institute *for* logic, language *and* computation

DANIEL OSHERSON, SCOTT WEINSTEIN,
DICK DE JONGH, ERIC MARTIN

Formal Learning Theory[1]

LP-94-08, *received*: June 1994

ILLC Research Report and Technical Notes Series
Series editor: Dick de Jongh

Logic, Semantics and Philosophy of Language (LP) Series, ISSN: 0928-3307

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam
Plantage Muidergracht 24
NL-1018 TV Amsterdam
The Netherlands
e-mail: illc@fwi.uva.nl

1. Draft of chapter in *Handbook 'Logic and Linguistics'*, J. van Benthem and A. ter Meulen (eds.), Elsevier Science Publishers

Formal Learning Theory*

Daniel Osherson
IDIAP

Scott Weinstein
University of Pennsylvania

Dick de Jongh
University of Amsterdam

Eric Martin
IDIAP

June 24, 1994

1 Introduction

The present chapter is devoted to formal models of language acquisition, and of empirical inquiry more generally. We begin by indicating the issues that motivate our study and then describe the scope of the chapter.

1.1 Empirical inquiry

Many people who have reflected about human intellectual development have noticed an apparent disparity. The disparity is between the information available to children about their environment, and the understanding they ultimately achieve about that environment. The former has a sparse and fleeting character whereas the latter is rich and systematic. This is especially so in the case of first language acquisition, as has been pointed out repeatedly.[1] A similar disparity characterizes other tasks of childhood. By an early age the child is expected to master the moral code of his household and community, to assimilate its artistic conventions and its humor, and at the same time to begin to understand the physical principles that shape the material environment. In each case the child is required to convert data of a happenstance character into the understanding (implicit or explicit) that renders his world predictable and intelligible.

Little is known about the mental processes responsible for children's remarkable intellectual achievements. Even elementary questions remain the subject of controversy and inconclusive findings. For example, there is little agreement about whether children use a general-purpose system to induce the varied principles bearing on language, social structure, etc.,

*Research support was provided by the Office of Naval Research under contracts Nos. N00014-87-K-0401 N00014-89-J-1725 and by the Swiss National Science Foundation under grant number 21-32399.91. Correspondence to D. Osherson, IDIAP, C.P. 609, CH-1920 Martigny, Switzerland. Electronic mail: dickdj@fwi.uva.nl, martin@maya.idiap.ch, osherson@maya.idiap.ch, weinstein@cis.upenn.edu

[1]See, for example, [Chomsky, 1975, Matthews, 1984, Hornstein & Lightfoot, 1981]. A review of empirical findings on first language acquisition is available in [Pinker, 1990].

or whether different domains engage special-purpose mechanisms in the mind.[2] Although some suggestive empirical findings are available [Gleitman, 1986, Johnson & Newport, 1989, Newport & Supalla, 1989], the matter still engenders controversy (e.g., [Bickerton, 1981]).

The disparity noted above for intellectual development has also been observed in the acquisition of scientific knowledge by adults. Like the child, scientists typically have limited access to data about the environment, yet are sometimes able to convert this data into theories of astonishing generality and (apparent) veracity. At an abstract level, the inquiries undertaken by child and adult may be conceived as a process of theory elaboration and test. From this perspective, both agents react to available data by formulating hypotheses, evaluating and revising old hypotheses as new data arrive. In the favorable case, the succession of hypotheses stabilizes to an accurate theory that reveals the nature of the surrounding environment. We shall use the term "empirical inquiry" to denote any enterprise that possesses roughly these features.

It is evident that both forms of empirical inquiry — achieved spontaneously in the early years of life, or more methodically later on — are central to human existence and cultural evolution. It is thus no accident that they have been the subject of speculation and inquiry for centuries, and of vigorous research programs within several contemporary disciplines (namely, Psychology, Artificial Intelligence, Statistics and Philosophy). We shall not here attempt to synthesize this vast literature but rather limit ourselves to a single line of investigation that descends from the pioneering studies [Putnam, 1965, Putnam, 1975, Solomonoff, 1964, Solomonoff, 1964, Gold, 1967, Blum & Blum, 1975]. It is this tradition that appears to have had the greatest impact on linguistics, and to a limited extent on epistemology.[3]

Our topic has been named in various ways, often as "Formal Learning Theory" which we adopt here usually without the qualifer "Formal". Central to the theory is the concept of a *paradigm* (or model) of empirical inquiry. The inquiry in question might be that of a child learning language, or of a scientist investigating nature. Every paradigm in the theory has essentially the same stock of component concepts, which we now explain.

1.2 Paradigms

A paradigm offers formal reconstruction of the following concepts, each central to empirical inquiry.

(1) (a) *a theoretically possible reality*
(b) *an intelligible hypothesis about reality*
(c) *the data available about any given reality, were it actual*
(d) *a scientist (or child)*
(e) *successful behavior by a scientist working in a given, possible reality*

The concepts figure in the following picture of scientific inquiry, conceived as a game between Nature and a scientist. First, a class of possible realities is specified in advance; the class is

[2]For discussion, see [Chomsky, 1975, Osherson & Wasow, 1976].

[3]Within linguistics, relevant papers include [Wexler & Culicover, 1980, Borer & Wexler, 1987, Lasnik, 1989, Matthews, 1989, Truscott & Wexler, 1989]. Within epistemology, see, for example, [Kelly & Glymour, 1993], [Earman, 1992, Ch. 9].

known to both players of the game. Nature is conceived as choosing one member from the class, to be the "actual world;" her choice is initially unknown to the scientist. Nature then provides a series of clues about this reality. These clues constitute the data upon which the scientist will base his hypotheses. Each time Nature provides a new clue, the scientist may produce a new hypothesis. The scientist wins the game if there is sufficient guarantee that his successive conjectures will stabilize to an accurate hypothesis about the reality Nature has chosen.

Different paradigms formalize this picture in different ways, resulting in different games. Whether a particular game is winnable depends, among other things, on the breadth of the set of possible realities. Wider sets make successful learning more difficult, to the point of impossibility. The dominant concern of Learning Theory is to formulate an illuminating characterization of the paradigms in which success is achievable.

1.3 Scope of the chapter

Contemporary Learning Theory has two principal branches, which may be termed "recursion theoretic," and "model theoretic." They are distinguished, as indicated, by the tools used to define and study paradigms. The recursion theoretic side of the discipline is older and better developed. The next three sections overview some principal results. A few proofs are lightly sketched, just for "feeling." The others may be found in [Osherson *et al.*, 1986c]. A more complete survey will be available in [Sharma *et al.*, 1995]. Concerns about recursion theoretic modeling are voiced in Section 5, and the alternative perspective is introduced. The subsequent five sections are devoted to Learning Theory from the point of view of model theory. We have chosen to follow one particular line of research, ending with some new results (proofs are given in the appendix). The material presented here is intended to be illustrative of central ideas and concepts; a comprehensive survey is not attempted. More systematic coverage is available in [Sharma *et al.*, 1995].

2 Identification

There is no better introduction to Learning Theory than presentation of its most fundamental paradigm. Such is the goal of the present section, whose essential ideas are due to [Gold, 1967]. To proceed, we consider in turn the components of paradigms listed in (1).

Realities. Possible realities are represented by nonempty, r.e. subsets of non-negative integers. (The non-negative integers are denoted by N in the sequel.) Thinking of such sets as potential natural languages, the paradigm is usually called *language identification,* and the sets themselves "languages." It will be convenient in what follows to drop the "language" qualifier when referring to identification.

Hypotheses. Intelligible hypotheses are the r.e. indices for languages, relative to some background, acceptable ordering of the Turing Machines (see [Machtey & Young, 1978] for "acceptable ordering").

Data. To specify the data that Nature makes available about a given language L, we rely on the following terminology. An ω-sequence of natural numbers is called a *text.* The set of numbers appearing in a text t is denoted *content*(t). Text t is said to be *for* L just in case *content*$(t) = L$. After choosing L as reality, Nature presents the scientist with an arbitrary text for L, that is, an infinite listing of L with no intrusions or omissions. If L has at least two elements, the class of texts for L is uncountable.

Let t be a text for L. The initial finite sequence of length n in t is denoted $t[n]$. $t[n]$ may be thought of as an "evidential position" since it contains all the data about L made available by t at the n^{th} moment of inquiry. The set $\{t[n] \mid n \in N$ and t is a text$\}$ of all evidential positions is denoted *SEQ*. Note that *SEQ* is the set of all finite sequences of natural numbers and hence is recursively isomorphic to N.

Scientists. A "scientist" is any function (not necessarily total or recursive) from *SEQ* to N, where the latter are conceived as r.e. indices. Thus, a scientist is a system that converts its current, evidential position into an hypothesis about the language giving rise to his text.

Success. Success is defined in stages.

(2) DEFINITION: Let scientist Ψ, text t, and $i \in N$ be given.

(a) Ψ *converges on* t to i just in case for all but finitely many $n \in N$, $\Psi(t[n]) = i$.

(b) Ψ *identifies* t just in case there is $i \in N$ such that Ψ converges to i on t, and i is an index for *content*(t).

(c) Ψ *identifies* language L just in case Ψ identifies all the texts for L.

(d) Ψ *identifies* a collection $\mathbf{L}$ of languages just in case Ψ identifies every $L \in \mathbf{L}$. In this case $\mathbf{L}$ is said to be *identifiable.*

Thus, Ψ identifies $\mathbf{L}$ just in case for every text t for any $L \in \mathbf{L}$, Ψ identifies t. Note that any singleton collection of languages is trivially identifiable (by a constant function). Scientists (and children) are challenged only by a wide range of theoretical possibilities.

To illustrate, the collection $\mathbf{F}$ of finite sets is identifiable by Ψ defined this way: For all $\sigma \in SEQ$, $\Psi(\sigma)$ is the smallest index for *content*(σ), where the latter is the set of numbers appearing in σ. $\mathbf{F}$ has the interesting property that no extension is identifiable [Gold, 1967], whereas every other identifiable collection can be extended to another one. The collection $\mathbf{L} = \{N\} \cup \{N - \{x\} \mid x \in N\}$ is also unidentifiable, whereas it is easy to define a scientist that identifies $\mathbf{L} - \{N\}$.

To prove the non-identifiability facts cited above, we rely on the "locking sequence" lemma. Its basic idea is due to [Blum & Blum, 1975].

(3) DEFINITION: Let scientist Ψ, language L, and $\sigma \in SEQ$ be given. σ is a *locking sequence* for Ψ and L just in case:

(a) $\Psi(\sigma)$ is defined; and

(b) for all $\tau \in SEQ$ drawn from L that extend σ, $\Psi(\tau) = \Psi(\sigma)$.

Intuitively, σ locks Ψ onto its conjecture $\Psi(\sigma)$, in the sense that no new data from L can lead Ψ to change its mind.

(4) LEMMA: Let language L and scientist Ψ be such that Ψ identifies L. Then there is a locking sequence σ for Ψ and L. Moreover, $\Psi(\sigma)$ is an index for L.

A proof is given in Section 12.1.

Now suppose that scientist Ψ identifies some infinite language L. By the lemma, let σ be a locking sequence for Ψ and L, and let t be a text that consists of endless repetitions of σ. By the choice of σ, Ψ converges on t to an index i for L. Since L is infinite, i is not for $content(t)$ since the latter is finite. Hence, Ψ fails to identify some text for a finite language, and thus does not identify $\mathbf{F}$. This is enough to show that no scientist identifies a proper extension of $\mathbf{F}$, as noted above. The nonidentifiability of $\{N\} \cup \{N - \{x\} \mid x \in N\}$ is shown similarly.

More generally, Lemma (4) allows us to provide the following characterization of identifiability (see [Osherson *et al.*, 1986c, Sec. 2.4] for the simple proof).

(5) PROPOSITION: [Angluin, 1980] Let collection $\mathbf{L}$ of languages be given. $\mathbf{L}$ is identifiable if and only if for all $L \in \mathbf{L}$ there is finite $D_L \subseteq L$ such that for all $L' \in \mathbf{L}$, if $D_L \subseteq L'$ then $L' \not\subset L$.

3 Remarks about the identification paradigm

Identification evidently provides a highly simplified portrait of first language acquisition and of empirical inquiry generally. Learning theorists have exercised considerable ingenuity in refining and elaborating the basic paradigm in view of more realistic models. Illustrations will be provided in the next section. First it may be useful to comment on a few aspects of the bare paradigm defined above.

3.1 Possible realities as sets of numbers

Limiting possible realities to r.e. subsets of N is mathematically convenient, and has been a feature of much work in Learning Theory.[4] The numbers are to be conceived as codes for objects and events found in scientific or developmental contexts. The details of such coding reflect substantive hypotheses concerning the kind of phonological, semantic, and other information available to children about the ambient language, or about the character of the data that drives scientific research. Unfortunately, mathematical studies of learning often neglect this aspect of formalization, simply starting with N as the base of inquiry. Until Section 6 we shall follow suit.

Some sets of numbers are "single-valued," in the sense of [Rogers, 1967, Sec. 5.7]. By limiting attention to collections of single-valued, r.e. sets, one treats the important problem of synthesizing a computer program from examples of its graph (as in [Shapiro, 1983]). Indeed, there have been more studies of function learning than of pure language learning (see [Sharma *et al.*, 1995]). In view of our present concern with natural language, no more will here be said about function learning (except for a remark in Section 4.9).

[4] An exception is [Kugel, 1977], who drops the r.e. requirement.

3.2 Reliability

The concepts of accuracy and stability are central to identification. Identifying a text t requires the scientist to ultimately issue an index i that enumerates *content*(t) and then to remain with i for the remainder of t, that is, it requires eventual accuracy and stability of the scientist's hypotheses. When we consider collections of languages a third concept arises. To identify collection $\mathbf{L}$, a scientist Ψ must succeed on any text for any member of $\mathbf{L}$. In this sense, Ψ is required to be a reliable agent of inquiry, succeeding not just on a happenstance collection of texts, but on all of them. Being able to reliably stabilize to an accurate conjecture is the hallmark of scientific competence in all of Learning Theory, and alternative paradigms provide varied reconstructions of these concepts. Kindred notions of reliability are studied in epistemology (e.g., [Goldman, 1986, Kornblith, 1985, Pappas, 1979]), which is one reason Learning Theory is considered pertinent to philosophical investigation (as in [Kelly, 1994]).

There is another aspect of successful performance that is pertinent to defining realistic models of language acquisition and of inquiry generally. Discovery should be reasonably rapid. The identification paradigm imposes no requirements in this connection, since successful scientists can begin convergence at any point in a text (and at different points for different texts, even for the same language). However, other paradigms build efficiency into the success criterion (as in [Daley & Smith, 1986]).[5]

One requirement on scientists that is usually *not* imposed by Learning Theory is worth noting. To succeed in identification, the scientist must produce a final, correct conjecture about the contents of the text he is facing. He is not required, however, to "know" that any specific conjecture is final. To see what is at issue, consider the problem of identifying $\mathbf{L} = \{N - \{x\} \mid x \in N\}$. Upon seeing $0, 2, 3, 4, \ldots, 1000$ there are no grounds for confidence in the appealing conjecture $N - \{1\}$ since the next bit of text might contradict this hypothesis. The identifiability of $\mathbf{L}$ does warrant a different kind of confidence, namely, that systematic application of an appropriate guessing rule will eventually lead to an accurate, stable conjecture on any text for a member of $\mathbf{L}$.

Distinguishing these two kinds of confidence allows us to focus on scientific success itself, rather than on the secondary question of warranted belief that success has been obtained. Thus, the fundamental question for Learning Theory is:

> What kind of scientist reliably succeeds on a given class of problems?

rather than:

> What kind of scientist "knows" when it is successful on a given class of problems?

Clarity about this distinction was one of the central insights that led to the mathematical study of empirical discovery (see [Gold, 1967, pp. 465-6]).[6]

---

[5]Efficiency is of paramount concern within the "PAC-learning" approach to inductive inference (see [Anthony & Biggs, 1992]). PAC-learning is less relevant than Formal Learning Theory to language acquisition by children, and is not treated here. For one attempt to relate the two approaches, see [Osherson *et al.*, 1991a].

[6]In "finite learning" scientists are allowed but a single conjecture so their attachment to it can be considered stronger than is the case for identification. See [Jain & Sharma, 1990b] for an illuminating study.

3.3 Comparative grammar

In the linguistic context, possible realities are the languages that children might be called upon to master. Now it seems evident to many linguists (notably, [Chomsky, 1975, Chomsky, 1986]) that children are not genetically prepared to acquire any, arbitrary language on the basis of the kind of casual linguistic exposure typically afforded the young. Instead, a relatively small class **H** of languages may be singled out as "humanly possible" on the basis of their amenability to acquisition by children, and it falls to the science of linguistics to propose a nontrivial description of **H**. Specifically, the discipline known as "comparative grammar" attempts to characterize the class of (biologically possible) natural languages through formal specification of their grammars; and a *theory* of comparative grammar is a specification of some definite collection. Contemporary theories of comparative grammar begin with [Chomsky, 1957, Chomsky, 1965]), but there are several different proposals currently under investigation (see [Wasow, 1989] and J. Higginbotham's chapter in this handbook).

Theories of linguistic development stand in an intimate relation to theories of comparative grammar inasmuch as a theory of comparative grammar is true only if it embraces a collection of languages 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 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 (see [van Riemsdijk & Williams, 1986]).

