a function $\varphi \in \mathscr{F}$ that identifies $\mathscr{L}$.)
b. $\mathscr{L} \subseteq$ RE is called *saturated* just in case $\mathscr{L}$ is identifiable and no proper superset of $\mathscr{L}$ is identifiable. Prove: $\mathscr{L} \subseteq$ RE is saturated if and only if $\mathscr{L} = \text{RE}_{\text{fin}}$.

***2.2F** $\mathscr{S} \subseteq \mathscr{F}$ is said to *team identify* $\mathscr{L} \subseteq$ RE just in case for every $L \in \mathscr{L}$ there is $\varphi \in \mathscr{S}$ such that φ identifies L. Show that no finite subset of $\mathscr{F}$ team identifies RE. (See also exercise 6.2.1I.)

2.3 A Comprehensive, Identifiable Collection of Languages

The collections of languages exhibited in proposition 2.2A are so simple as to encourage the belief that only impoverished subsets of RE are identifiable. The next proposition shows this belief to be mistaken. To state it, a definition is required.

DEFINITION 2.3A Let $L, L' \in$ RE be such that $(L - L') \cup (L' - L)$ is finite. Then L and L' are said to be *finite variants* (of each other).

That is, finite variants differ by only finitely many members. Thus $E \cup \{3, 5, 7\}$ and $E - \{2, 4, 6, 8\}$ are finite variants, where E is the set of even numbers. Note that any pair of finite languages are finite variants.

PROPOSITION 2.3A (Wiehagen 1978) There is $\mathscr{L} \subseteq$ RE such that (i) for every $L \in$ RE there is $L' \in \mathscr{L}$ such that L and L' are finite variants and (ii) $\mathscr{L}$ is identifiable.

Proposition 2.3A asserts the existence of an identifiable collection that is "nearly all" of RE. To prove the proposition, two important lemmata are required.

LEMMA 2.3A (Recursion Theorem) Let total $f \in \mathscr{F}^{\text{rec}}$ be given. Then there exists $n \in N$ such that $\varphi_n = \varphi_{f(n)}$ (and so $W_{f(n)} = W_n$).

Proof See Rogers (1967, sec. 11.2, theorem I). □

DEFINITION 2.3B

i. $L \in$ RE is said to be *self-describing* just in case the smallest $x \in L$ is such that $L = W_x$.
ii. The collection $\{L \in \text{RE} \mid L$ is self-describing$\}$ is denoted: RE_{sd}.

LEMMA 2.3B For every $L \in$ RE there is $L' \in \text{RE}_{\text{sd}}$ such that L and L' are finite variants.

Proof Fix $L \in$ RE. Define a recursive function f by the condition that for all $n \in N$, $W_{f(n)} = (L \cup \{n\}) \cap \{n, n+1, n+2, \ldots\}$. That such an f exists is a consequence of the definition of acceptable indexing definition 1.21A(ii). To see this, if $L = W_{i_0}$, there is $j_0 \in N$ such that for all $n, x \in N$:

$$\varphi_{j_0}(\langle n, x\rangle) = \begin{cases} \varphi_{i_0}(x), & \text{if } x > n, \\ 1, & \text{if } x = n, \\ \uparrow, & \text{if } x < n. \end{cases}$$

Now by definition 1.2.1A(ii) there is a function g such that $\varphi_{g(\langle j_0,n\rangle)}(x) = \varphi_{j_0}(\langle n,x\rangle)$. By setting $f(n) = g(\langle j_0, n\rangle)$ for all $n \in N$, $W_{f(n)}$ has the desired properties.

Now by the recursion theorem there is $n \in N$ such that $W_n = W_{f(n)} = (L \cup \{n\}) \cap \{n, n+1, n+2, \ldots\}$. Clearly W_n is self-describing and is a finite variant of L. □

Proof of Proposition 2.3A By lemma 2.3B it suffices to show that $\mathrm{RE}_{\mathrm{sd}}$ is identifiable. But this is trivial. Define $f \in \mathscr{F}$ such that for all $\sigma \in \mathrm{SEQ}$, $f(\sigma)$ is the smallest number in $\mathrm{rng}(\sigma)$. Then f identifies $\mathrm{RE}_{\mathrm{sd}}$. □

Exercises

2.3A Show that for no $L \in \mathrm{RE}$ does $\mathrm{RE}_{\mathrm{sd}}$ include every finite variant of L.