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 largely devoid of explicit information about the nonsentences of the target language (see [Brown & Hanlon, 1970, Demetras *et al.*, 1986, Hirsh-Pasek *et al.*, 1984, Penner, 1987]). The acquisition process, moreover, is relatively insensitive to the order in which language is addressed to children (see [Newport *et al.*, 1977, Schieffelin & Eisenberg, 1981]). Finally, certain clinical cases suggest that a child's own linguistic productions are not essential to mastery of the incoming language ([Lenneberg, 1967]). These facts lend a modicum of plausibility to the use of texts as a model of the child's linguistic input. Other pertinent findings bear on the character of immature grammar, which appears not to be a simple subset of the rules of adult grammar but rather incorporates distinctive rules that will be abandoned later (see [Pinker, 1990]).

For all their interest, 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 or whether it lies beyond the limits of their 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 assimilable by young children. Such information can in turn be used to evaluate theories of comparative grammar.

To fulfill its inferential role, Learning Theory offers a range of models of language acquisition. The models arise by precisely construing concepts generally left vague in studies of child language, namely, the five concepts listed in (1). The interesting paradigms from the point of view of comparative grammar are those that best represent the circumstances of actual linguistic development in children. The deductive consequences of such models yield information about the class of possible natural languages.

Many of the paradigms investigated within the theory have little relevance to comparative grammar, for example, studies bearing on team-learning [Daley, 1986, Jain & Sharma, 1990b, Pitt, 1989]. On the other hand, considerable effort has been devoted to paradigms which bear on aspects of language acquisition. For purposes of illustration, the next section is devoted to refinements of the Identification paradigm.[7]

4 More refined paradigms

Refinements of identification can alter any or all of the five components of paradigms, (1)a-e. We limit ourselves here to some simple illustrations bearing on the concepts:

- scientist (or child);
- data made available;
- successful inquiry.

More comprehensive surveys are available in [Angluin & Smith, 1983, Osherson *et al.*, 1986c, Sharma *et al.*, 1995]. The latter two references provide proofs for claims made in this section.

4.1 Memory limitation

It seems evident that children have limited memory for the sentences presented to them. Once processed, sentences are likely to be quickly erased from the child's memory. Here we shall consider scientists that undergo similar information loss. The following notation is used. Let $\sigma \in SEQ$ be given (SEQ is defined in Section 2). The result of removing the last member of σ is denoted by σ^- (if $length(\sigma) = 0$, then $\sigma^- = \sigma = \emptyset$). The last member of σ is denoted by σ_{last} (if $length(\sigma) = 0$, then σ_{last} is undefined).

The following definition says that a scientist is memory limited if his current conjecture depends on no more than his last conjecture and the current datum.

(6) DEFINITION: [Wexler & Culicover, 1980] Scientist Ψ is *memory limited* just in case for all $\sigma, \tau \in SEQ$, if $\Psi(\sigma^-) = \Psi(\tau^-)$ and $\sigma_{last} = \tau_{last}$, then $\Psi(\sigma) = \Psi(\tau)$.

Intuitively, a child is memory limited if her conjectures arise from the interaction of the current input sentence with the latest grammar that she has formulated and stored. The stored grammar, of course, may provide information about other sentences seen to date. To illustrate, it is not hard to prove that the class of finite languages is identifiable by memory limited scientist.

[7] For further discussion of the role of Learning Theory in comparative grammar see [Osherson *et al.*, 1984, Wexler & Culicover, 1980]. Other constraints on theories of comparative grammar might be adduced from biological considerations, or facts about language change. See [Lightfoot, 1982] for discussion.

Thus, it is sometimes possible to compensate for memory limitation by retrieving past data from current conjectures. Nonetheless, memory limitation places genuine restrictions on the identifiable collections of languages, as shown by the following proposition.

(7) PROPOSITION: There is an identifiable collection of languages that is not identified by any memory limited scientist.

We give an idea of the proof (for details, see [Osherson *et al.*, 1986c, Prop. 4.4.1B]). Let E be the set of even numbers, and consider the collection $\mathbf{L}$ of languages consisting of:

(a) E,

(b) for every $n \in N$, $\{2n+1\} \cup E$, and

(c) for every $n \in N$, $\{2n+1\} \cup E - \{2n\}$.

It is easy to verify that $\mathbf{L}$ is identifiable without memory limitation. In contrast, suppose that memory limited Ψ identifies E, and let $\sigma \in SEQ$ be a locking sequence for Ψ and E. Pick $n \in N$ such that $2n \notin content(\sigma)$. Then, Ψ will have the same value on σ and σ extended by $2n$. From this point it is not difficult to see that Ψ will fail to identify at least one text for either $\{2n+1\} \cup E$ or $\{2n+1\} \cup E - \{2n\}$. Hence, Ψ does not identify $\mathbf{L}$. As is common in results of this form one may now further establish that there are uncountably many such identifiable classes of languages not identified by any memory limited scientist.

Proposition (7) shows that, compared to the original paradigm, the memory limited model of linguistic development makes a stronger claim about comparative grammar, imposing a more stringent condition on the class of human languages. According to the refined paradigm, the human languages are not just identifiable, but identifiable by a memory limited learner. Of course, this greater stringency represents progress only if children are in fact memory limited in something like the fashion envisioned by Definition (6).

4.2 Fat text

It may be that in the long run every sentence of a given human language will be uttered indefinitely often. What effect would this have on learning?

(8) DEFINITION:

(a) A text t is *fat* just in case for all $x \in content(t)$, $\{n \mid t(n) = x\}$ is infinite.

(b) Let scientist Ψ and collection $\mathbf{L}$ of languages be given. Ψ *identifies* $\mathbf{L}$ *on fat text* just in case for every fat text t for any $L \in \mathbf{L}$, Ψ identifies t. In this case, $\mathbf{L}$ is *identifiable on fat text.*

Thus, every number appearing in a fat text appears infinitely often. It is easy to prove that every identifiable collection $\mathbf{L}$ of languages is identifiable on fat text, and conversely.

Fat text is more interesting in the context of memory limitation. The following proposition shows that the former entirely compensates for the latter.

(9) PROPOSITION: Suppose that collection $\mathbf{L}$ of languages is identifiable. Then some memory limited scientist identifies $\mathbf{L}$ on fat text.

4.3 Computability

The Turing simulability of human thought is a popular hypothesis in Cognitive Science, and the bulk of Learning Theory has focussed on scientists that implement computable functions. Obviously, any collection of languages that is identifiable by computable scientist is identifiable *tout court*. The converse question is settled by the following.

(10) PROPOSITION: Let $\mathcal{S}$ be any countable collection of functions from SEQ to N (conceived as scientists). Then there is an identifiable collection $\mathbf{L}$ of languages such that no member of $\mathcal{S}$ identifies $\mathbf{L}$.

One argument for (10) proceeds by constructing for each $Q \subseteq N$ an identifiable collection $\mathbf{L}_Q$ of languages such that no single scientist can identify two such classes. The proposition then follows from the fact that there are uncountably many subsets of N but only countably many Turing machines. (See [Osherson *et al.*, 1986c, Prop. 4.1A] for details.)

The assumption that children are Turing simulable is thus a substantive hypothesis for comparative grammar inasmuch as it renders unlearnable some otherwise identifiable collections of languages (assuming the empirical fidelity of the other components of the identification paradigm, which is far from obvious). On the other hand, under suitable assumptions of uniform recursivity of the class of languages, the characterization of (ineffective) identifiability offered by (5) can be transformed into a characterization of identifiability witnessed by Turing-computable scientist (see [Angluin, 1980] and for applications [Kapur, 1991, Kapur & Bilardi, 1992]).

It might be thought that Proposition (10) points to a complexity bound on the languages that co-inhabit collections identifiable by computable scientist. However, the following proposition shows that such a bound cannot be formulated in terms of the usual notions of computational complexity, as developed in [Blum, 1967].

(11) PROPOSITION: [Wiehagen, 1978] There is a collection $\mathbf{L}$ of languages with the following properties.

(a) Some computable scientist identifies $\mathbf{L}$.

(b) For every r.e. $S \subseteq N$ there is $L \in \mathbf{L}$ such that S and L differ by only finitely many elements (that is, the symmetric difference of S and L is finite).

One such collection turns out to consist of all languages L whose least member is an index for L. This collection is easily identified (indeed, by a Turing Machine that runs in time linear in the length of the input), and an application of the recursion theorem shows it to satisfy (11)b. This argument is Wiehagen's (see [Osherson *et al.*, 1986c, Prop. 2.3A]).

Once alternative hypotheses about scientists have been defined and investigated it is natural to consider their interaction. We illustrate with the following fact about memory limitation (Definition (6)).

(12) PROPOSITION: There is a collection $\mathbf{L}$ of languages with the following properties.

(a) Some memory limited scientist (not computable) identifies $\mathbf{L}$.

(b) Some computable scientist identifies $\mathbf{L}$.

(c) No computable, memory-limited scientist identifies $\mathbf{L}$.

4.4 Consistency, Conservatism, Prudence

At the intuitive level, learning theorists use the term "strategy" to refer to a policy for choosing hypotheses in the face of data. Formally, a strategy is just a subset of scientists, such as the class of memory-limited scientists. Further illustration is provided by the next definition, which relies on the following notation. The finite set of numbers appearing in $\sigma \in SEQ$ is denoted $content(\sigma)$. If scientist Ψ is defined on σ, then the language hypothesized by Ψ on σ is denoted $W_{\Psi(\sigma)}$ (notation familiar from [Rogers, 1967]).

(13) DEFINITION: Let scientist Ψ be given.

(a) [Angluin, 1980] Ψ is *consistent* just in case for all $\sigma \in SEQ$, $content(\sigma) \subseteq W_{\Psi(\sigma)}$.

(b) [Angluin, 1980] Ψ is *conservative* just in case for all $\sigma \in SEQ$, if $content(\sigma) \subseteq W_{\Psi(\sigma^-)}$ then $\Psi(\sigma) = \Psi(\sigma^-)$.

(c) [Osherson *et al.*, 1982] Ψ is *prudent* just in case for all $\sigma \in SEQ$, if $\Psi(\sigma)$ is defined then Ψ identifies $W_{\Psi(\sigma)}$.

Thus, the conjectures of a consistent scientist always generate the data seen so far. A conservative scientist never abandons a locally successful conjecture. A prudent scientist only conjectures hypotheses for languages he is prepared to learn.

Conservatism has been the focus of considerable interest within linguistics and developmental psycholinguistics.[8] The prudence hypothesis is suggested by "prestorage" models of linguistic development (as in [Chomsky, 1965]). A prestorage model posits an internal list of candidate grammars that coincides exactly with the natural languages; at any moment in language acquisition, the child is assumed to respond to available data by selecting a grammar from the list. Regarding consistency, it is likely not a strategy adopted by children since early grammars are inconsistent with most everything the child hears; on the other hand, consistency is a property of learners that has attracted the attention of epistemologists (e.g., [Juhl, 1993, Kelly, 1994]).

Consistency and conservatism are substantive strategies in the following sense.

(14) PROPOSITION:

(a) There is a collection of languages that is identifiable by computable scientist but by no consistent, computable scientist.[9]

(b) [Angluin, 1980] There is a collection of languages that is identifiable by computable scientist but by no conservative, computable scientist.[10]

[8] See [Berwick, 1986, Baker & McCarthy, 1981, Mazurkewich & White, 1984, Pinker, 1989, Pinker, 1990].

[9] For more information about consistency and cognate notions, see [Fulk, 1988].

[10] See [Kinber, 1994] for thorough analysis of conservatism and related concepts.

In contrast, we have the following fact about prudence.

(15) PROPOSITION: [Fulk, 1990] Suppose that collection $\mathbf{L}$ of languages can be identified by computable scientist. Then $\mathbf{L}$ can be identified by computable, prudent scientist.

Indeed, the prudent scientist can be constructed uniformly from an index for the original one [Kurtz & Royer, 1988]. Fulk's proof proceeds by showing that every class of languages identified by a computable scientist can be extended to a similarly identifiable collection with an r.e. index set. Proposition (15) then follows easily (see [Osherson *et al.*, 1986c, Lemmas 4.3.4A,B]).

4.5 Noisy and incomplete texts

Although it appears that children's linguistic environments are largely free of grammatical error [Newport *et al.*, 1977], imperfections of two sorts are bound to arise. On the one hand, ungrammatical strings might find their way into the corpus; on the other hand, certain grammatical strings might be systematically withheld. Texts with simple forms of these defects may be defined as follows.

(16) DEFINITION: Let language L and text t be given.

- (a) t is a *noisy* text for L just in case there is finite $D \subset N$ such that t is an (ordinary) text for $L \cup D$.
- (b) t is an *incomplete* text for L just in case there is finite $D \subset N$ such that t is an (ordinary) text for $L - D$.
- (c) Scientist Ψ identifies L *on noisy text* just in case for every noisy text t for L, Ψ converges on t to an index for L. Ψ identifies collection $\mathbf{L}$ of languages *on noisy text* just in case Ψ identifies every $L \in \mathbf{L}$ on noisy text.
- (d) Scientist Ψ identifies L *on incomplete text* just in case for every incomplete text t for L, Ψ converges on t to an index for L. Ψ identifies collection $\mathbf{L}$ of languages *on incomplete text* just in case Ψ identifies every $L \in \mathbf{L}$ on incomplete text.

It is easy to see that noise and incompletion interfere with learning languages differing only finitely from each other. A more substantial fact is the following.

(17) PROPOSITION: There is a collection $\mathbf{L}$ of languages with the following properties.

- (a) Every $L \in \mathbf{L}$ is infinite.
- (b) Every distinct pair of languages in $\mathbf{L}$ is disjoint.
- (c) Some computable scientist identifies $\mathbf{L}$ (on ordinary text).
- (d) No computable scientist identifies $\mathbf{L}$ on noisy text.

A parallel fact holds for incompletion. Indeed, it is shown in [Fulk *et al.*, 1992, Theorem 1] that incompletion is substantially more disruptive for identification than is noise.

4.6 Exact identification

The dictum that natural languages are learnable by children (via casual exposure, etc.) has a converse, namely, that nonnatural languages are not learnable. We are thus led to consider a variant of identification in which successfully learning collection **L** entails identifying **L** and no more. But a complication arises. It may be that certain degenerate languages (e.g., containing but a single word) can be learned by children, even though we do not wish to classify them as natural.

There are findings to suggest, however, that children are not inclined to learn profoundly inexpressive languages. Some of the evidence comes from studies of children raised in pidgin dialects [Sankoff & Brown, 1976]; other work involves the linguistic development of sensorily deprived children [Feldman & Goldin-Meadow, 1978, Landau & Gleitman, 1985]. If we accept the thesis that learnability implies expressiveness, then it is appropriate to define the natural languages as exactly the collection of learnable languages.

Within Learning Theory these ideas give rise to the following definition.

(18) DEFINITION: [Osherson & Weinstein, 1982a] Let scientist Ψ and collection **L** be given. Ψ identifies **L** *exactly* just in case Ψ identifies **L** and identifies no proper superset of **L**.

The requirement of exact identification interacts with hypotheses about strategies. This is illustrated by comparing Proposition (15) with the following.

(19) PROPOSITION: There is a collection **L** of languages with the following properties.

(a) Some computable scientist exactly identifies **L**,

(b) No prudent, computable scientist exactly identifies **L**.

More generally, exact identifiability by computable scientist is possible only in the circumstances described below.

(20) PROPOSITION: Let collection **L** of languages be given. Some computable scientist exactly identifies **L** if and only if **L** is Π^1_1 indexable and some computable scientist identifies **L**.

The Π^1_1 indexability of **L** here means that there is a Π^1_1 subset of N that holds indexes for just the members of **L**. We note that (19) is a corollary to (20). For, there are computably identifiable, properly Π^1_1 collections of languages whereas any collection that is identified by prudent, computable scientist is r.e. indexable. (See [Osherson *et al.*, 1986c, Sec. 7] for discussion.)

4.7 Efficiency

First language acquisition by children has struck many observers as remarkably rapid.[11] It is thus pertinent to examine paradigms in which success requires efficient use of data. To define a simple paradigm of this character, we use the following terminology. Let scientist Ψ, text t, and

[11] But not everyone. See [Putnam, 1980].

$n \in N$ be given. Suppose that Ψ converges on t to index $i \in N$. Then n is called the *convergence point* for Ψ on t just in case n is smallest such that Ψ conjectures i on all initial segments of t of length n or greater. If Ψ does not converge on t we take the convergence point to be ∞.

(21) DEFINITION: [Gold, 1967] Let scientists Ψ_0 and Ψ_1, and collection $\mathbf{L}$ of languages be given.

(a) Ψ_0 identifies $\mathbf{L}$ *strictly faster* than Ψ_1 just in case:

i. both Ψ_0 and Ψ_1 identify $\mathbf{L}$;

ii. for every text t for every $L \in \mathbf{L}$, the convergence point for Ψ_0 on t is no greater than that for Ψ_1 on t;

iii. for some text t for some $L \in \mathbf{L}$, the convergence point for Ψ_0 on t is smaller than that for Ψ_1 on t.

(b) Ψ_0 identifies $\mathbf{L}$ *efficiently* just in case Ψ_0 identifies $\mathbf{L}$, and no scientist Ψ_1 identifies $\mathbf{L}$ strictly faster than Ψ_0.

The next proposition shows that the three strategies examined in Section 4.4 guarantee efficient learning.

(22) PROPOSITION: Suppose that scientist Ψ identifies collection $\mathbf{L}$ of languages. If Ψ is consistent, conservative and prudent then Ψ identifies $\mathbf{L}$ efficiently.

The preceding proposition can be used to show that in the absence of computability constraints, efficiency imposes no restriction on identification (see [Osherson *et al.*, 1986c, Sec. 4.5.1]). In contrast, the work of computable scientists cannot always be delegated to efficient, computable ones.

(23) PROPOSITION: There is a collection $\mathbf{L}$ of languages with the following properties.

(a) Some computable scientist identifies $\mathbf{L}$.

(b) For every computable scientist Ψ that identifies $\mathbf{L}$ there is a computable scientist that identifies $\mathbf{L}$ strictly faster than Ψ.

A rough idea of the proof may be given as follows (see [Osherson *et al.*, 1986c, Prop 8.2.3A] for details). Suppose that $Q \subset N$ is an r.e., nonrecursive set, and that Ψ's speed is aided by quickly deciding whether $n \in N$ belongs to Q. Then Ψ cannot do this for at least one n since otherwise Q would be recursive. Hence, there is a scientist strictly faster than Ψ which has built-in information about this n but which otherwise behaves like Ψ.

4.8 Stability and accuracy liberalized

Identification proposes strict criteria of hypothesis stability and accuracy (in the sense of Section 3.2), and many liberalizations have been examined. For example, weaker criteria of stability might allow successful learners to switch indefinitely often among indices for the same language, or alternatively, to cycle among some finite set of them [Osherson & Weinstein, 1982b, Jain *et al.*, 1989]. Weaker criteria of accuracy might allow a finite number of errors into the final conjecture [Case & Smith, 1983], or else allow the final conjecture to "approximate" the target in a variety of senses [Fulk & Jain, 1992, Royer, 1986]. These and other liberalizations have been studied extensively, both separately and in combination. For a review of findings, see [Sharma *et al.*, 1995].

4.9 Identifying the child's program for language acquisition

Whereas the child's task is to discover a grammar for the ambient language, the task of developmental psycholinguists is to discover the mental program animating the child's efforts. By focusing on the child's learning program rather than on what it learns, we may attempt to define paradigms that illuminate the prospects for success in discovering the mechanisms of first language acquisition. In this case the learner is the psycholinguist and her data may be conceived as the graph of the acquisition function implemented by the child. Successful inquiry consists of converging on the graph to an index for the child's learning function. A less stringent requirement is convergence to a program that identifies at least as many languages as children do, irrespective of its similarity to the child's method. This latter success criterion is called "weak delimitation."

We would like to know how wide a class of potential children can be identified or weakly delimited. If the class is narrow, there may be no reliable means of investigating first-language acquisition. Success in psycholinguistics would depend in this case upon the fortuitous circumstance that the child's learning function falls into the small class of possibilities for which our scientific methods are adapted.

In [Osherson & Weinstein, in pressb] it is shown that some narrow classes of potential children can be neither identified nor weakly delimited. One such class consists of just those children that identify less than three, nonempty languages, none of them finite.

5 The need for complementary approaches

A quarter century of research within Formal Learning Theory has provided suggestive findings for both epistemology and linguistics. It seems fair to say, however, that its impact on the latter discipline has as yet been meager, despite efforts to confront theories of comparative grammar with results about learning (as in [Berwick, 1986, Osherson *et al.*, 1984, Truscott & Wexler, 1989, Wexler & Culicover, 1980]). One reason for the lack of interaction is the abstract character of learning theoretic results. Indeed, the majority of findings remain true under recursive permutation of N, and hence have little to do with the grammatical structure of natural language.

A more recent tradition of research on learning shows greater promise in this regard. For example, [Shinohara, 1990] considers languages defined via elementary formal systems (EFS's) in the sense of [Smullyan, 1961]. He proves that for any $n \in N$, the class of languages definable by length-bounded EFS's with at most n axioms is computably identifiable. From this it follows that for any $n \in N$, the class of languages with context-sensitive grammars of at most n rules is similarly identifiable. Another notable finding is due to [Kanazawa, 1993]. He shows that the class of classical categorial grammars assigning at most k types to each symbol is identifiable by computable scientist in the sense of Definition (2), above. As Kanazawa notes, it follows that the entire class of context-free languages is similarly learnable, provided that texts are enriched with information about the type-ambiguity of each symbol. (For further results, see [Kanazawa, 1994].)

Results like the foregoing are of potentially greater interest to linguistic theory than those bearing on arbitrary r.e. sets. However, research in the new tradition has yet to investigate the special character of children's learning, e.g., its memory-limitation and resistance to noise. These are just the topics given greatest attention in the older literature.

To understand a second reason for Learning Theory's lack of impact on linguistics, let us recall that comparative grammar is supposed to contribute to the theory of innate ideas. In particular, the universal elements of grammar, invariant across natural languages, correspond to what the prelinguistic child already knows about the language into which he is plunged. Extensive debate has arisen about the form in which such knowledge might be lodged in the infant's mind — and even whether it should be called "knowledge" at all, instead of simply "predisposition" (see, for example, [Chomsky, 1975, Matthews, 1984, Putnam, 1967, Stich, 1978]). To address the issue squarely, let us conceive of the child's innate preparation to learn language as a prestored message that characterizes the class of potential natural languages. Then it is difficult to locate this message within the learning paradigms of the Putnam/Gold/Solomonoff tradition. There are just classes of languages in play, under no particular description. Given specific assumptions about data-presentation and so on, either the child can learn the languages or not. There is no innate starting point in sight.[12]

To remedy this shortcoming, some recent paradigms have conceived of innate knowledge as a first-order theory in a countable language (e.g., [Osherson *et al.*, 1991b, Osherson *et al.*, 1992]). In the usual case, the innate theory is not complete; otherwise, there is nothing to learn and there would be no linguistic variation across cultures. So the child's task is to extend the innate theory via new axioms that are true of the particular language spoken in his environment. Consequently, these paradigms consider a single sentence in the language of the original theory, and ask what sort of learning device could determine the truth-value of the sentence by examining data from the environment. The environment is assumed to be consistent with the child's background theory, which thus serves as prior information about the range of theoretical possibilities.

The remainder of the chapter provides details about this approach. To keep the discussion manageable, it is limited to a single strand of inquiry, leaving several relevant studies aside (e.g., [Glymour & Kelly, 1989, Kelly & Glymour, 1993]). The work to be discussed was stimulated by the seminal papers [Glymour, 1985, Shapiro, 1981, Shapiro, 1991].

We proceed as follows. Background ontology and basic concepts occupy Section 6. An elementary but fundamental paradigm is described in Section 7 and some basic facts presented. More sophisticated paradigms are advanced in Sections 8 and 9. Their relevance to first language acquisition is taken up in Section 10. Unless noted otherwise, verification of examples and proofs of propositions are given in the appendix to this chapter.

[12] A preliminary attempt to communicate "starting points" to learners within a recursion theoretic framework is reported in [Osherson *et al.*, 1988].

6 Ontology and basic concepts

6.1 Overview

The paradigms in the remainder of this chapter are embedded in a first-order logical framework. By this is meant that the "possible worlds" in which the scientist might find herself are represented by relational structures for a first-order language. Moreover, the hypotheses that the scientist advances about her world are limited to sentences drawn from the same language. Generalizations are of course possible (as in [Kelly & Glymour, 1992, Osherson & Weinstein, 1989a], for example), but our aim here is to exhibit significant results within the simplest framework possible.

6.2 Language, structures, assignments

We fix a countably infinite collection $\mathbf{D}$ of individuals $\mathbf{d}_0, \mathbf{d}_1 \ldots$. $\mathbf{D}$ is the domain of all structures to be considered in the sequel. In particular, given a set T of first-order sentences, $mod(T)$ denotes the class of structures with domain $\mathbf{D}$ that satisfy T. The exclusion of finite models from the remainder of the discussion is only for convenience. In contrast, the exclusion of uncountable models is necessary to avoid unresolved conceptual questions (see [Osherson & Weinstein, 1986, Section 6.1]).

By a "$\mathbf{D}$-sequence" is meant an ω-sequence onto $\mathbf{D}$ (i.e., with range equal to all of $\mathbf{D}$). Given $\mathbf{D}$-sequence d and $i \in N$, d_i denotes the i^{th} member of d, and $d[i]$ denotes the initial segment of length i in d. The set $\{d[i] \mid d \text{ is a } \mathbf{D}\text{-sequence and } i \in N\}$ of all finite initial segments of $\mathbf{D}$-sequences is denoted $\mathbf{D}^{<\omega}$.

We also fix a language $\mathcal{L}$ with a countable set VAR $= \{v_i \mid i \in N\}$ of variables. The vocabulary of $\mathcal{L}$ is assumed to be finite and include only constants and relation symbols (including identity).[13] The sets of $\mathcal{L}$-formulas and $\mathcal{L}$-sentences are denoted by $\mathcal{L}_{form}$ and $\mathcal{L}_{sen}$, respectively. The set of free variables occurring in $\varphi \in \mathcal{L}_{form}$ is denoted $var(\varphi)$. We use BAS to denote the set of basic formulas, that is, the subset of $\mathcal{L}_{form}$ consisting of atomic formulas and negations thereof.

A $\mathbf{D}$-sequence d will be used to assign objects from $\mathbf{D}$ to variables in VAR. In particular, for every $i \in N$, $d(v_i) = d_i$. Similarly, the finite sequence $\bar{d} = (d_0, \ldots, d_n) \in \mathbf{D}^{n+1}$ corresponds to the finite assignment $\{(v_0, d_0), \ldots, (v_n, d_n)\}$. By $domain(\bar{d})$ is meant the set of variables that $\bar{d}$ interprets, i.e., $\{v_i \in \text{VAR} \mid i < length(\bar{d})\}$.

6.3 Environments

(24) DEFINITION: Let structure $\mathcal{S}$ and $\mathbf{D}$-sequence d be given. By the *environment* for $\mathcal{S}$ and d is meant the ω-sequence e such that for all $i \in N$, $e_i = \{\beta \in BAS \mid var(\beta) \subseteq domain(d[i]) \text{ and } \mathcal{S} \models \beta[d[i]]\}$. An *environment* for $\mathcal{S}$ is an environment for $\mathcal{S}$ and d, for some $\mathbf{D}$-sequence d. An *environment* is an environment for some structure.

[13] The exclusion of function symbols is for convenience only. Their presence would slightly complicate the definition of environments, below.

Thus, an environment is a sequence of ever-more-inclusive, finite, consistent sets of basic formulas. (The sets are finite by our choice of $\mathcal{L}$.) It is as if Nature chooses elements from $\mathbf{D}$ one by one, and after each selection tells us everything she can about the new element and its relation to all the previously chosen elements. For example, suppose that the predicates of $\mathcal{L}$ are $\{=, R\}$, and that structure $\mathcal{S}$ interprets R as $\{(\mathbf{d}_i, \mathbf{d}_j) \mid i < j\}$. If $\mathbf{D}$-sequence d is $\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2 \ldots$ then the environment for $\mathcal{S}$ and d begins this way:

$$\left\{ \begin{array}{c} v_0 = v_0 \\ \neg Rv_0v_0 \end{array} \right\} \left\{ \begin{array}{cc} v_0 = v_0 & v_1 \neq v_0 \\ v_1 = v_1 & Rv_0v_1 \\ v_0 \neq v_1 & \neg Rv_1v_0 \\ \neg Rv_0v_0 & \neg Rv_1v_1 \end{array} \right\} \left\{ \begin{array}{ccccc} v_0 = v_0 & v_2 = v_2 & v_2 \neq v_1 & Rv_0v_1 & \neg Rv_1v_1 \\ v_1 = v_1 & v_0 \neq v_2 & \neg Rv_1v_0 & Rv_0v_2 & \neg Rv_2v_2 \\ v_0 \neq v_1 & v_2 \neq v_0 & \neg Rv_2v_0 & Rv_1v_2 & \\ v_1 \neq v_0 & v_1 \neq v_2 & \neg Rv_2v_1 & \neg Rv_0v_0 & \end{array} \right\} \cdots$$

The following lemma is straightforward (a proof appears in [Osherson & Weinstein, 1986]).

(25) LEMMA: Let environment e and structures $\mathcal{S}$ and $\mathcal{U}$ be given. If e is for both $\mathcal{S}$ and $\mathcal{U}$ then $\mathcal{S}$ and $\mathcal{U}$ are isomorphic.

6.4 Scientists

The finite segment of length i in environment e is denoted $e[i]$, and the set $\{e[i] \mid e$ is an environment and $i \in N\}$ is denoted *SEQ* (there is no risk of confusion with our previous use of *SEQ* in Section 2). Since $\mathcal{L}$ is a finite relational language, *SEQ* is a collection of finite sequences of finite subsets of a fixed countable set; hence, *SEQ* is countable.

A (formal) scientist is defined to be any function from *SEQ* to $\mathcal{L}_{sen}$. According to this conception, scientists examine the data embodied in finite initial segments of environments, and emit hypotheses about the underlying structure in the guise of first-order sentences.

(26) DEFINITION: Let $\theta \in \mathcal{L}_{sen}$, environment e, and scientist Ψ be given. Ψ *converges* on e to θ just in case $\Psi(e[i]) = \theta$ for all but finitely many $i \in N$.

6.5 Solvability for environments

To succeed in a given environment, we require the scientist's hypotheses to stabilize to a single, true, interesting sentence. The idea of stabilization is defined by (26), above. Rather than attempt to formalize the concept of "interesting sentence," we leave it as a parameter in the definition of scientific success. The parameter takes the form of a subset X of sentences, which count as the interesting ones.

(27) DEFINITION:

Let $\mathrm{X} \subseteq \mathcal{L}_{sen}$, scientist Ψ and structure $\mathcal{S}$ be given. Suppose that environment e is for $\mathcal{S}$. Then Ψ X-*solves* e just in case there is $\theta \in \mathrm{X}$ such that:

(a) Ψ converges on e to θ, and

(b) $\mathcal{S} \models \theta$.

It is Lemma (25) that renders clause (b) unambiguous: up to isomorphism, $\mathcal{S}$ is the unique structure for which e is an environment.

(28) EXAMPLE: For $\theta \in \mathcal{L}_{sen}$, let $X = \{\theta, \neg\theta\}$. Then, scientist Ψ X-solves environment e for structure $\mathcal{S}$ just in case Ψ converges on e to whichever of θ, $\neg\theta$ is true in $\mathcal{S}$. This choice of X yields the paradigm of "truth-detection," analyzed in [Glymour & Kelly, 1989, Osherson *et al.*, 1991b].

Other choices of X are discussed in [Osherson & Weinstein, in pressa, Osherson *et al.*, 1992].

6.6 Solvability for structures

All of the paradigms discussed below share the foregoing apparatus. They differ only in the definition given to the idea of solving a given structure $\mathcal{S}$. In each case a scientist will be credited with solving $\mathcal{S}$ if she solves enough environments for $\mathcal{S}$, but the paradigms differ in their interpretation of "enough." The first (and simplest) paradigm conceives the matter in absolute terms: To solve $\mathcal{S}$ the scientist must be able to solve all of its environments. Subsequent paradigms offer probabilistic conceptions.

A scientist Ψ solves a collection $\mathcal{K}$ of structures just in case Ψ solves all the structures in $\mathcal{K}$. This is a constant feature of our paradigms, regardless of how the solution of individual structures is defined. Of particular interest is the case of elementary classes of structures, picked out by a first-order theory. The results discussed below bear principally on this case.

6.7 Relation to language acquisition

Let us relate the concepts discussed above to the child's acquisition of a first language.

The collection X of sentences represents alternative, finitely axiomatized theories of some circumscribed linguistic realm, for example, well-formedness or pragmatic force. Each member of X provides an adequate description of a potential human language (relative to the realm in question). The description is "adequate" in the sense of representing the implicit knowledge accessible to mature speakers. The child's task is to find a member of X that is true of the particular language presented to him.

The class $\mathcal{K}$ of structures embodies the range of linguistic realities for which children are genetically prepared. These realities are the "human" or "natural" ones, in the terms of Section 3.3. If $\mathcal{K}$ is elementary, then the child is assumed competent for any linguistic situation that satisfies a certain theory. The theory can thus be conceived as a component of Universal Grammar, embodying linguistic information available innately to the child at the start of language acquistion.

Environments represent the linguistic data from which a theory can be inferred. In this perspective, **D** might consist of vocalic events (perhaps with associated context) which are classified by the predicates of $\mathcal{L}$. For example, if the theories in X bear on pragmatic force, then predicates might code the intonational contours of utterances, the apparent emotional state of the speaker, etc. Note that environments give direct access to "negative data," whereas this is often assumed not to be a feature of linguistic input to children (see the discussion in Section 3.3, above). To exclude negative data from environments it suffices to restrict their content to atomic formulas, suppressing basic formulas containing negations. We have not adopted this convention since it is unclear whether negative evidence is lacking in linguistic realms other than syntax; in learning the semantics of quantifiers, for example, negative feedback might be available from the failure to communicate an intended meaning. In any event, it remains to determine how well our theorems transfer to the case of "positive environments."

Formal scientists play the role of children. Their mission is to stabilize to a true theory drawn from X. In the model of Section 7 it will be assumed that children achieve such stability with perfect reliability, i.e., no matter how the data are presented. The models of Section 8 and 9 admit the possibility that language acquisition fails when data are presented in an unlikely order.

Suppose that we've established a linguistic realm of interest (e.g., well-formedness). Suppose furthermore that X holds the kind of theories achieved by adults for that realm. Then, a nontrivial property can be attributed to the class of natural languages, namely, X-solvability in the relevant sense. The paradigms now presented provide alternative definitions of X-solvability.

7 First paradigm: Absolute solvability

The idea of solving an environment was formulated in Definition (27) above. To solve a structure, our first paradigm requires the scientist to solve all of its environments. Subsequent paradigms adopt a probabilistic stance.

7.1 Solving arbitrary collections of structures

(29) Definition: Let $X \subseteq \mathcal{L}_{sen}$ and scientist Ψ be given.