2.3B Specify $\mathscr{L} \subseteq \mathrm{RE}$ such that (a) for all $L, L' \in \mathscr{L}$, if $L \neq L'$, then L and L' are not finite variants, and (b) $\mathscr{L}$ is not identifiable.

2.4 Identifiable Collections Characterized

The next proposition provides a necessary and sufficient condition for the identifiability of a collection of languages.

PROPOSITION 2.4A (Angluin 1980) $\mathscr{L} \subseteq \mathrm{RE}$ is identifiable if and only if for all $L \in \mathscr{L}$ there is a finite subset D of L such that for all $L' \in \mathscr{L}$, if $D \subseteq L'$, then $L' \not\subset L$.

Proof First suppose that $\mathscr{L} \subseteq \mathrm{RE}$ is identifiable, and let $\varphi \in \mathscr{F}$ witness this. By proposition 2.1A, for each $L \in \mathscr{L}$ choose a locking sequence σ_L for φ and L. Since for each $L \in \mathscr{L}$, $\mathrm{rng}(\sigma_L)$ is a finite subset of L, it suffices to prove that for all $L' \in \mathscr{L}$, if $\mathrm{rng}(\sigma_L) \subseteq L'$, then $L' \not\subset L$. Suppose otherwise for some $L, L' \in \mathscr{L}$, and let t be a text for L' such that $\bar{t}_{\mathrm{lh}(\sigma_L)} = \sigma_L$. Then φ converges on t to $L \neq L' = \mathrm{rng}(t)$. Thus φ fails to identify L', contradicting our choice of φ.

For the other direction, suppose that for every $L \in \mathscr{L}$ there is a finite $D_L \subseteq L$ such that $D_L \subseteq L'$ and $L' \in \mathscr{L}$ implies $L' \not\subset L$. We define $f \in \mathscr{F}$ as follows. For all $\sigma \in \mathrm{SEQ}$,

$$f(\sigma) = \begin{cases} \text{least } i \text{ such that } i \text{ is an index for some} & \\ L \in \mathscr{L} \text{ such that } D_L \subseteq \mathrm{rng}(\sigma) \subseteq L, & \text{if such an } i \text{ exists,} \\ 0, & \text{otherwise.} \end{cases}$$

To see that f identifies $\mathscr{L}$, fix $L \in \mathscr{L}$, and let t be a text for L. Let i be the least index for L. Then there is an $n \in N$ such that

1. $\mathrm{rng}(\bar{t}_n) \supseteq D_L$,
2. if $j < i$, $W_j \in \mathscr{L}$, and $L \not\subseteq W_j$, then $\mathrm{rng}(\bar{t}_n) \not\subseteq W_j$.

We claim that $f(\bar{t}_m) = i$ for all $m \geq n$. By 1 and the fact that t is a text for L, f will conjecture i on $\bar{t}_m$ unless there is $j < i$ such that $W_j = L' \in \mathscr{L}$ and $D_{L'} \subseteq \mathrm{rng}(\bar{t}_m) \subset L'$. If $\mathrm{rng}(\bar{t}_m) \supseteq D_{L'}$, then $L \supseteq D_{L'}$ so by the condition on $D_{L'}$, $L \not\subset L'$. But then by 2, $\mathrm{rng}(\bar{t}_m) \not\subseteq L'$. Thus on $\bar{t}_m$, f will not conjecture j for any $j < i$. □

Exercises

2.4A Specify a collection of finite sets meeting the conditions of proposition 2.4A with respect to

a. $\mathrm{RE}_{\mathrm{fin}}$.
b. $\{N - \{x\} \mid x \in N\}$.

2.4B

a. Use proposition 2.4A to provide alternative proofs of propositions 1.4.3A, 1.4.3B, and 1.4.3C.
b. Use proposition 2.4A to provide an alternative proof of proposition 2.2A.