- (a) Ψ X-*solves* structure $\mathcal{S}$ just in case Ψ X-solves every environment for $\mathcal{S}$.
- (b) Ψ X-*solves* collection $\mathcal{K}$ of structures just in case Ψ X-solves every $\mathcal{S} \in \mathcal{K}$. In this case, $\mathcal{K}$ is said to be X-*solvable.*

For the examples to follow, we suppose that $\mathcal{L}$ is limited to a sole binary relation symbol R (plus identity).

(30) Example: Let $X = \{\theta, \neg\theta\}$ for $\theta = \forall x \exists y Rxy$ ("there is no greatest point"). We describe the extensions of R in a collection $\mathcal{K} = \{\mathcal{S}_j \mid j \in N\}$. $R^{\mathcal{S}_0}$ is the successor function $\{(\mathbf{d}_i, \mathbf{d}_{i+1}) \mid i \in N\}$. For $j > 0$, $R^{\mathcal{S}_j}$ is $\{(\mathbf{d}_i, \mathbf{d}_{i+1}) \mid i < j\}$. Then $\mathcal{K}$ is not X-solvable.

(31) Example: Let $\mathcal{K}$ be as defined in Example (30). Given $n \in N - \{0\}$, let $\theta_n =$

$$\exists x_1 \ldots x_{n+1}(Rx_1x_2 \ \wedge \ldots \wedge \ Rx_nx_{n+1} \ \wedge \ \forall y \neg Ryx_1 \ \wedge \ \forall z \neg Rx_{n+1}z)$$

i.e., there is an R-chain of length exactly n. Then, for all $n \in N - \{0\}$, $\mathcal{K}$ is X_n-solvable, where $X_n = \{\theta_n, \neg\theta_n\}$. The simple proof is left for the reader.

(32) Example: Let T be the theory of linear orders (with respect to R). Let $\Lambda = \exists x \forall y Rxy$ ("there is a least point"), $\Gamma = \exists x \forall y Ryx$ ("there is a greatest point"), and $X = \{\Lambda, \neg\Lambda\}$. Then $mod(\ T \cup \{\ \Lambda \vee \Gamma,\ \neg(\Lambda \wedge \Gamma)\ \}\ )$ is X-solvable whereas $mod(T)$ is not.

For verification of Example (32), see [Osherson *et al.*, 1991b, Example 5]. Additional examples are given in [Osherson & Weinstein, 1989b, Osherson & Weinstein, in pressa].

Example (32) reveals that inductive inference within our paradigm does not amount to "waiting for deduction to work." For, no $\sigma \in SEQ$ implies either Λ or $\neg\Lambda$ in the models of $T \cup \{\ \Lambda \vee \Gamma,\ \neg(\Lambda \wedge \Gamma)\ \}$. The latter class is nonetheless $\{\Lambda, \neg\Lambda\}$-solvable.

7.2 Solving elementary classes of structures

The theory of solvability has a simple character when limited to first-order definable classes of structures (as in Example (32), above). The theory defining such a class may be conceived as a scientific "starting point" since it embodies all the prior information that is available about a potential environment. In this case there is a computable learning method that is optimal, even compared to methods embodied by noncomputable scientists. We state the matter precisely in the following proposition (whose formulation presupposes familiarity with the arithmetical hierarchy and in particular with the notion of a Σ^0_2 subset of $\mathcal{L}_{sen}$).

(33) Proposition: Suppose that $X \subseteq \mathcal{L}_{sen}$ is Σ^0_2. Then there is an oracle machine M such that for all $T \subseteq \mathcal{L}_{sen}$, if $mod(T)$ is X-solvable, then M^T X-solves $mod(T)$.

The proposition follows immediately from the following lemmas. Their statement requires a preliminary definition, along with the following notation: $\varphi \in \mathcal{L}_{sen}$ will be called "$\exists\forall$" if it is existential-universal in form; either or both sets of quantifiers may be null.

(34) Definition: Let $X \subseteq \mathcal{L}_{sen}$ and $T \subseteq \mathcal{L}_{sen}$ be given. X is *confirmable* in T just in case for all $\mathcal{S} \in mod(T)$ there is $\varphi \in \mathcal{L}_{sen}$ such that:

(a) φ is $\exists\forall$,

(b) $\mathcal{S} \models \varphi$, and

(c) for some $\theta \in X$, $T \cup \{\varphi\} \models \theta$.

(35) Lemma: Let a Σ^0_2 subset X of $\mathcal{L}_{sen}$ be given. Then there is an oracle machine M such that for all $T \subseteq \mathcal{L}_{sen}$, if X is confirmable in T then M^T X-solves $mod(T)$.

(36) Lemma: Let $X \subseteq \mathcal{L}_{sen}$ be given. For all $T \subseteq \mathcal{L}_{sen}$, if $mod(T)$ is X-solvable, then X is confirmable in T.

Lemma (35) is an exercise in "dovetailing" and Σ^0_2-programming, some of the basic ideas already appearing in [Gold, 1965, Putnam, 1965]. A complete proof in a closely related paradigm is given in [Osherson & Weinstein, in pressa]. We do not repeat it here. (Lemma (36) is proved in Section 12.5, below.)

In [Osherson *et al.*, 1991b] the following corollary is derived from Lemma (36) and a weaker version of Lemma (35).

(37) COROLLARY: Let $\theta \in \mathcal{L}_{sen}$ and $T \subseteq \mathcal{L}_{sen}$ be given. Then $mod(T)$ is $\{\theta, \neg\theta\}$-solvable if and only if both θ and $\neg\theta$ are equivalent over T to existential-universal sentences.

As an immediate consequence of Corollary (37) and [Chang & Keisler, 1977, Theorem 3.1.16], we obtain the following fact, demonstrated independently in [Kelly, 1994] (cited in [Earman, 1992, Chapter 9]).

(38) COROLLARY: Let $\theta \in \mathcal{L}_{sen}$ and $T \subseteq \mathcal{L}_{sen}$ be given. Then $mod(T)$ is $\{\theta, \neg\theta\}$-solvable if and only if θ is equivalent over T to a Boolean combination of existential sentences.

We note in passing that Proposition (33) can be extended to no regular logic stronger than the predicate calculus which meets the Löwenheim-Skolem condition. See [Osherson *et al.*, 1991b, Section 4].

8 Second paradigm: Probabilistic solvability

In the present section and the next we conceive of environments as created by a stochastic process. In particular, the entities in our universal domain $\mathbf{D}$ are assumed to be delivered for inspection via independent, identically distributed sampling according to a probability law which may be unknown to the scientist. The associated paradigm measures successful performance in probabilistic rather than all-or-none fashion, and thus differs from most earlier investigations of scientific discovery within a model-theoretic context. It also takes a different approach than that offered in [Gaifman & Snir, 1982] inasmuch as probabilities are attached to the countable set $\mathbf{D}$ rather than to uncountable classes of structures. Within the recursion-theoretic literature on inductive inference, related paradigms are treated by [Angluin, 1988] and [Osherson *et al.*, 1986b, Ch. 10.5].

The core idea of our paradigm is to allow scientists to fail on "small" sets of environments, namely, of measure 0. It will be seen that such liberalization has no effect on the solvability of elementary classes of structures. Moreover, the universal machine for absolute solvability is universal in the present setting as well.

8.1 Measures over environments

The class of all positive probability distributions over $\mathbf{D}$ is denoted $\mathbf{P}$. ($P \in \mathbf{P}$ is positive just in case $P(d) > 0$ for all $d \in \mathbf{D}$.) Given $P \in \mathbf{P}$, we extend P to the product measure over $\mathbf{D}^\omega$ (as reviewed, for example, in [Levy, 1979, Section VII.3]). Given a structure $\mathcal{S}$, this measure is extended to sets E of environments for $\mathcal{S}$ via their underlying $\mathbf{D}$-sequences. That is, the P-measure of E is the P-measure of $\{d \in \mathbf{D}^\omega \mid \text{ for some } e \in E, e \text{ is for } \mathcal{S} \text{ and } d\}$. (All sets of environments measured below are Borel.)

In what follows we ignore members of $\mathbf{D}^\omega$ that are not onto $\mathbf{D}$. This is because the class of such sequences has measure zero for any $P \in \mathbf{P}$, by the positivity of P (for discussion see [Billingsley, 1986, Chapter 4]). Recall from Section 6.2 that $\mathbf{D}$-sequences are, by definition, onto $\mathbf{D}$. The following lemma is easy to demonstrate.

(39) LEMMA: Let structure $\mathcal{S}$ be given, and let E be the class of environments for $\mathcal{S}$. Then for all $P \in \mathbf{P}$, E has P-measure 1.

8.2 Success criterion

To give probabilistic character to scientific success we modify only the concept of solving a structure. The same success criterion as before applies to individual environments (see Definition (27)).

(40) DEFINITION: Let $\mathrm{X} \subseteq \mathcal{L}_{sen}$, $\mathbf{P}_0 \subseteq \mathbf{P}$, and scientist Ψ be given.

(a) Let structure $\mathcal{S}$ be given. Ψ X-*solves* $\mathcal{S}$ *on* $\mathbf{P}_0$ just in case for every $P \in \mathbf{P}_0$, the set of environments for $\mathcal{S}$ that Ψ X-solves has P-measure 1.

(b) Let collection $\mathcal{K}$ of structures be given. Ψ X-*solves* $\mathcal{K}$ *on* $\mathbf{P}_0$ just in case Ψ X-solves every $\mathcal{S} \in \mathcal{K}$ on $\mathbf{P}_0$. In this case, $\mathcal{K}$ is said to be X-*solvable* on $\mathbf{P}_0$.

If $\mathbf{P}_0$ is a singleton set $\{P\}$, we drop the braces when employing the foregoing terminology.

Of course, if $\mathbf{P}_0, \mathbf{P}_1$ are classes of distributions with $\mathbf{P}_0 \subseteq \mathbf{P}_1$ then X-solvability on $\mathbf{P}_1$ implies X-solvability on $\mathbf{P}_0$. Lemma (39) implies that if Ψ X-solves $\mathcal{K}$ (in the absolute sense), then Ψ X-solves $\mathcal{K}$ on $\mathbf{P}$. Definition (40) thus generalizes the absolute conception of solvability.

(41) EXAMPLE: Let $\mathcal{L}$, X, and $\mathcal{K}$ be as described in Example (30). Let $\mathbf{P}_0 \subset \mathbf{P}$ be any class of distributions such that for all $i \in N$, $glb\{P(\mathbf{d}_i) \mid P \in \mathbf{P}_0\} > 0$. Then $\mathcal{K}$ is X-solvable on $\mathbf{P}_0$.

A recursion-theoretic analogue of the contrast between Examples (30) and (41) appears in [Osherson *et al.*, 1986c, Prop. 10.5.2.A]. Further analysis is provided by [Angluin, 1988].

8.3 Comparison with absolute solvability

Examples (30) and (41) show that absolute and probabilistic solvablility do not coincide for arbitrary collections of structures. However, for *elementary* collections of structures things are different. In this case the same concept of confirmability (Definition (34)) governs solvability in both the absolute and probabilistic senses. This is revealed by the next two lemmas, which parallel (35) and (36). The first is an immediate consequence of (35) and (39).

(42) LEMMA: Let a Σ_2^0 subset X of $\mathcal{L}_{sen}$ be given. Then there is an oracle machine M such that for all $T \subseteq \mathcal{L}_{sen}$, if X is confirmable in T then M^T X-solves $mod(T)$ on $\mathbf{P}$.

(43) LEMMA: Let $\mathrm{X} \subseteq \mathcal{L}_{sen}$ be given. Then for all $P \in \mathbf{P}$ and $T \subseteq \mathcal{L}_{sen}$, if $mod(T)$ is X-solvable on P then X is confirmable in T.

Lemmas (42) and (43) directly yield the following proposition.

(44) PROPOSITION: Suppose that $\mathrm{X} \subseteq \mathcal{L}_{sen}$ is Σ^0_2. Then there is an oracle machine M such that for all $P \in \mathbf{P}$ and $T \subseteq \mathcal{L}_{sen}$, if $mod(T)$ is X-solvable on P then M^T X-solves $mod(T)$ in the absolute sense (hence M^T X-solves $mod(T)$ on $\mathbf{P}$, as well).

As a corollary we obtain:

(45) COROLLARY: Let $\theta \in \mathcal{L}_{sen}$ be given. Then for all $T \subseteq \mathcal{L}_{sen}$ the following conditions are equivalent.

(a) $mod(T)$ is $\{\theta, \neg\theta\}$-solvable.

(b) $mod(T)$ is $\{\theta, \neg\theta\}$-solvable on $\mathbf{P}$.

(c) For some $P \in \mathbf{P}$, $mod(T)$ is $\{\theta, \neg\theta\}$-solvable on P.

(d) θ is equivalent over T to a Boolean combination of existential sentences.

9 Third paradigm: Solvability with specified probability

So far in our discussion we have considered the natural-nonnatural boundary to be sharp. A more liberal attitude would define the natural languages as those for which there is some positive probability of successful acquisition by children, and recognize that different members of this class are associated with different probabilities. Such is the approach of the present section. We preserve the assumption of a sharp distinction between success and failure in any given environment, but allow the class of environments that lead to success to have measure between 0 and 1.

Formulation of this idea requires reflection about the case in which success is not achieved. In particular, we rely on the following hypothesis, which is substantive but strikes us as plausible. When the acquisition process breaks down, we assume that the child fails to converge to any grammar, rather than stabilizing to an incorrect one.

It may be interesting to view the foregoing hypothesis from a normative perspective (that is, independently of the empirical question of its veridicality for children). A scientist who solves a given structure with small probability is worse than useless if he exhibits high probability of misleading an external observer. In particular, it is misleading to converge to a false theory; for in this case the mistaken theory appears to be held with confidence, and risks being accredited. If the probability that the scientist misleads us this way is high, and the probability of genuine success low, it might be better to show him no data at all.

9.1 Definitions and principal theorem

These considerations suggest the following definitions.

(46) DEFINITION: Let scientist Ψ, structure $\mathcal{S}$ and environment e for $\mathcal{S}$ be given. Ψ *is misleading* on e just in case Ψ converges on e to $\theta \in \mathcal{L}_{sen}$ such that $\mathcal{S} \not\models \theta$.

Given $X \subseteq \mathcal{L}_{sen}$, if Ψ X-solves structure $\mathcal{S}$ then Ψ is not misleading on any environment for $\mathcal{S}$. Definition (46) is inspired by the concept of "reliability" from the recursion theoretic literature (see [Blum & Blum, 1975]).

(47) DEFINITION: Let $r \in [0, 1]$, $X \subseteq \mathcal{L}_{sen}$, $\mathbf{P}_0 \subseteq \mathbf{P}$, and scientist Ψ be given.

(a) Given structure $\mathcal{S}$, we say that Ψ *X-solves $\mathcal{S}$ on $\mathbf{P}_0$ with probability* r just in case the following conditions hold for all $P \in \mathbf{P}_0$.

  i. The set of environments for $\mathcal{S}$ that Ψ X-solves has P-measure at least r.[14]

  ii. The set of environments for $\mathcal{S}$ on which Ψ is misleading has P-measure 0.

(b) Given collection $\mathcal{K}$ of structures, we say that Ψ *X-solves $\mathcal{K}$ on $\mathbf{P}_0$ with probability* r just in case Ψ X-solves every $\mathcal{S} \in \mathcal{K}$ on $\mathbf{P}_0$ with probablity r. In this case, $\mathcal{K}$ is said to be *X-solvable* on $\mathbf{P}_0$ with probability r.

Clause a-ii of the definition embodies our hypothesis that acquisition failure results in nonconvergence. On the normative side, it renders useful any scientist whose chance of success is positive. In particular, the hypotheses of such a scientist lend themselves to aggregation within a larger scientific community (see [Jain & Sharma, 1990a, Osherson *et al.*, 1986a, Pitt & Smith, 1988] for discussion of aggregating scientific competence).[15]

Definition (47) generalizes the earlier paradigms. This is shown by the following lemma, which follows immediately from our definitions.

(48) LEMMA: Let $P \in \mathbf{P}$, scientist Ψ, $X \subseteq \mathcal{L}_{sen}$, and structure $\mathcal{S}$ be given. If either

(a) Ψ X-solves $\mathcal{S}$ or

(b) Ψ X-solves $\mathcal{S}$ on P

then Ψ X-solves structure $\mathcal{S}$ on P with probability 1.

The present conception of scientific success has a "zero-one" character, as revealed by the following proposition.

(49) PROPOSITION: Let $X \subseteq \mathcal{L}_{sen}$, $\mathbf{P}_0 \subseteq \mathbf{P}$, and collection $\mathcal{K}$ of structures be given. Then $\mathcal{K}$ is X-solvable on $\mathbf{P}_0$ with probability greater than 0 if and only if $\mathcal{K}$ is X-solvable on $\mathbf{P}_0$.

From Proposition (44), Corollary (45), and Proposition (49) we have the following immediate corollaries.

[14] Recall from Section 8.1 that the measure of a set of environments is defined via their underlying **D**-sequences.

[15] We note that the aggregation problem is distinct from "team learning" in the sense of [Daley, 1986, Jain & Sharma, 1990b, Pitt, 1989]. The latter paradigm requires only that a single scientist arrive at the truth, not that divergent opinions be unified into a correct one.

(50) COROLLARY: Suppose that $\mathrm{X} \subseteq \mathcal{L}_{sen}$ is Σ^0_2. Then there is an oracle machine M such that for all $P \in \mathbf{P}$ and $T \subseteq \mathcal{L}_{sen}$, if $mod(T)$ is X-solvable on P with probability greater than 0, then M^T X-solves $mod(T)$ in the absolute sense.

(51) COROLLARY: Let $\theta \in \mathcal{L}_{sen}$ be given. Then for all $T \subseteq \mathcal{L}_{sen}$ the following condition is equivalent to (a) - (d) of Corollary (45).