2.5 Identifiability of Single-Valued Languages

Every language may be paired with a structurally identical single-valued language in the following way.

DEFINITION 2.5A We let S be the function from RE to RE defined as follows. For all $L \in \mathrm{RE}$, $S(L) = \{\langle x, 0\rangle \mid x \in L\}$. For $\mathscr{L} \subseteq \mathrm{RE}$, we define $S(\mathscr{L})$ to be $\{S(L) \mid L \in \mathscr{L}\}$.

Example 2.5A

a. Let L be the finite language $\{2,4,6\}$. Then $S(L)$ is the finite, single-valued language $\{\langle 2,0\rangle, \langle 4,0\rangle, \langle 6,0\rangle\}$.
b. $S(N)$ is the set of numbers $\langle x,y\rangle$ such that $y = 0$. Note that $S(N)$ is total, whereas for all other $L \in \text{RE}$, $S(L)$ is not total.

PROPOSITION 2.5A $\mathscr{L} \subseteq \text{RE}$ is identifiable if and only if $S(\mathscr{L})$ is identifiable.

Proof Given $\sigma \in \text{SEQ}$, say $\sigma = (x_0, \ldots, x_n)$, define $S(\sigma) = (\langle x_0, 0\rangle, \ldots, \langle x_n, 0\rangle)$. Similarly, if $\sigma = (\langle x_0, y_0\rangle, \ldots, \langle x_n, y_n\rangle)$, define $P(\sigma) = (x_0, \ldots, x_n)$. Let $g, h \in \mathscr{F}$ be such that for all $i \in N$, $W_{g(i)} = S(W_i)$ and $W_{h(i)} = P(W_i)$.

Now suppose that $\mathscr{L} \subseteq \text{RE}$ is identified by $\varphi \in \mathscr{F}$. Let $\psi \in \mathscr{F}$ be such that for all $\sigma \in \text{SEQ}$,

$$\psi(\sigma) = g(\varphi(P(\sigma))).$$

It is clear that ψ identifies $S(\mathscr{L})$.

Similarly, if $\psi \in \mathscr{F}$ identifies $S(\mathscr{L})$, let $\varphi \in \mathscr{F}$ be such that for all $\sigma \in \text{SEQ}$,

$$\varphi(\sigma) = h(\psi(S(\sigma))).$$

Then φ identifies $\mathscr{L}$. □

The technique used in the foregoing proof is important. It might be called "internal simulation." For instance, in the first part, ψ works by simulating the action of φ on a text constructed from the text given to φ.

COROLLARY 2.5A The collection of all single-valued languages is not identifiable.

Corollary 2.5A should be compared with proposition 1.4.3C.

Proposition 2.5A (along with the method of its proof) shows that the collection of single-valued languages presents nothing new from the point of view of identification. In contrast, proposition 1.4.3C shows that the collection of total, single-valued languages has learning-theoretic properties that distinguish it from RE. For this reason, when considering single-valued languages, we shall generally restrict attention to RE_{svt}, the collection of total, single-valued r.e. sets.

What makes RE_{svt} identifiable? Recall from section 1.3.3 that texts do

not, in general, allow the learner to infer directly the nonoccurrence of sentences. In contrast, if t is a text for an unspecified language in RE_{svt}, then for every $x \in N$ there is an $n \in N$ such that examination of $\bar{t}_n$ is sufficient to determine whether or not $x \in \text{rng}(t)$. To see this, suppose that $x = \langle i,j\rangle$. Then some number y occurs in t such that $y = \langle i,k\rangle$ (since $\text{rng}(t)$ is total). As soon as $\langle i,k\rangle$ appears in t, the question "$\langle i,j\rangle \in \text{rng}(t)$?" can be answered, for $\langle i,j\rangle \in \text{rng}(t)$ just in case $j = k$ (since $\text{rng}(t)$ is single-valued). If $j \neq k$, the presence of $\langle i,k\rangle$ in t may be thought of as "indirect negative evidence" for $\langle i,j\rangle$ in t, in the sense discussed by Pinker (1984). In sum, texts for total, single-valued languages offer information about both the presence and the absence of sentences. The learning function h of proposition 1.4.3C exploits this special property of RE_{svt}.

Exercises

***2.5A** Is there a price for self-knowledge? We restrict attention to recursive learning functions. Call $\varphi \in \mathscr{F}^{\text{rec}}$ *Socratic* just in case φ identifies the language $L_\varphi = \{\langle x,y\rangle \mid \varphi(x) = y\}$. (Since $\varphi \in \mathscr{F}^{\text{rec}}$, $L_\varphi \in \text{RE}$.)

a. Specify a collection $\mathscr{L}$ of single-valued languages such that some $\varphi \in \mathscr{F}^{\text{rec}}$ identifies $\mathscr{L}$, but no $\varphi \in \mathscr{F}^{\text{rec}}$ identifies $\mathscr{L} \cup \{L_\varphi\}$. (*Hint:* See exercise 2.2E.) Conclude that (recursive) Socratic learning functions are barred from identifying certain identifiable collections.
b. Prove: Let $\mathscr{L} \subseteq \text{RE}_{\text{svt}}$ be given. Then some $\varphi \in \mathscr{F}^{\text{rec}}$ identifies $\mathscr{L}$ if and only if some Socratic $\varphi \in \mathscr{F}^{\text{rec}}$ identifies $\mathscr{L}$. (*Hint:* Use the recursion theorem, lemma 2.3A.) Philosophize about all this.

3 Learning Theory and Natural Language

We interrupt the formal development of learning theory in order to motivate the technicalities that follow. Specifically we attempt to locate learning-theoretic considerations in the context of theories of the human language faculty. Toward this end section 3.1 presents the perspective that animates this book; it derives from Chomsky (1975) and Wexler and Culicover (1980, ch. 2). Section 3.2 examines several issues that complicate the use of learning theory in linguistics.

3.1 Comparative Grammar

Comparative grammar is the attempt to characterize the class of (biologically possible) natural languages through formal specification of their grammars; a *theory* of comparative grammar is such a specification of some definite collection of languages. Contemporary theories of comparative grammar begin with Chomsky (e.g., 1957, 1965), but there are several different proposals currently under investigation.

Theories of comparative grammar stand in an intimate relation to theories of linguistic development. If anything is certain about natural language, it is this: children can master any natural language in a few years time on the basis of rather casual and unsystematic exposure to it. This fundamental property of natural language can be formulated as a necessary condition on theories of comparative grammar: such a theory is true only if it embraces a collection of languages that is learnable by children.

For this necessary condition to be useful, however, it must be possible to determine whether given collections of languages are learnable by children. How can this information be acquired? Direct experimental approaches are ruled out for obvious reasons. Investigation of existing natural languages is indispensable, since such languages have already been shown to be learnable by children; as revealed by recent studies, much knowledge can be gained by examining even a modest number of languages. We might hope for additional information about learnable languages from the study of children acquiring a first language. Indeed, many relevant findings have emerged from child language research. For example, the child's linguistic environment appears to be devoid of explicit information about the nonsentences of her language (see section 1.3.3). As another example, the rules in a child's immature grammar are not simply a subset of the rules of the adult grammar but appear instead to incorporate distinctive rules that will be abandoned later.

However, such findings do not directly condition theories of comparative grammar. They do not by themselves reveal whether some particular class of languages is accessible to children, nor whether some other particular class lies beyond the limits of child learning. Learning theory may be conceived as an attempt to provide the inferential link between the results of acquisitional studies and theories of comparative grammar. It undertakes to translate empirical findings about language acquisition into information about the kinds of languages accessible to young children. Such information in turn can be used to evaluate theories of comparative grammar.

To fulfill its inferential role, learning theory provides precise construals of concepts generally left informal in studies of child language, notably the four concepts of Section 1.1 as well as the criterion of successful acquisition to which children are thought to conform. Each such specification constitutes a distinctive learning paradigm, as discussed in Section 1.1. The scientifically interesting paradigms are those that best represent the circumstances of actual linguistic development in children. The deductive consequences of such paradigms yield information about the class of possible natural languages. Such information in turn imposes constraints on theories of comparative grammar.