(e) For some $P \in \mathbf{P}$, $mod(T)$ is $\{\theta, \neg\theta\}$-solvable on P with probability greater than 0.

10 Empirical evaluation

The paradigms discussed above provide at best a crude picture of first language acquisition by children. We provide a partial list of their deficiencies.

(a) The linguistic data available to children are not adequately represented by the formal concept of environment. The issue of negative information was already noted in Section 6.7, above. In addition, the concept of probabilistic solvability portrays data as arising via identically distributed, stochastically independent sampling. It is easy to see that real language does not arise in this way (for discussion see [Angluin, 1988, Osherson *et al.*, 1986b]).

(b) Except for computability, our paradigms provide no constraint on the class of formal scientists whereas the inductive mechanisms of children surely operate under severe limitations. At the least, we can assume that children have limited memory for the precise form of spoken sentences, and that the time devoted to processing any given datum is recursively bounded. Building these constraints into formal scientists alters the collections of structures that can be solved.[16]

(c) The criterion of solvability is both too weak and too strong compared to actual language acquisition. It is too strong in requiring selection of $\theta \in \mathrm{X}$ that is "exactly" true in the underlying structure. Since the grammatical theories issuing from normal language acquistion are not likely to be entirely accurate reflections of the input language, more realistic paradigms would incorporate a suitable notion of "approximate truth" (for discussion of this notion, see [Kuipers, 1987, Osherson *et al.*, 1989]). On the other hand, solvability is too weak inasmuch as it imposes no requirements on the number of data that must be examined before convergence begins. In contrast, the rapidity of first language acquisition is one of its striking features. Note also that solvability for individual environments is defined here as an all-or-nothing affair. In reality, children might harbor random processes that yield only probable success within any fixed set of circumstances.[17]

As seen in Section 4, the foregoing issues (among others) have begun to be addressed within the recursion theoretic tradition in Learning Theory. In contrast, their exploration within a first-order framework has hardly been initiated.

[16]Preliminary work on restricted classes of scientists within the model theoretical perspective is reported in [Gaifman *et al.*, 1990, Osherson & Weinstein, 1986].

[17]For an analysis of random processes in learning, see [Daley, 1986, Pitt, 1989].

11 Concluding remarks

Apart from concerns about first language acquisition, the model theoretic paradigms discussed in this chapter may be examined from an epistemological point of view. For example, Proposition (33) indicates that there is an upper bound on scientific competence, at least for elementarily defined starting points (in the sense of Section 7.2). Moreover, this bound is already reached by a Turing Machine whose sole recourse to an oracle is to determine the axioms of the background theory. The theorem might thus be relevant to the thesis T according to which human mentation is computer simulable. Although T might imply various bounds on human knowledge or capacity, Proposition (33) provides one sense in which the scope of scientifically attainable knowledge is not affected by the status of T.[18] Corollary (50) provides an even stronger sense.

Theorem (33) raises questions about the character of first-order logic itself. To what extent is the theorem linked to the special properties of the predicate calculus? Are there analogous theorems for stronger logics? Inversely, are all of the deductive consequences of first-order logic necessary for conducting scientific inquiry, including such inferences as $p \models p \vee q$ (sometimes thought to have an odd character [Schurz & Weingartner, 1987])? Some preliminary results that bear on these questions are presented in [Osherson *et al.*, 1991b, Section 4], [Osherson & Weinstein, 1993].

12 Appendix: Proofs

12.1 Proof of Lemma (4)

We restrict attention to scientists that are total functions; that no generality is lost follows from [Osherson *et al.*, 1986c, Props. 4.3.1A,B]. Assume that Ψ identifies L but no locking sequence for Ψ and L exists. Moreover assume that $a_0, a_1, a_2, \ldots$ is an enumeration of L. We now construct in stages a special text t for L.

> **Stage 0:** Start t with a_0.
>
> **Stage $n+1$:** Suppose that $t[m_0]$ has been constructed at stage n. By assumption, this sequence is not a locking sequence. So, it can be extended by elements of L to some τ such that either $\Psi(\tau)$ is not an index for L or $\Psi(\tau) \neq \Psi(t[m_0])$. Let τ followed by a_{n+1} be the segment of t constructed in the present stage.

It is easy to see that t is a text for L, and that Ψ does not converge on t to an index for L. Hence Ψ does not identify L, contradicting our assumption.

12.2 Notation

The following notation will be helpful in the sequel. Given **D**-sequence d and structure $\mathcal{S}$, we let $[\mathcal{S}, d]$ denote the environment for $\mathcal{S}$ and d. Given structure $\mathcal{S}$ and $\bar{d} \in \mathbf{D}^{<\omega}$ of length $n \in N$,

[18]For discussion of the machine simulability of thought, see [Glymour, 1992, Chapter 13] and references cited there.

we let $[\mathcal{S}, \bar{d}]$ denote $e[n]$, where $e = [\mathcal{S}, d]$ and d extends $\bar{d}$. For example, with $\mathcal{S}$ and d as in Section 6.3, $[\mathcal{S}, d[3]]$ is displayed just above Lemma (25) (ignoring the ...). It is helpful to note that for structures $\mathcal{S}, \mathcal{U}$, and $\bar{d}, \bar{u} \in \mathbf{D}^{<\omega}$, $[\mathcal{S}, \bar{d}] = [\mathcal{U}, \bar{u}]$ iff $\mathcal{S}$ restricted to $\bar{d}$ is isomorphic to $\mathcal{U}$ restricted to $\bar{u}$.

12.3 Model-theoretic locking sequences

In the model-theoretic paradigms the following version of the locking sequence lemma is used. It has been demonstrated elsewhere in diverse forms (e.g., [Osherson & Weinstein, 1982b, Lemma B], [Osherson *et al.*, 1991b, Lemma 24]). The proof resembles that for Lemma (4), and we do not rehearse it here.

(52) DEFINITION: Let scientist Ψ, structure $\mathcal{S}$, and $\bar{d} \in \mathbf{D}^{<\omega}$ be given. $\bar{d}$ is a *locking sequence* for $(\Psi, \mathcal{S})$ just in case:

- (a) $\Psi([\mathcal{S}, \bar{d}]) \in \mathcal{L}_{sen}$, i.e., Ψ is defined on $[\mathcal{S}, \bar{d}]$, and
- (b) for all $\bar{d}' \in \mathbf{D}^{<\omega}$ that extend $\bar{d}$, $\Psi([\mathcal{S}, \bar{d}']) = \Psi([\mathcal{S}, \bar{d}])$.

(53) LEMMA: Let $X \subseteq \mathcal{L}_{sen}$, scientist Ψ, and structure $\mathcal{S}$ be given. Suppose that Ψ X-solves every environment for $\mathcal{S}$. Then there is a locking sequence $\bar{d}$ for $(\Psi, \mathcal{S})$. Moreover, $\mathcal{S} \models \Psi([\mathcal{S}, \bar{d}])$.

12.4 Proof of Example (30)

Suppose that Ψ X-solves $\mathcal{S}_0$. Then, because $\mathcal{S}_0 \models \theta$, Lemma (53) implies the existence of $\bar{d} \in \mathbf{D}^{<\omega}$ such that:

(54) for all $\bar{d}' \in \mathbf{D}^{<\omega}$ that extend $\bar{d}$, $\Psi([\mathcal{S}_0, \bar{d}']) = \theta$.

Choose $i \in N$ large enough so that $\mathcal{S}_i \models [\mathcal{S}_0, \bar{d}]$. Let $\mathbf{D}$-sequence h extend $\bar{d}$. Then it is easy to verify that:

(55) for all $j \geq \mathit{length}(\bar{d})$ there is $\bar{d}' \in \mathbf{D}^{<\omega}$ of length j such that:

- (a) $\bar{d}'$ extends $\bar{d}$, and
- (b) $[\mathcal{S}_i, \bar{h}_j] = [\mathcal{S}_0, \bar{d}']$.

By (54) and (55), Ψ converges on $[\mathcal{S}_i, h]$ to θ. It follows that Ψ does not X-solve $\mathcal{S}_i$ since $\mathcal{S}_i \not\models \theta$.

12.5 Proof of Lemma (36)

We rely on the following notation.

(56) DEFINITION: Let structure $\mathcal{S}$ and $\bar{d} \in \mathbf{D}^{<\omega}$ be given.

(a) The set

$$\{\ \pi \in \mathcal{L}_{form} \mid \pi \text{ is universal, } var(\pi) \subseteq domain(\bar{d}), \text{ and } \mathcal{S} \models \pi[\bar{d}]\ \}$$

is denoted by $\forall$-*type*$(\bar{d},\mathcal{S})$.

(b) The set

$$\{\ \pi \in \mathcal{L}_{form} \mid \pi \text{ is existential, } var(\pi) \subseteq domain(\bar{d}), \text{ and } \mathcal{S} \models \pi[\bar{d}]\ \}$$

is denoted by $\exists$-*type*$(\bar{d},\mathcal{S})$.

Let scientist Ψ, $\mathrm{X} \subseteq \mathcal{L}_{sen}$, and $T \subseteq \mathcal{L}_{sen}$ be such that Ψ X-solves *mod*(T). We suppose that T is satisfiable and $\mathrm{X} \neq \emptyset$ (the other cases are trivial). By Lemma (53) choose $\bar{d} \in \mathbf{D}^{<\omega}$ and $\theta \in \mathrm{X}$ such that:

(57) (a) $\bar{d}$ is a locking sequence for $(\Psi, \mathcal{S})$, and
(b) $\Psi([\mathcal{S}, \bar{d}]) = \theta$.

It is sufficient to show that there is $\varphi \in \mathcal{L}_{sen}$ such that:

(58) (a) φ is of form $\exists\forall$,
(b) $\mathcal{S} \models \varphi$, and
(c) $T \cup \{\varphi\} \models \theta$.

(59) FACT: Suppose that $\mathcal{U} \in mod(T)$ and sequence $\bar{u}$ are such that $length(\bar{u}) = length(\bar{d})$ and $\exists$-*type*$(\bar{u},\mathcal{U}) \subseteq \exists$-*type*$(\bar{d},\mathcal{S})$. Then $\mathcal{U} \models \theta$.

Proof: Suppose that $\mathcal{U}$, $\bar{u}$ satisfy the assumptions, and let $\bar{u'}$ extend $\bar{u}$. Let $\chi \in \mathcal{L}_{form}$ be the conjunction of the basic formulas in $[\mathcal{U}, \bar{u'}]$. Then $\mathcal{U} \models \exists x_{length(\bar{u})} \cdots \exists x_{length(\bar{u'})-1} \chi[\bar{u}]$. Hence, because $\exists$-*type*$(\bar{u},\mathcal{U}) \subseteq \exists$-*type*$(\bar{d},\mathcal{S})$, $\mathcal{S} \models \exists x_{length(\bar{u})} \cdots \exists x_{length(\bar{u'})-1} \chi[\bar{d}]$. Hence, some extension $\bar{d'}$ of $\bar{d}$ of the same length as $\bar{u'}$ satisfies $[\mathcal{S}, \bar{d'}] = [\mathcal{U}, \bar{u'}]$. So, by (57)a, $\Psi([\mathcal{U}, \bar{u'}]) = \Psi([\mathcal{S}, \bar{d'}]) = \Psi([\mathcal{S}, \bar{d}])$. We infer immediately that:

(60) $\bar{u}$ is a locking sequence for $(\Psi,\mathcal{U})$.

From the same equality (with $\bar{u'} = \bar{u}$) and (57)b, we obtain:

(61) $\Psi([\mathcal{U}, \bar{u}]) = \theta$.

Finally, (59) follows from (60), (61) and the assumptions that $\mathcal{U} \in mod(T)$ and Ψ X-solves *mod*(T). ■

Using (59), we now show that:

(62) FACT: $T \cup \forall$-*type*$(\bar{d},\mathcal{S}) \models \theta$

Proof: By the Löwenheim-Skolem theorem it is sufficient to show that θ holds in any countable model $\mathcal{U} \in mod(T)$ in which $\forall$-*type*$(\bar{d},\mathcal{S})$ is satisfied by a sequence $\bar{u}$ of the same length as $\bar{d}$. So assume that $\mathcal{U}$ and $\bar{u}$ are such a model and sequence. Then $\exists$-*type*$(\bar{u},\mathcal{U}) \subseteq \exists$-*type*$(\bar{d},\mathcal{S})$. Hence, by (59), θ holds in $\mathcal{U}$. ■

By compactness there is a finite subset Π of $\forall$-*type*$(\bar{d},\mathcal{S})$ such that

(63) $T \cup \Pi \models \theta$

To witness (58), let φ be the existential closure of the conjunction of Π. Then, φ can immediately be seen to satisfy (58)a,b. That φ satisfies (58)c follows directly from (63).

12.6 Proof of Example (41)

Let $\mathbf{P}_0$, X and $\mathcal{K}$ be as specified in the example.

Given $i, j \in N$, let $A_{i,j} \subset \mathbf{D}^\omega$ be the collection of $\mathbf{D}$-sequences d such that not all of $\mathbf{d}_0 \ldots \mathbf{d}_i$ occur in $d[j]$. By the assumption on $\mathbf{P}_0$, let strictly increasing $f : N \to N$ be such that for all $i \in N$ and $P \in \mathbf{P}_0$, $P(A_{i,f(i)}) < \frac{1}{2^i}$. So for each $P \in \mathbf{P}_0$, $\Sigma_i P(A_{i,f(i)})$ converges. Hence, by the first Borel-Cantelli lemma ([Billingsley, 1986, Theorem 4.3]):

(64) $P(\limsup_i A_{i,f(i)}) = 0$ for every $P \in \mathbf{P}_0$.

Via the definition of $A_{i,j}$, (64) yields:

(65) For every $P \in \mathbf{P}_0$ the class of $\mathbf{D}$-sequences d such that

$$\{\mathbf{d}_0, \ldots, \mathbf{d}_i\} \not\subseteq range(d[f(i)])$$

for infinitely many $i \in N$ has P-measure 0.

Define scientist Ψ as follows. For all environments e and all $j \in N$:

(a) if $j \in range(f)$ and $\bigwedge e[j]$ implies the existence of an R-chain of length at least $f^{-1}(j)$, then $\Psi(e[j]) = \theta$; (since f is strictly increasing, $f^{-1}(j)$ is well-defined).

(b) if $j \in range(f)$ and $\bigwedge e[j]$ does not imply the existence of an R-chain of length at least $f^{-1}(j)$, then $\Psi(e[j]) = \neg\theta$;

(c) if $j \notin range(f)$, then $\Psi(e[j]) = \Psi(e[j-1])$ (unless $j = 0$, in which case $\Psi(e[j]) = \forall x(x = x)$).

Now let $\mathcal{S}_j \in \mathcal{K}$ be given, with $j > 0$. Let $\mathbf{D}$-sequence d and environment e for $\mathcal{S}_j$ and d also be given. Then for all but finitely many $i \in N$, $\bigwedge e[f(i)]$ does not imply the existence of an R-chain of length at least i (because there is no such chain). Hence, case (b) above arises for a cofinite subset of $range(f)$ whereas case (a) arises for only a finite subset of $range(f)$. It follows that Ψ converges to $\neg\theta$ on e. Hence, Ψ X-solves $\mathcal{S}_j$, so Ψ X-solves $\mathcal{S}_j$ on $\mathbf{P}_0$.

Regarding $\mathcal{S}_0$, let e be for $\mathcal{S}_0$ and d. Call e "bad" just in case for infinitely many $i \in N$, $\bigwedge e[f(i)]$ does not imply the existence of an R-chain of length at least i. It follows directly from (65) that for every $P \in \mathbf{P}_0$, the class of bad environments for $\mathcal{S}_0$ has P-measure 0. Hence, the P-probability is 1 that case (a) arises for a cofinite subset of $range(f)$ whereas case (b) arises for only a finite subset of $range(f)$. Thus, for every $P \in \mathbf{P}_0$, Ψ converges to θ on a class of environments for $\mathcal{S}_0$ of P-measure 1, so Ψ X-solves $\mathcal{S}_0$ on $\mathbf{P}_0$. ■

12.7 Proof of Lemma (43)

Our demonstration of Lemma (43) proceeds via the following definition and propositions.

(66) Definition: Let $P \in \mathbf{P}$ be given. The class of $P' \in \mathbf{P}$ such that for some permutation $\pi : N \to N$, $P' = \{(\mathbf{d}_i, P(\mathbf{d}_{\pi(i)})) \mid i \in N\}$ is denoted $PERM(P)$.

(67) Proposition: Let $X \subseteq \mathcal{L}_{sen}$, $P \in \mathbf{P}$, and $T \subseteq \mathcal{L}_{sen}$ be given. If $mod(T)$ is X-solvable on P then $mod(T)$ is X-solvable on $PERM(P)$.

(68) Proposition: Let $X \subseteq \mathcal{L}_{sen}$, $P \in \mathbf{P}$, and $T \subseteq \mathcal{L}_{sen}$ be given. If X is not confirmable in T then for some $P' \in PERM(P)$, $mod(T)$ is not X-solvable on P'.

To obtain Lemma (43) from the foregoing, let $X \subseteq \mathcal{L}_{sen}$, $P \in \mathbf{P}$, and $T \subseteq \mathcal{L}_{sen}$ be given, and suppose that $mod(T)$ is X-solvable on P. Then by Proposition (67), $mod(T)$ is X-solvable on $PERM(P)$. So, by Proposition (68), X is confirmable in T.

It remains to prove the two propositions.