To illustrate, the identification paradigm represents languages as r.e. sets and environments as texts; children are credited with the ability to identify any text for any natural language. If normal linguistic development is correctly construed as a species of identification, then proposition 2.2A yields nonvacuous constraints on theories of comparative grammar; no such theory, for example, could admit as natural some infinite and all finite languages.

Unfortunately identification is far from adequate as a representation of normal linguistic development. Children's linguistic environments, for example, are probably not arbitrary texts for the target language: on the one hand, texts do not allow for the grammatical omissions and ungrammatical intrusions that likely characterize real environments; on the other hand, many texts constitute bizarre orderings of sentences, orderings that are unlikely to participate in normal language acquisition. In addition the identification paradigm provides no information about the special character of the child's learning function. To claim that this latter function is some member of $\mathscr{F}$ is to say essentially nothing at all. Even the criterion of successful learning is open to question because linguistic development does not always culminate in the perfectly accurate, perfectly stable grammar envisioned in the definition of identification.

The defects in the identification paradigm can be remedied only in light of detailed information about children's linguistic development. For the most part, the needed information seems not to be currently available. Consequently we shall not propose a specific model of language acquisition. Rather, the chapters that follow survey a variety of learning paradigms of varying relevance to comparative grammar. The survey, it may be hoped, will suggest questions about linguistic development whose answers can be converted into useful constraints on theories of comparative grammar.

Our survey of learning paradigms occupies parts II and III of this book. Before turning to it, we discuss some potential difficulties associated with the research program just described.

3.2 Learning Theory and Linguistic Development

3.2.1 How Many Grammars for the Young Child?

If $\varphi \in \mathscr{F}$ identifies $t \in \mathscr{T}$, then φ is defined on t (see definition 1.4.1A); thus $\varphi(\bar{t}_n)\downarrow$ for all $n \in N$. This feature of identification will be carried forward through almost all of the paradigms to be studied in this book. Yet it is easy to imagine that newborn infants do not form grammars in response to the first sentence they hear (perhaps: "It's a boy!"); similarly, *bona fide* grammars might be lacking during early stages of linguistic production. The empirical interest of learning theory might seem to be compromised by this possibility.

To respond to this problem, we may adopt a new convention concerning indexes. According to the new convention all indexes are increased by 1, leaving the number 0 without an associated grammar. Zero may then be used to represent any output that does not constitute a grammar. Then for $n \in N$, $\varphi(n) = 0$ implies $\varphi(n)\downarrow$, as before. Plainly, $\mathscr{L} \subseteq \text{RE}$ is identifiable if and only if $\mathscr{L}$ is identifiable under the new convention. The result is that identification of a text t need not be compromised by the failure to conjecture a grammar at early stages of t.

In similar fashion it is possible to envision the following possibility. Children may respond to linguistic input not with one grammar but with a finite array of grammars, each associated with some (rational) subjective probability. To represent this possibility, the numbers put out by learning functions can be interpreted not as r.e. indexes but as codes for such finite arrays, since finite arrays of the sort envisioned are readily coded as single natural numbers. On the other hand, we might simply choose as the child's "official" conjecture at a given moment the grammar assigned highest subjective probability at that moment.

Consider next children growing up in multilingual environments. Such children simultaneously master more than one language and hence convert their current linguistic input into more than one, noncompeting grammatical hypothesis. To represent this situation, we must assume that inputs from different languages are segregated by the child prior to grammatical analysis (perhaps by superficial characteristics of the wave form or the speaker). Linguistic development may then be conceived as the simultaneous application of the same learning function to texts for different languages.

Clearly the general framework of learning theory can be adapted to a wide variety of empirical demands of the kind just considered. Consequently in the sequel we shall not pause to refine our models in these directions; specifically, we shall continue to treat conjectures straightforwardly as (single) r.e. indexes.