12.7.1 Proof of Proposition (67)

Suppose that scientist Ψ X-solves $mod(T)$ on P, and let $\mathcal{S} \in mod(T)$ be given. Let π be any permutation of N, and let $P' = \{(\mathbf{d}_i, P(\mathbf{d}_{\pi(i)})) \mid i \in N\}$. It suffices to show that Ψ X-solves $\mathcal{S}$ on P'.

Given $\mathbf{D}$-sequence d, let $\pi(d)$ be its permutation (via the indexes of $\mathbf{D}$) under π. Given subset D of $\mathbf{D}$-sequences, let $\pi(D) = \{\pi(d) \mid d \in D\}$. Since the measure of a collection of $\mathbf{D}$-sequences is determined only by the probabilities applying to each coordinate (and not by their names), we have:

(69) For every set D of $\mathbf{D}$-sequences, $P(\pi(D)) = P'(D)$.

Let $\mathcal{S}'$ be the structure whose vocabulary is interpreted in the following way. For individual constant c, $c^{\mathcal{S}'} = \mathbf{d}_{\pi(j)}$ iff $c^{\mathcal{S}} = \mathbf{d}_j$. For n-ary relation symbol R, $(d_{\pi(j_1)}, \ldots, d_{\pi(j_n)}) \in R^{\mathcal{S}'}$ iff $(d_{j_1}, \ldots, d_{j_n}) \in R^{\mathcal{S}}$. Evidently, $\mathcal{S}' \in mod(T)$ since $\mathcal{S}$ and $\mathcal{S}'$ are isomorphic. Hence:

(70) Ψ X-solves $\mathcal{S}'$ on P.

Let D be the set of $\mathbf{D}$-sequences d such that Ψ X-solves $[\mathcal{S}', \pi(d)]$ on P. By (70), $P(\pi(D)) = 1$, so by (69) the proof is completed if we show that for almost every $d \in D$, Ψ X-solves $[\mathcal{S}, d]$ on P'. However, this follows immediately from the following fact, easy to verify: for all $\mathbf{D}$-sequences d, $[\mathcal{S}', \pi(d)] = [\mathcal{S}, d]$.

12.7.2 Proof of Proposition (68)

We begin by collecting some facts.

(71) FACT: Let $P \in \mathbf{P}$ be given. For every $q \in [0, 1)$ there is finite $D \subset \mathbf{D}$ with $P(D) > q$.

Proof: Immediate from [Billingsley, 1986, Theorem 2.1]. ■

(72) FACT: Let $\mathrm{X} \subseteq \mathcal{L}_{sen}$ and $T \subseteq \mathcal{L}_{sen}$ be given. Suppose that X is not confirmable in T. Then there is $\mathcal{S} \in mod(T)$ such that:

> for every $\varphi \in \mathcal{L}_{sen}$, if φ is $\exists\forall$ and $\mathcal{S} \models \varphi$ then for all $\theta \in \mathrm{X}$, $T \cup \{\varphi\} \cup \{\neg\theta\}$ is satisfiable.

Proof: Directly from Definition (34). ■

We call $\bar{d} \in \mathbf{D}^{<\omega}$ "repetition free" just in case $length(\bar{d}) = card(range(\bar{d}))$. A $\mathbf{D}$-sequence h is repetition free just in case all of its initial segments are. In what follows, repetition free sequences will code the domains of finite functions that can be extended to a probability distribution.

(73) FACT: Let $\mathrm{X} \subseteq \mathcal{L}_{sen}$ and $T \subseteq \mathcal{L}_{sen}$ be given. If X is not confirmable in T, then there is $\mathcal{S} \in mod(T)$ such that for every $\theta \in \mathrm{X}$ and repetition free $\bar{u} \in \mathbf{D}^{<\omega}$ there is structure $\mathcal{U}$ and repetition free $\bar{u'} \in \mathbf{D}^{<\omega}$ such that:

(a) $\mathcal{U} \in mod(T)$,
(b) $length(\bar{u'}) = length(\bar{u})$,
(c) $\mathcal{U} \models \neg\theta$, and
(d) $\exists$-*type*$(\bar{u'},\mathcal{U}) \subseteq \exists$-*type*$(\bar{u},\mathcal{S})$.

Proof: Suppose that X is not confirmable in T. Then by Definition (34) there is $\mathcal{S} \in mod(T)$ such that:

(74) for every $\varphi \in \mathcal{L}_{sen}$, if φ is $\exists\forall$ and $\mathcal{S} \models \varphi$ then for all $\theta \in \mathrm{X}$, $T \cup \{\varphi\} \cup \{\neg\theta\}$ is satisfiable.

Now let $\theta \in \mathrm{X}$ and repetition free $\bar{u} \in \mathbf{D}^{<\omega}$ be given. By (74), for every finite $\{\pi_1, \ldots, \pi_m\} \subseteq \forall$-*type*$(\bar{u},\mathcal{S})$, $T \cup \{\pi_1 \wedge \ldots \wedge \pi_m \wedge \exists x_1 \ldots x_m \bigwedge_{i \neq j} x_i \neq x_j\} \cup \{\neg\theta\}$ is satisfiable. This is because the existential closure of $\pi_1 \wedge \ldots \wedge \pi_m \wedge \exists x_1 \ldots x_m \bigwedge_{i \neq j} x_i \neq x_j$ is equivalent to an $\exists\forall$ sentence, and $\mathcal{S}$ satisfies $\exists x_1 \ldots x_m \bigwedge_{i \neq j} x_i \neq x_j$, since $|\mathcal{S}| = \mathbf{D}$ which is infinite. It follows from the compactness and Löwenheim-Skolem theorems that there is (countably infinite) $\mathcal{U} \in mod(T)$ and repetition free $\mathbf{D}$-sequence h such that $\mathcal{U} \models \{\neg\theta\} \cup \forall\text{-}type(\bar{u},\mathcal{S})[h]$. Let $\bar{u}' \in \mathbf{D}^{<\omega}$ be the initial segment of h with the same length as $\bar{u}$. Then $\mathcal{U}, \bar{u}'$ are easily seen to satisfy the Fact. ■

To demonstrate Proposition (68), fix $\mathrm{X} \subseteq \mathcal{L}_{sen}$, $P \in \mathbf{P}$, and $T \subseteq \mathcal{L}_{sen}$ for the remainder of the proof. Assume that $\mathrm{X} \neq \emptyset$ and that T is satisfiable (the other cases are trivial). Suppose that:

(75) X is not confirmable in T.

For a reductio, let scientist Ψ satisfy:

(76) Ψ X-solves $mod(T)$ on $PERM(P)$.

(77) FACT: Let $q \in [0,1)$, $P' \in PERM(P)$, and $\mathcal{S} \in mod(T)$ be given. Then there are cofinitely many $n \in N$ such that

$$P'(\ \{\ d \ \mid \ \Psi([\mathcal{S}, d[n]]) \in \mathrm{X} \text{ and } \mathcal{S} \models \Psi([\mathcal{S}, d[n]])\ \}\ ) > q.$$

Intuitively, as Ψ moves further and further down a randomly constructed environment, the probability goes to 1 that Ψ begins to emit adequate conjectures.

Proof: For a reductio, suppose that the conclusion to Fact (77) is false. Then for infinitely many $n \in N$, $P'(Z_n) \geq 1 - q$, where Z_n is the complement of

$$\{\ d \ \mid \ \Psi([\mathcal{S}, d[n]]) \in \mathrm{X} \text{ and } \mathcal{S} \models \Psi([\mathcal{S}, d[n]])\ \}.$$

The Bolzano-Weierstrass theorem thus yields $\limsup_n P'(Z_n) \geq 1 - q$. By [Billingsley, 1986, Theorem 4.1(i)], this implies:

(78) $P'(\limsup_n Z_n) \geq 1 - q$.

Now suppose that $d \in \limsup_n Z_n$, and that e is for $(\mathcal{S}, d)$. Then Ψ fails to X-solve e since for infinitely many $n \in N$, $\Psi(e[n]) \notin \mathrm{X}$ or $\mathcal{S} \not\models \Psi(e[n])$. Hence, (78) implies that Ψ fails to X-solve some collection C of environments for $\mathcal{S}$ with $P'(C) > 0$. This contradicts (76) ■

By (75), choose structure $\mathcal{S}$ that conforms to Fact (73). Given $\theta \in \mathrm{X}$ and repetition free $\bar{u} \in \mathbf{D}^{<\omega}$, let $\mathcal{U}_{\theta,\bar{u}} = \mathcal{U}$ be as guaranteed in Fact (73).

(79) DEFINITION: Let finite function $f : \mathbf{D} \to (0,1)$ be given. f is *extendable* in $PERM(P)$ just in case there is $P' \in PERM(P)$ that extends f.

(80) FACT: Let $q \in [0,1)$, $m \in N$, $\theta \in \mathrm{X}$, repetition free $\bar{u} \in \mathbf{D}^{<\omega}$ and $f : range(\bar{u}) \to (0,1)$ be given. Suppose that f is extendable in $PERM(P)$. Then there is $n > m$, and finite function F with the following properties.

(a) F extends f and is extendable in $PERM(P)$.

(b) Let $\widehat{\mathbf{d}}$ be the class of $\mathbf{D}$-sequences b with $\Psi([\mathcal{S}, b[n]]) \in \mathrm{X}$ and $\mathcal{U}_{\theta,\bar{u}} \models \Psi([\mathcal{S}, b[n]])$. Then $F(\widehat{\mathbf{d}}) > q$, i.e., for every $P^* \in PERM(P)$ that extends F, $P^*(\widehat{\mathbf{d}}) > q$.

Proof: Let θ, $\bar{u}$, and f be as described. Let repetition free $\bar{u'} \in \mathbf{D}^{<\omega}$ be as specified in Fact (73). Choose $P' \in PERM(P)$ such that for all $i < length(\bar{u})$, $P'(\bar{u'}_i) = f(\bar{u}_i)$. By Facts (71), (77), and the additivity of P', choose $n > m$ and $\bar{r'} \in \mathbf{D}^n$ so that:

(81) (a) $\bar{r'}$ is repetition free and $range(\bar{u'}) \cap range(\bar{r'}) = \emptyset$.

(b) If $\widehat{\mathbf{d}'}$ is the class of $\mathbf{D}$-sequences b' with $range(b'[n]) \subseteq range(\bar{u'}) \cup range(\bar{r'})$, $\Psi([\mathcal{U}_{\theta,\bar{u}}, b'[n]]) \in \mathrm{X}$ and $\mathcal{U}_{\theta,\bar{u}} \models \Psi([\mathcal{U}_{\theta,\bar{u}}, b'[n]])$, then $P'(\widehat{\mathbf{d}'}) > q$.

By Fact (73)d, choose $\bar{r} \in \mathbf{D}^{<\omega}$ and mapping M such that:

(82) (a) $\bar{r}$ is repetition free and $range(\bar{u}) \cap range(\bar{r}) = \emptyset$.

(b) $length(\bar{r}) = length(\bar{r'})$.

(c) For all finite or infinite $\mathbf{D}$-sequences b, $M(b)$ is the result of simultaneously interchanging $\bar{u'}_i$ with $\bar{u}_i$ and $\bar{r'}_j$ with $\bar{r}_j$ in the coordinates of b, for $i < length(\bar{u})$ and $j < length(\bar{r})$.

(d) For all $\bar{g} \in \mathbf{D}^{<\omega}$ with $range(\bar{g}) \subseteq range(\bar{u'}) \cup range(\bar{r'})$, $[\mathcal{U}_{\theta,\bar{u}}, \bar{g}] = [\mathcal{S}, M(\bar{g})]$.

Let $F = f \cup \{(\bar{r}_i, P'(\bar{r'}_i)) \mid i < length(\bar{r})\}$. It is clear that F meets condition (a) of Fact (80). For (b), let $\widehat{\mathbf{d}}_0 = \{M(b') \mid b' \in \widehat{\mathbf{d}'}\}$, where $\widehat{\mathbf{d}'}$ is as specified in (81)b. Since $P'(\widehat{\mathbf{d}'}) > q$, $F(\widehat{\mathbf{d}}_0) > q$. So the proof is completed by showing that $\widehat{\mathbf{d}}_0 \subseteq \widehat{\mathbf{d}}$, i.e., that:

(83) For all $b \in \widehat{\mathbf{d}}_0$, $\Psi([\mathcal{S}, b[n]]) \in \mathrm{X}$ and $\mathcal{U}_{\theta,\bar{u}} \models \Psi([\mathcal{S}, b[n]])$.

So suppose that $b = M(b')$ for $b' \in \widehat{\mathbf{d}'}$. By (81)b and (82)d, respectively, we obtain

(84) (a) $\Psi([\mathcal{U}_{\theta,\bar{u}}, b'[n]]) \in \mathrm{X}$ and $\mathcal{U}_{\theta,\bar{u}} \models \Psi([\mathcal{U}_{\theta,\bar{u}}, b'[n]])$, and

(b) $[\mathcal{U}_{\theta,\bar{u}}, b'[n]] = [\mathcal{S}, b[n]]$.

from which (83) follows immediately. ■

We shall now construct $P' \in PERM(P)$ to satisfy:

(85) For every $\theta \in \mathrm{X}$ there are infinitely many $m \in N$ with the following property. Let $\widehat{\mathbf{d}}_m$ be the class of $\mathbf{D}$-sequences b such that $\Psi([\mathcal{S}, b[m]]) \neq \theta$. Then $P'(\widehat{\mathbf{d}}_m) > \frac{2^{m+1}-1}{2^{m+1}}$.

To see that this suffices to complete the proof, suppose (85) and let $\theta \in \mathrm{X}$ be given. Since $\lim_m \frac{2^{m+1}-1}{2^{m+1}} = 1$, $P'(\limsup_m \widehat{\mathbf{d}}_m) = 1$ by [Billingsley, 1986, Theorem 4.1(i)]. Hence:

$P'(\{b \mid \Psi$ does not converge on $[\mathcal{S}, b]$ to $\theta\}) = 1$.

Since this holds of arbitrary $\theta \in \mathrm{X}$, the countable additivity of P' yields:

$$P'(\{b \mid \Psi \text{ converges on } [\mathcal{S}, b] \text{ to no } \theta \in \mathrm{X}\}) = 1$$

which contradicts our choice of Ψ.

For the construction of the needed P', let $\{\theta_i \mid i \in N\}$ index X, and be such that:

(86) For every $\theta \in \mathrm{X}$ there are infinitely many $i \in N$ with $\theta_i = \theta$.

Given $n, m \in N$, Let $\widehat{\mathbf{d}}_{n,m}$ be the class of $\mathbf{D}$-sequences b such that $\Psi([\mathcal{S}, b[n]]) \not\models \theta_m$.

The desired P' is constructed in stages. For $m \geq -1$, we let f_m denote the finite initial segment of P' constructed at stage m. $f_{-1} = \emptyset$. Let $\{r_i \mid i \in N\}$ enumerate *range*(P). For $m \geq 0$ it will be the case that:

(87) (a) $\mathbf{d}_m \in \mathit{domain}(f_m)$;
(b) $\mathit{card}(f_m^{-1}(r_m)) = \mathit{card}(P^{-1}(r_m))$;
(c) f_m is extendable in *PERM*(P);
(d) $f_m \subset f_{m+1}$; and
(e) for some $n > m$, $f_m(\widehat{\mathbf{d}}_{n,m}) > \frac{2^{m+1}-1}{2^{m+1}}$.

We take $P' = \bigcup_m f_m$. Condition (87)e and (86) imply (85). Conditions (87)a-d imply that $P' \in \mathit{PERM}(P)$.

Stage $m \geq 0$: Suppose that f_{m-1} is extendable in *PERM*(P).

- Let $i \in N$ be least with $\mathbf{d}_i \notin \mathit{domain}(f_{m-1})$ and let $g = f_{m-1} \cup \{(\mathbf{d}_i, r_j)\}$ where $j \in N$ is least with $r_j \in \mathit{range}(P) - \mathit{range}(f_{m-1})$ (it is easy to see that such an r_j exists). This step ensures (87)a.

- Let $x = \mathit{card}(P^{-1}(r_m)) - \mathit{card}(g^{-1}(r_m))$ (it is obvious that x is finite, possibly 0). Let $\{\mathbf{d}_{i_1}, \ldots, \mathbf{d}_{i_x}\}$ be chosen from outside *domain*(g). Let $f = g \cup \{(\mathbf{d}_{i_j}, r_m) \mid j \leq x\}$ (if $x = 0$, $f = g$). This step ensures (87)b.

- Let repetition free $\bar{u} \in \mathbf{D}^{<\omega}$ be such that $\mathit{range}(\bar{u}) = \mathit{domain}(f)$. By Fact (80), choose $n > m$, and finite function F with the following properties.

  (α) F extends f and is extendable in *PERM*(P).
  (β) Let $\widehat{\mathbf{d}}$ be the class of $\mathbf{D}$-sequences b with $\Psi([\mathcal{S}, b[n]]) \in \mathrm{X}$ and $\mathcal{U}_{\theta_m, \bar{u}} \models \Psi([\mathcal{S}, b[n]])$. Then $F(\widehat{\mathbf{d}}) > \frac{2^{m+1}-1}{2^{m+1}}$.

  By the definition of $\mathcal{U}_{\theta_m,\bar{u}}$ and Fact (73)c, $\mathcal{U}_{\theta_m,\bar{u}} \not\models \theta_m$. Hence, ($\beta$) implies $F(\widehat{\mathbf{d}}_{n,m}) > \frac{2^{m+1}-1}{2^{m+1}}$. This step ensures (87)e.

- Let $f_m = F$. This step ensures (87)c,d.

It is clear that $\bigcup_m f_m$ satisfies the conditions of (87).

12.8 Proof of Proposition (47)

The right-to-left direction of the proposition is immediate. For the other direction, Let $X \subseteq \mathcal{L}_{sen}$, collection $\mathcal{K}$ of structures, and $\mathbf{P}_0 \subseteq \mathbf{P}$ be given. Suppose that scientist Ψ X-solves $\mathcal{K}$ on $\mathbf{P}_0$ with probability greater than 0. We shall exhibit scientist Ψ_0 that X-solves $\mathcal{K}$ on $\mathbf{P}_0$. For this purpose some definitions and facts are needed.

(88) DEFINITION:

(a) Given $n \in N$ and $\mathbf{D}$-sequence d, the tail of d that begins at d_n is denoted d^n.

(b) Given structure $\mathcal{S}$, we denote by $\widehat{\mathbf{d}}_{\mathcal{S}}$ the class of $\mathbf{D}$-sequences d such that for some $n \in N$:

i. $\mathit{range}(d^n) = \mathbf{D}$, and

ii. Ψ X-solves $[\mathcal{S}, d^n]$.

(89) FACT: For all $\mathcal{S} \in \mathcal{K}$ and $P \in \mathbf{P}_0$, $P(\widehat{\mathbf{d}}_{\mathcal{S}}) = 1$.

Proof: Let $P \in \mathbf{P}_0$ and $\mathcal{S} \in \mathcal{K}$ be given. By choice of Ψ, let $\widehat{\mathbf{d}}$ be a measurable class of $\mathbf{D}$-sequences such that:

(90) (a) for all $d \in \widehat{\mathbf{d}}$,

i. $\mathit{range}(d) = \mathbf{D}$,

ii. Ψ X-solves $[\mathcal{S}, d]$.

(b) $P(\widehat{\mathbf{d}}) > 0$.

By (90)a, $\widehat{\mathbf{d}} \subseteq \widehat{\mathbf{d}}_{\mathcal{S}}$, so (90)b implies $P(\widehat{\mathbf{d}}_{\mathcal{S}}) > 0$. It follows immediately from Kolmogorov's zero-one law for tail events [Billingsley, 1986, Theorem 4.5] that $P(\widehat{\mathbf{d}}_{\mathcal{S}}) = 1$. ■

By Definition (47) and the choice of Ψ we also have:

(91) FACT: Let $P \in \mathbf{P}_0$, $n \in N$, and $\mathcal{S} \in \mathcal{K}$ be given. Let $\widehat{\mathbf{d}}$ be the class of $\mathbf{D}$-sequences d such that Ψ is misleading on $[\mathcal{S}, d^n]$. Then $P(\widehat{\mathbf{d}}) = 0$.

As a final preliminary, we make use of the following definition and fact.

(92) DEFINITION: Let σ be the initial finite segment of length m in some environment e. Let scientist Ψ also be given. The *score for Ψ on σ* is the smallest $j \in N$ such that $\Psi(e[i]) = \Psi(\sigma)$ for all i with $j \leq i \leq m$.

Intuitively, the lower the score for Ψ on σ, the greater sign Ψ gives of having begun its convergence within σ. We record two obvious facts about score.

(93) FACT:

(a) Suppose that scientist Ψ converges on environment e. Then there is $s \in N$ such that for all $k \in N$, the score for Ψ on $e[k]$ is bounded by s.

(b) Suppose that scientist Ψ does not converges on environment e. Then for every $s \in N$ there is $\ell \in N$ such that the score for Ψ on $e[k]$ exceeds s if $k \geq \ell$.

In what follows we assume the existence of a uniform recursive procedure which converts finite initial segments of environments into new segments that "start over" at a specified position n. That is, for all structures $\mathcal{S}$, $\mathbf{D}$-sequences d, and $m \in N$, the procedure converts $[\mathcal{S}, d[m]]$ into $[\mathcal{S}, b[m-n]]$ where $b = d^n$ (if $m < n$, we take $b = d$, so $[\mathcal{S}, d[m]]$ remains unchanged). It is easy to verify the existence of this procedure. Facts (89), (91) and (93) make it clear how to construct the desired scientist Ψ_0. We provide an informal description. Let E be an enumeration of $N \times N$.

Given incoming environment e, Ψ_0 works in stages to create ever longer initial segments $e^0, e^1, \ldots$ of e, where e^i is the environment that "starts over" at position i of e. At stage zero we have $\Psi_0(\emptyset) = \Psi(\emptyset)$. Between two stages, while not enough of e is available to proceed to the next stage, Ψ_0 repeats its last conjecture on each new position in e. At the n^{th} stage ($n \geq 1$), Ψ_0 has examined enough of e to construct the initial segments of length n for each of $e^0, \ldots, e^n$. Let these initial segments be denoted $\sigma^0 \ldots \sigma^n$. Let (x, y) be the first pair enumerated in E such that $x, y \leq n$ and Ψ's score on σ^x is y. Then at the n^{th} stage Ψ_0 conjectures $\Psi(\sigma^x)$. Intuitively, Ψ_0 looks for a tail of e on which Ψ's conjectures appear eventually to stop changing. For each (x, y) in turn, Ψ_0 conjectures that one such tail is e^x and that Ψ's hypotheses on e^x stopped changing at position y. This conjecture is kept until Ψ's score on some initial segment of e^x is shown to be greater than y.

To see that Ψ_0 X-solves $\mathcal{K}$ on $\mathbf{P}_0$, let structure $\mathcal{S} \in \mathcal{K}$, environment e for $\mathcal{S}$, and $P \in \mathbf{P}_0$ be given. By Fact (89) the probability (according to P) is 1 that there is $n \in N$ such that e^n is also an environment for $\mathcal{S}$ and that Ψ X-solves e^n. Call such a tail of e "good," the others "bad." By Facts (91) and (93) the probability is 1 that Ψ's scores on initial segments of any bad tail e^x eventually defeat any hypothesis (x, y). It is thus straightforward to verify that Ψ_0 converges on e to $\Psi(\sigma)$ for some σ with the following properties:

(a) σ is an initial segment of a good tail e^n of e, and

(b) for all extensions τ of σ in e^n, $\Psi(\tau) = \Psi(\sigma)$.

Since e^n is a good tail, for this σ we have $\Psi(\sigma) \in \mathrm{X}$ and $\mathcal{S} \models \Psi(\sigma)$. Hence with probability 1, Ψ_0 converges on e to a sentence of X true in the underlying structure $\mathcal{S}$. This completes the proof.

We note that straightforward modifications to the foregoing proof demonstrate the following version of Proposition (49).

(94) PROPOSITION: Let $\mathbf{P}_0 \subseteq \mathbf{P}$, recursive $\mathrm{X} \subseteq \mathcal{L}_{sen}$, and collection $\mathcal{K}$ of structures be given. Suppose that some computable scientist Ψ X-solves $\mathcal{K}$ on $\mathbf{P}_0$ with probability greater than 0. Then there is computable scientist Ψ_0, constructible in uniform recursive fashion from Ψ, that X-solves $\mathcal{K}$ on $\mathbf{P}_0$.

References

[Angluin & Smith, 1983] D. Angluin & C. H. Smith. A survey of inductive inference: Theory and methods. *ACM Comput. Surv.*, 15(3):237–269, September 1983.

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

[Angluin, 1988] D. Angluin. Identifying languages from stochastic examples. Technical Report YALEU/DCS/RR-614, Yale University Dept. of Computer Science, New Haven, CT, 1988.

[Anthony & Biggs, 1992] M. Anthony & N. Biggs. *Computational Learning Theory.* Cambridge Tracts in Theoretical Computer Science (30). Cambridge University Press, 1992.

[Baker & McCarthy, 1981] C. L. Baker & J. McCarthy, editors. *The logical problem of language acquisition.* MIT Press, 1981.

[Berwick, 1986] R. Berwick. *The acquisition of syntactic knowledge.* MIT Press, 1986.

[Bickerton, 1981] D. Bickerton. *The Roots of Language.* Karoma, Ann Arbor, MI, 1981.

[Billingsley, 1986] P. Billingsley. *Probability and Measure.* John Wiley & Sons, second edition, 1986.

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

[Blum, 1967] M. Blum. A machine independent theory of the complexity of the recursive functions. *Journal of the Association for Computing Machinery*, 14(2):322–336, 1967.

[Borer & Wexler, 1987] H. Borer & K. Wexler. The maturation of syntax. In T. Roeper & E. Williams, editors, *Parameter Setting.* Reidel, 1987.

[Brown & Hanlon, 1970] R. Brown & C. Hanlon. Derivational complexity and the order of acquisition in child speech. In J. R. Hayes, editor, *Cognition and the Development of Language.* Wiley, 1970.

[Case & Smith, 1983] J. Case & C. Smith. Comparison of identification criteria for machine inductive inference. *Theoret. Comput. Sci.*, 25:193–220, 1983.

[Chang & Keisler, 1977] C. C. Chang & H. J. Keisler. *Model Theory.* North Holland, second edition, 1977.

[Chomsky, 1957] N. Chomsky. *Syntactic Structures.* Mouton & Co., 1957.

[Chomsky, 1965] N. Chomsky. *Aspects of the Theory of Syntax.* MIT Press, 1965.

[Chomsky, 1975] N. Chomsky. *Reflections on Language.* Pantheon, 1975.

[Chomsky, 1986] N. Chomsky. *Knowledge of Language: Its Nature, Origin and Use.* Praeger, 1986.

[Daley & Smith, 1986] R. Daley & C. Smith. On the complexity of inductive inference. *Inform. Control*, 69:12–40, 1986.

[Daley, 1986] R. P. Daley. Inductive inference hierarchies: Probabilistic vs pluralistic. *Lecture Notes in Computer Science*, 215:73–82, 1986.

[Demetras *et al.*, 1986] M. Demetras, K. Post, & C. Snow. Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child Language*, 13:275–292, 1986.

[Earman, 1992] J. Earman. *Bayes or Bust?* MIT Press, Cambridge, MA, 1992.

[Feldman & Goldin-Meadow, 1978] H. Feldman & L. Goldin-Meadow, S.and Gleitman. Beyond herotodus: The creation of language by linguistically deprived deaf children. In A. Lock, editor, *Action, Symbol, and Gesture: The Emergence of Language*. Academic Press, San Diego, 1978.

[Fulk & Jain, 1992] M. A. Fulk & S. Jain. Approximate inference and scientific method. *Inform. Comput.*, 1992. To Appear.

[Fulk *et al.*, 1992] M. Fulk, S. Jain, & D. Osherson. Open problems in systems that learn. *J. of Comput. Syst. Sci.*, 1992.

[Fulk, 1988] M. Fulk. Saving the phenomenon: Requirements that inductive machines not contradict known data. *Inform. Comput.*, 79(3):193–209, 1988.

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

[Gaifman & Snir, 1982] H. Gaifman & M. Snir. Probabilities over rich languages. *Journal of Symbolic Logic*, 47:495–548, 1982.

[Gaifman *et al.*, 1990] H. Gaifman, D. Osherson, & S. Weinstein. A reason for theoretical terms. *Erkenntnis*, 32:149–159, 1990.

[Gleitman, 1986] L. Gleitman. Biological dispositions to learn language. In W. Demopoulos & A. Marras, editors, *Language Learning and Concept Acquisition*. ABLEX, 1986.

[Glymour & Kelly, 1989] C. Glymour & K. Kelly. On convergence to the truth and nothing but the truth. *Philosophy of Science*, 56, 1989.

[Glymour, 1985] C. Glymour. Inductive inference in the limit. *Erkenntnis*, 22:23–31, 1985.

[Glymour, 1992] C. Glymour. *Thinking Things Through*. MIT Press, Cambridge MA, 1992.

[Gold, 1965] E. M. Gold. Limiting recursion. *J. Symbolic Logic*, 30(1):28–48, 1965.

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

[Goldman, 1986] A. Goldman. *Epistemology and Cognition*. Harvard University Press, Cambridge, MA, 1986.

[Hirsh-Pasek *et al.*, 1984] K. Hirsh-Pasek, R. Treiman, & M. Schneiderman. Brown and hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, 11:81–88, 1984.

[Hornstein & Lightfoot, 1981] N. Hornstein & D. Lightfoot, editors. *Explanation in Linguistics*. Longmans, London, 1981.

[Jain & Sharma, 1990a] S. Jain & A. Sharma. Finite learning by a *team*. In *Proc. 3rd Annu. Workshop on Comput. Learning Theory*, pages 163–177. Morgan Kaufmann, San Mateo, CA, 1990.

[Jain & Sharma, 1990b] S. Jain & A. Sharma. Language learning by a team. In M. S. Paterson, editor, *Proceedings of the 17th International Colloquium on Automata, Languages and Programming*, pages 153–166. Springer-Verlag, July 1990.

[Jain *et al.*, 1989] S. Jain, A. Sharma, & J. Case. Convergence to nearly minimal size grammars by vacillating learning machines. In R. Rivest, D. Haussler, & M. Warmuth, editors, *Proc. 2nd Annu. Workshop on Comput. Learning Theory*, pages 189–199. Morgan Kaufmann, San Mateo, CA, 1989.

[Johnson & Newport, 1989] J. S. Johnson & E. L. Newport. Critical period effects in second language learning: The influence of maturational state on the acquisition of english as a second language. *Cognitive Psychology*, 21:60–99, 1989.

[Juhl, 1993] C. Juhl. Bayesianism and reliable scientific inquiry. *Philosophy of Science*, 60:302–319, 1993.

[Kanazawa, 1993] M. Kanazawa. Identification in the limit of categorial grammars. Technical report, Department of Linguistics, Stanford University, 1993.

[Kanazawa, 1994] M. Kanazawa. *Learnable classes of categorial grammars*. PhD thesis, Department of Linguistics,Stanford University, 1994.

[Kapur & Bilardi, 1992] S. Kapur & G. Bilardi. On uniform learnability of language families. *Information Processing Letters*, 44:35–38, 1992.

[Kapur, 1991] S. Kapur. Computational learning of languages. Technical Report TR 91-1234, Department of Computer Science, Cornell University, 1991.

[Kelly & Glymour, 1992] K. T. Kelly & C. Glymour. Inductive inference and theory-laden data. *Journal of Philosophical Logic*, 21(4), 1992.

[Kelly & Glymour, 1993] K.T. Kelly & C. Glymour. Theory discovery from data with mixed quantifiers. *Journal of Philosophical Logic*, 1993.

[Kelly, 1994] K. T. Kelly. *The Logic of Reliable Inquiry*. MIT Press, 1994.

[Kinber, 1994] E. Kinber. Monotonicity versus efficiency for learning languages from texts. Technical Report 94-22, University of Delaware, Department of Computer and Information Sciences, January 1994.

[Kornblith, 1985] H. Kornblith, editor. *Naturalizing Epistemology.* MIT Press, 1985.

[Kugel, 1977] P. Kugel. Induction, pure and simple. *Inform. Control*, 35:276–336, 1977.

[Kuipers, 1987] T. A. Kuipers, editor. *What is Closer-to-the-Truth?* Rodopi, Amsterdam, 1987.

[Kurtz & Royer, 1988] S. A. Kurtz & J. S. Royer. Prudence in language learning. In D. Haussler & L. Pitt, editors, *Proc. 1st Annu. Workshop on Comput. Learning Theory*, pages 143–156. Morgan Kaufmann, Inc., 1988.

[Landau & Gleitman, 1985] B. Landau & L. Gleitman. *Language and Experience.* Harvard University Press, 1985.

[Lasnik, 1989] H. Lasnik. On certain substitutes for negative data. In R. J. Matthews & W. Demopoulos, editors, *Learnability and Linguistic Theory.* Kluwer, 1989.

[Lenneberg, 1967] E. Lenneberg. *Biological Foundations of Language.* Wiley, 1967.

[Levy, 1979] A. Levy. *Basic Set Theory.* Springer-Verlag, Berlin, 1979.

[Lightfoot, 1982] D. Lightfoot. *The Language Lottery.* MIT Press, 1982.

[Machtey & Young, 1978] M. Machtey & P. Young. *An Introduction to the General Theory of Algorithms.* North-Holland, New York, 1978.

[Matthews, 1984] R. J. Matthews. The plausibility of rationalism. *Journal of Philosophy*, 81:492–515, 1984.

[Matthews, 1989] R. J. Matthews. Learnability and linguistic theory. In R. J. Matthews & W. Demopoulos, editors, *Learnability and Linguistic Theory.* Kluwer, 1989.

[Mazurkewich & White, 1984] I. Mazurkewich & L. White. The acquisition of dative-alternation: unlearning overgeneralizations. *Cognition*, 16(3):261–283, 1984.

[Newport & Supalla, 1989] E. L. Newport & T. Supalla. A critical period effect in the acquisition of a primary language. *Science*, 1989.

[Newport *et al.*, 1977] E. Newport, L. Gleitman, & H. Gleitman. Mother i'd rather do it myself: Some effects and noneffects of maternal speech style. In C. Snow & C. Ferguson, editors, *Talking to children: Language input and acquisition.* Cambridge University Press, 1977.

[Osherson & Wasow, 1976] D. Osherson & T. Wasow. Species specificity and task specificity in the study of language: a methodological note. *Cognition*, 4(2):203–214, 1976.

[Osherson & Weinstein, 1982a] D. Osherson & S. Weinstein. A note on formal learning theory. *Cognition*, 11:77–88, 1982.

[Osherson & Weinstein, 1982b] D. N. Osherson & S. Weinstein. Criteria of language learning. *Inform. Control*, 52:123–138, 1982.

[Osherson & Weinstein, 1986] D. N. Osherson & S. Weinstein. Identification in the limit of first order structures. *Journal of Philosophical Logic*, 15:55–81, 1986.

[Osherson & Weinstein, 1989a] D. Osherson & S. Weinstein. Identifiable collections of countable structures. *Philosophy of Science*, 18:1–42, 1989.