3.2.2 Are the Child's Conjectures a Function of Linguistic Input?

As discussed in section 1.3.4, learning functions are conceived as mappings from finite linguistic corpora (represented as members of SEQ) into grammatical hypotheses. It is possible, however, that children's linguistic conjectures depend on more than their linguistic input; that is, the same finite corpus might lead to different conjectures by the same child depending on such nonlinguistic inputs as the physical affection afforded the child that day or the amount of incident sunlight. Put another way, children may not implement any function from finite linguistic corpora into grammatical hypotheses; rather, the domain of the function that produces children's linguistic conjectures might include nonlinguistic elements.

This issue must not be confused with the problem of individual differences. It is possible that different children implement distinct learning functions, but the present question concerns the nature of a single child's function. We shall in fact proceed on the assumption that children are more or less identically endowed with respect to first language acquisition.

The present issue is also independent of the possibility that the child's learning function undergoes maturational change. To see this, let $\psi \in \mathscr{F}$ be considered the maturational successor to $\varphi \in \mathscr{F}$, and let ψ begin its operation at the nth moment of childhood. Then the child may be thought of as implementing the single function $\theta \in \mathscr{F}$ such that for all $\sigma \in \mathrm{SEQ}$,

$$\theta(\sigma) = \begin{cases} \varphi(\sigma), & \text{if } \mathrm{lh}(\sigma) < n, \\ \psi(\sigma), & \text{otherwise.} \end{cases}$$

θ is a function of linguistic input if φ and ψ are such functions. This schema may be refined in several ways, and any number of maturational changes may be envisioned.

Finally, the problem of nonlinguistic inputs to the learning function is not the same as the problem of utterance context. As noted in section 1.3.1, any finite aspect of context may be built into the representation of a sentence. What is at issue here, in contrast, are inputs that play no evident communicative role, such as the child's diet or interaction with pets.

The possibility that the child's grammatical hypotheses are a function of more than just linguistic input can be accommodated in a straightforward way. Specifically, the interpretation of SEQ can be extended to allow both sentences and other kinds of inputs to figure in the finite sequences presented to learning functions. Such extension would require a compensatory change in the definition of successful learning since convergence on a text t to $\mathrm{rng}(t)$ would no longer be appropriate; rather, success would consist in convergence to the linguistic subset of $\mathrm{rng}(t)$.

In practice, such amended definitions seem unmotivated since there is no available information about the role of nonlinguistic inputs in children's grammatical hypotheses, if indeed there is any such role. As a consequence learning theory has developed under the assumption (usually tacit) that the only inputs worth worrying about are linguistic. We shall follow suit.

3.2.3 What Is a Natural Language?

Comparative grammar aims at an illuminating characterization of the class of natural languages. But what independent characterization of this latter class gives content to particular theories of comparative grammar? The question may be put this way: What is a natural language, other than that which is characterized by a true theory of comparative grammar?

Inevitably considerations of learnability enter into any "pretheoretical" specification of the natural languages. Even if we revert to the partly ostensive definition "The natural languages are English, Japanese, Russian, and other languages *like those*," the italicized expression must bear on the

ease of language acquisition if the resulting concept is to have much interest for linguistics. The following formula thus suggests itself:

> A highly expressive linguistic system is *natural* just in case it can be easily acquired by normal human infants in the linguistic environments typically afforded the young.

The role of the qualification "highly expressive" in the foregoing formula is discussed in section 7.1, so we do not consider the matter here. Rather, we examine the remaining concepts, beginning with "normal human infant."

What content can be given to the concept of a nomal infant that does not render the preceding formula a tautology? Plainly it is no help to qualify a child as "normal" just in case he or she is capable of acquiring natural language (easily and in a typical environment). It is equally useless to appeal to majority criteria such as: a language is natural just in case a majority of the world's actual children can acquire it (easily, etc.). The reason is that the world's actual children might all have accidental properties (e.g., the same subtle infection), rendering them inappropriate as the intended standard. What was wanted were normal children, not the possibly unlucky sample actually at hand.

It is tempting to here invoke neurological considerations by stipulating that a child is normal just in case his or her brain meets certain neurophysiological conditions laid down by some successful (and future) neurophysiological theory. The difficulty with this suggestion is that the choice of such neurological conditions must depend partly on information about the normal linguistic capacities of the newborn, for a brain cannot be judged normal if it is incapable of performing the tasks normally assigned to it. And of course invocation of normal capacities leads back to our starting point. Quite similar problems arise if we attempt to identify normal children with those children implementing the "human" learning function (or a "normal" learning function).

Consider next the concept "typical linguistic environment." Majoritarian construals of this idea are ruled out for reasons similar to before. Rather, "typical" must be read as "normal" or "natural." It is of course unhelpful to stipulate that an environment is natural just in case it allows (normal) children to acquire (easily) a natural language. Nor is it admissible to characterize the natural languages as those acquirable (easily, etc.) in some environment or other, for in that case the notion of natural language will vary with our ability to imagine increasingly exotic environments (e.g., environments that modify the brain in "abnormal" ways). We leave it to the reader to formulate parallel concerns with respect to the concept of "easy acquisition."

None of this discussion is intended to suggest that comparative grammar suffers from unique conceptual problems foreign to other sciences. As in other sciences, we must hope for gradual and simultaneous clarification of all the concepts in play. Thus examination of central cases of natural language will constrain our conjectures about the human learning function, which can then be expected to sharpen questions about environments, criteria of successful learning, and, eventually, natural language itself. As in other sciences a natural language will eventually be construed in the terms offered by the most interesting linguistic theory. Within this perspective learning theory may be understood as the study of the deductive constraints that bind together the various concepts discussed earlier. These concepts are thus in no worse shape than comparable concepts in other emerging sciences. Our discussion is intended to show only that they are not in much better shape either.

3.2.4 Idealization

Texts are infinitely long, and convergence takes forever. These features of identification will be generalized to all the paradigms discussed in this book. However, language acquisition is a finite affair, so learning theory (at least as developed here) might seem from the outset to have little bearing on linguistic development and comparative grammar.

Two replies to this objection may be considered. First, although convergence is an infinite process, the onset of convergence occurs only finitely far into an identified text. What is termed "language acquisition" may be taken to be the acquisition of a grammar that is accurate and stable in the face of new inputs from the linguistic environment; such a state is reached at the onset of convergence not at the end. Moreover, although it is true that identification places no bound on the time to convergence, we shall later consider paradigms that do begin to approximate the constraints on time and space under which the acquisition of natural language actually takes place. Further development of the theory in this direction may be possible as more information about children becomes available.

This first reply notwithstanding, convergence involves grammatical stability over infinitely many inputs, and such ideal behavior may seem removed from the reality of linguistic development. We therefore reply, second, that learning theory is best interpreted as relevant to the design of a language acquisition system, not to the resources (either spatial or temporal) made available to the system that implements that design. Analogously, a computer implementing a standard multiplication algorithm is limited to a finite class of calculations whereas the algorithm itself is designed to determine products of arbitrary size. In this light, consider the learning function φ of the three-year-old child. However mortal the child, φ is timeless and eternal, forever three years old in design. Various questions can be raised about φ, for example: What class of languages does φ identify? If comparative grammar is cast as the study of the design of the human language faculty—as abstracted from various features of its implementation—then such questions are central to linguistic theory.

Evidently, the foregoing argument presupposes that a design-implementation distinction can be motivated in the case of human cognitive capacity. Now Kripke (1982), in an exegesis of Wittgenstein (1953), has offered apparently persuasive arguments against the coherence of the predicate "nervous system (or mind) ... represents rule ...." If the latter predicate is indeed incoherent, then not much can be made of the program-hardware distinction invoked above.

We decline the present opportunity to examine Kripke's argument in detail. The issue, after all, is quite general since it bears on all representational theories in cognitive science, in the sense of Fodor (1976). We note only that Kripke's challenge must eventually be faced if cognitive science, and learning theory in particular, are to rest on firm conceptual foundations.

Having done no more than raise some of the conceptual and philosophical complexities surrounding the application of learning theory to the study of natural language, we now return to formal development of the theory itself.