[Osherson & Weinstein, 1989b] D. Osherson & S. Weinstein. Paradigms of truth-detection. *Journal of Philosophical Logic*, 18:1–42, 1989.

[Osherson & Weinstein, 1993] D. Osherson & S. Weinstein. Relevant consequence and scientific discovery. *Journal of Philosophical Logic*, 22:437–448, 1993.

[Osherson & Weinstein, in pressa] D. Osherson & S. Weinstein. On the danger of half-truths. *Journal of Philosophical Logic*, in press.

[Osherson & Weinstein, in pressb] D. Osherson & S. Weinstein. On the study of first language acquisition. Journal of Mathematical Psychology, in press.

[Osherson *et al.*, 1982] D. Osherson, M. Stob, & S. Weinstein. Learning strategies. *Inform. Control*, 53(1):32–51, 1982.

[Osherson *et al.*, 1984] D. N. Osherson, M. Stob, & S. Weinstein. Learning theory and natural language. *Cognition*, 17(1):1–28, 1984.

[Osherson *et al.*, 1986a] D. Osherson, M. Stob, & S. Weinstein. Aggregating inductive expertise. *Inform. Control*, 70(1):69–95, 1986.

[Osherson *et al.*, 1986b] D. Osherson, M. Stob, & S. Weinstein. Analysis of a learning paradigm. In W. Demopoulos & A. Marras, editors, *Language Learning and Concept Acquisition.* Ablex Publ. Co., New Jersey, 1986.

[Osherson *et al.*, 1986c] D. N. Osherson, M. Stob, & S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists.* MIT Press, Cambridge, MA, 1986.

[Osherson *et al.*, 1988] D. Osherson, M. Stob, & S. Weinstein. Synthesising inductive expertise. *Inform. Comput.*, 77:138–161, 1988.

[Osherson *et al.*, 1989] D. N. Osherson, M. Stob, & S. Weinstein. On approximate truth. In *Proc. 2nd Annu. Workshop on Comput. Learning Theory*, pages 88–101. Morgan Kaufmann, San Mateo, CA, 1989.

[Osherson *et al.*, 1991a] D. Osherson, M. Stob, & S. Weinstein. New directions in automated scientific discovery. *Information Sciences*, 1991.

[Osherson *et al.*, 1991b] D. Osherson, M. Stob, & S. Weinstein. A universal inductive inference machine. *Journal of Symbolic Logic*, 56(2):661–672, 1991.

[Osherson *et al.*, 1992] D. Osherson, M. Stob, & S. Weinstein. A universal method of scientific inquiry. *Machine Learning*, 9:261–271, 1992.

[Pappas, 1979] G. Pappas, editor. *Justification and Knowledge.* Reidel, Dordrecht, 1979.

[Penner, 1987] S. Penner. Parental responses to grammatical and ungrammatical child utterances. *Child Development*, 58:376–384, 1987.

[Pinker, 1989] S. Pinker. Markedness and language development. In R. J. Matthews & W. Demopoulos, editors, *Learnability and Linguistic Theory.* Kluwer, 1989.

[Pinker, 1990] S. Pinker. Language acquisition. In D. Osherson & H. Lasnik, editors, *Invitation to Cognitive Science: Language*, pages 199–241. MIT Press, Cambridge, MA, 1990.

[Pitt & Smith, 1988] L. Pitt & C. Smith. Probability and plurality for aggregations of learning machines. *Inform. Comput.*, 77:77–92, 1988.

[Pitt, 1989] L. Pitt. Probabilistic inductive inference. *J. ACM*, 36(2):383–433, 1989.

[Putnam, 1965] H. Putnam. Trial and error predicates and a solution to a problem of mostowski. *Journal of Symbolic Logic*, 30(1):49–57, 1965.

[Putnam, 1967] H. Putnam. The 'innateness hypothesis' and explanatory models in linguistics. *Synthese*, 17(1):12–22, 1967.

[Putnam, 1975] H. Putnam. Probability and confirmation. In *Mathematics, Matter and Method.* Cambridge University Press, 1975.

[Putnam, 1980] H. Putnam. 'what is innate and why,. In M. Piatelli-Palmarini, editor, *The Debate between Jean Piaget and Noam Chomsky.* Harvard University Press, 1980.

[Rogers, 1967] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, New York, 1967.

[Royer, 1986] J. Royer. Inductive inference of approximations. *Inform. Control*, 70:156–178, 1986.

[Sankoff & Brown, 1976] G. Sankoff & P. Brown. The origins of syntax in discourse: A case study of tok pisin relatives. *Language*, 52:631–666, 1976.

[Schieffelin & Eisenberg, 1981] B. Schieffelin & A. Eisenberg. Cultural variation in children's conversations. In R. Schiefelbusch & D. Bricker, editors, *Early language: Acquisition and Intervention.* University Park Press, Baltimore, 1981.

[Schurz & Weingartner, 1987] G. Schurz & P. Weingartner. Verisimilitude defined by relevant consequence-elements. a new reconstruction of Popper's idea. In T. A. Kuipers, editor, *What is Closer-to-the-Truth?* Rodopi, Amsterdam, 1987.

[Shapiro, 1981] E. Y. Shapiro. A general incremental algorithm that infers theory from facts. In *Proc. of the Seventh Internat. Joint Conf. on Artif. Intell.*, volume 1, pages 446–451. Morgan Kaufmann, 1981.

[Shapiro, 1983] E. Y. Shapiro. *Algorithmic Program Debugging.* MIT Press, 1983.

[Shapiro, 1991] E. Y. Shapiro. Inductive inference of theories from facts. In Jean-Louis Lassez & Gordon Plotkin, editors, *Computational Logic: Essays in honor of Alan Robinson.* The MIT Press, Cambridge MA, 1991.

[Sharma *et al.*, 1995] A. Sharma, S. Jain, J. Royer, E. Martin, D. de Jongh, D. Osherson, & S. Weinstein. *Systems that Learn.* MIT Press, second edition, 1995. To appear.

[Shinohara, 1990] T. Shinohara. Inductive inference from positive data is powerful. In *Proc. 3rd Annu. Workshop on Comput. Learning Theory*, pages 97–110. Morgan Kaufmann, 1990.

[Smullyan, 1961] R. M. Smullyan. *Theory of Formal Systems.* Princeton University Press, 1961.

[Solomonoff, 1964] R. J. Solomonoff. A formal theory of inductive inference: Part 1. *Inform. Control*, 7:1–22, 1964.

[Stich, 1978] S. Stich. Empiricism, innateness, and linguistic universals. *Philosophical Studies*, 33(3):273–286, 1978.

[Truscott & Wexler, 1989] J. Truscott & K. Wexler. Some problems in the parametric analysis of learnability. In R. J. Matthews & W. Demopoulos, editors, *Learnability and Linguistic Theory.* Kluwer, 1989.

[van Riemsdijk & Williams, 1986] H. van Riemsdijk & E. Williams. *Introduction to the Theory of Grammar.* MIT Press, 1986.

[Wasow, 1989] T. Wasow. Grammatical theory. In M. Posner, editor, *Foundations of Cognitive Science.* MIT Press, 1989.

[Wexler & Culicover, 1980] K. Wexler & P. Culicover. *Formal Principles of Language Acquisition.* MIT Press, Cambridge, Mass., 1980.

[Wiehagen, 1978] R. Wiehagen. Characterization problems in the theory of inductive inference. In *Proc. of 5th Colloquium on Automata, Languages, and Programming*, pages 494–508. Springer-Verlag, 1978.

institute *for* logic, language *and* computation

ILLC Research Reports and Technical Notes

Coding for Research Reports: *Series-Year-Number*, with LP = Linguistics and Philosophy of Language; ML = Mathematical Logic and Foundations; CL = Compuational Linguistics; CT = Computation and Complexity Theory; X = Technical Notes.
All previous ILLC-publications are available from the ILLC bureau. For prepublications before 1993, contact the bureau.

LP-93-01 *Martijn Spaan, Parallel Quantification*
LP-93-02 *Makoto Kanazawa, Dynamic Generalized Quantifiers and Monotonicity*
LP-93-03 *Nikolai Pankrat'ev, Completeness of the Lambek Calculus with respect to Relativized Relational Semantics*
LP-93-04 *Jacques van Leeuwen, Identity, Quarrelling with an Unproblematic Notion*
LP-93-05 *Jaap van der Does, Sums and Quantifiers*
LP-93-06 *Paul Dekker, Updates in Dynamic Semantics*
LP-93-07 *Wojciech Buszkowski, On the Equivalence of Lambek Categorial Grammars and Basic Categorial Grammars*
LP-93-08 *Zisheng Huang, Peter van Emde Boas, Information Acquisition from Multi-Agent resources; abstract*
LP-93-09 *Makoto Kanazawa, Completeness and Decidability of the Mixed Style of Inference with Composition*
LP-93-10 *Makoto Kanazawa, Weak vs. Strong Readings of Donkey Sentences and Monotonicity Inference in a Dynamic Setting*
LP-93-11 *Friederike Moltmann, Resumptive Quantifiers in Exception Sentences*
LP-93-12 *Jaap van der Does, On Complex Plural Noun Phrases*
LP-93-13 *Natasha Alechina, Binary Quantifiers and Relational Semantics*
LP-93-14 *Mati Pentus, Lambek Calculus is L-complete*
LP-93-15 *David Ian Beaver, What comes first in Dynamic Semantics*

ML-93-01 *Maciej Kandulski, Commutative Lambek Categorial Grammars*
ML-93-02 *Johan van Benthem, Natasha Alechina, Modal Quantification over Structured Domains*
ML-93-03 *Mati Pentus, The Conjoinablity Relation in Lambek Calculus and Linear Logic*
ML-93-04 *Andreja Prijatelj, Bounded Contraction and Many-Valued Semantics*
ML-93-05 *Raymond Hoofman, Harold Schellinx, Models of the Untyped l-calculus in Semi Cartesian Closed Categories*
ML-93-06 *J. Zashev, Categorial Generalization of Algebraic Recursion Theory*
ML-93-07 *A.V. Chagrov, L.A. Chagrova, Algorithmic Problems Concerning First-Order Definability of Modal Formulas on the Class of All Finite Frames*
ML-93-08 *Raymond Hoofman, Ieke Moerdijk, Remarks on the Theory of Semi-Functors*
ML-93-09 *A.S. Troelstra, Natural Deduction for Intuitionistic Linear Logic*
ML-93-10 *Vincent Danos, Jean-Baptiste Joinet, Harold Schellinx, The Structure of Exponentials: Uncovering the Dynamics of Linear Logic Proofs*
ML-93-11 *Lex Hendriks, Inventory of Fragments and Exact Models in Intuitionistic Propositional Logic*
ML-93-12 *V.Yu. Shavrukov, Remarks on Uniformly Finitely Precomplete Positive Equivalences*
ML-93-13 *V.Yu. Shavrukov, Undecidability in Diagonizable Algebras*
ML-93-14 *Dick de Jongh, Albert Visser, Embeddings of Heyting Algebras*
ML-93-15 *G.K. Dzhaparidze, Effective Truth*
ML-93-16 *Maarten de Rijke, Correspondence Theory for Extended Modal Logics*
ML-93-17 *Alexander Chagrov, Michael Zakharyaschev, On the Independent Axiomatizability of Modal and Intermediate Logics*
ML-93-18 *Jaap van Oosten, Extensional Realizability*
ML-93-19 *Raymond Hoofman, Comparing Models of the Non-Extensional Typed l-Calculus*
ML-93-20 *L.A. Chagrova, Dick de Jongh, The Decidability of Dependency in Intuitionistic Propositional Logic*
ML-93-21 *Max I. Kanovich, The Relational Knowledge-Base Interpretation and Feasible Theorem Proving for Intuitionistic Propositional Logic*

ML-93-22 *Andreja Prijatelj, Connectification for n-contraction*

CT-93-01 *Marianne Kalsbeek, The Vanilla Meta-Interpreter for Definite Logic Programs and Ambivalent Syntax*
CT-93-02 *Sophie Fischer, A Note on the Complexity of Local Search Problems*
CT-93-03 *Johan van Benthem, Jan Bergstra, Logic of Transition Systems*
CT-93-04 *Karen L. Kwast, Sieger van Denneheuvel, The Meaning of Duplicates in the Relational Database Model*
CT-93-05 *Erik Aarts, Proving Theorems of the Lambek Calculus of Order 2 in Polynomial Time*
CT-93-06 *Krzysztof R. Apt, Declarative programming in Prolog*
CT-93-07 *Janusz A. Pomykala, Approximation, Similarity and Rough Constructions, Part I. Elementary Introduction*

CL-93-01 *Noor van Leusen, László Kálmán, Computaional Linguistics*
CL-93-02 *Theo M.V. Janssen, An Algebraic View On Rosetta*
CL-93-03 *Patrick Blackburn, Claire Gardent, Wilfried Meyer-Viol, Talking about Trees*

X-93-01 *Paul Dekker, Existential Disclosure, revised version*
X-93-02 *Maarten de Rijke, What is Modal Logic?*
X-93-03 *Michiel Leezenberg, Gorani Influence on Central Kurdish: Substratum or Prestige Borrowing*
X-93-04 *A.S. Troelstra (editor), Metamathematical Investigation of Intuitionistic Arithmetic and Analysis, Corrections to the First Edition*
X-93-05 *A.S. Troelstra (editor), Metamathematical Investigation of Intuitionistic Arithmetic and Analysis, Second, corrected Edition*
X-93-06 *Michael Zakharyashev, Canonical Formulas for K4. Part II: Cofinal Subframe Logics*

ML-94-01 *Domenico Zambella, Notes on polynomially bounded arithmetic*
ML-94-02 *Domenico Zambella, End Extensions of Models of Linearly Bounded Arithmetic*
ML-94-03 *Johan van Benthem, Dick de Jongh, Gerard Renardel de Lavalette, Albert Visser, NNIL, A Study in Intuitionistic Propositional Logic*
ML-94-04 *Michiel van Lambalgen, Independence Structures in Set Theory*
ML-94-05 *V. Kanovei, IST is more than an Algorithm to prove ZFC Theorems*

LP-94-01 *Dimitar Gelev, Introducing Some Classical Elements of Modal Logic to the Propositional Logics of Qualitative Probabilities*
LP-94-02 *Andrei Arsov, Basic Arrow Logic with Relation Algebraic Operators*
LP-94-03 *Jerry Seligman, An algebraic appreciation of diagrams*
LP-94-04 *Kazimierz Świrydowicz, A Remark on the Maximal Extensions of the Relevant Logic* R
LP-94-05 *Natasha Kurtonina, The Lambek Calculus: Relational Semantics and the Method of Labelling*
LP-94-06 *Johan van Benthem, Dag Westerståhl, Directions in Generalized Quantifier Theory*
LP-94-07 *Nataša Rakić, Absolute Time, Special Relativity and* ML^{ν}
LP-94-08 *Daniel Osherson, Scott Weinstein, Dick de Jongh, Eric Martin, Formal Learning Theory*

CT-94-01 *Harry Buhrman and Leen Torenvliet, On the Cutting Edge of Relativization: the Resource Bounded Injury Method*
CT-94-02 *Alessandro Panconesi, Marina Papatriantafilou, Philippas Tsigas, Paul Vitányi, Randomized Wait-Free Distributed Naming*
CT-94-03 *Ming Lee, John Tromp,*
Paul Vitányi, Sharpening Occam's Razor (extended abstract)
CT-94-04 *Ming Lee and Paul Vitányi, Inductive Reasoning*

CT-94-05 *Tao Jiang, Joel I. Seiferas, Paul M.B. Vitányi, Two heads are Better than Two Tapes*
CT-94-06 *Guido te Brake, Joost N. Kok, Paul Vitányi, Model Selection for Neural Networks: Comparing MDL and NIC*
CT-94-07 *Charles H. Bennett, Péter Gács, Ming Li, Paul M.B. Vitányi, Wojciech H. Zurek, Thermodynamics of Computation and Information Distance*
CT-94-08 *Krzysztof R. Apt, Peter van Emde Boas and Angelo Welling, The STO-problem is NP-hard*
CT-94-09 *Klaus Ambos-Spies, Sebastiaan A. Terwijn, Zheng Xizhong, Resource Bounded Randomness and Weakly Complete Problems*
CT-94-10 *Klaus Ambos-Spies, Hans-Christian Neis, Sebastiaan A. Terwijn, Genericity and Measure for Exponential Time*

X-94-01 *Johan van Benthem, Two Essays on Semantic Modelling*
X-94-02 *Vladimir Kanovei, Michiel van Lambalgen, Another Construction of Choiceless Ultrapower*
X-94-03 *Natasha Alechina, Michiel van Lambalgen, Correspondence and Completeness for Generalized Quantifiers*

Titles in the ILLC Dissertation Series:

1993-1 *Transsentential Meditations; Ups and downs in dynamic semantics, Paul Dekker*
1993-2 *Resource Bounded Reductions, Harry Buhrman*
1993-3 *Efficient Metamathematics, Rineke Verbrugge*
1993-4 *Extending Modal Logic, Maarten de Rijke*
1993-5 *Studied Flexibility, Herman Hendriks*
1993-6 *Aspects of Algorithms and Complexity, John Tromp*
1994-1 *The Noble Art of Linear Decorating, Harold Schellinx*
1994-2 *Generating Uniform User-Interfaces for Interactive Programming Environments, Jan Willem Cornelis Koorn*
1994-3 *Process Theory and Equation Solving, Nicoline Johanna Drost*
1994-4 *Calculi for Constructive Communication, a Study of the Dynamics of Partial States, Jan Jaspars*
1994-5 *Executable Language Definitions, Case Studies and Origin Tracking Techniques, Arie van Deursen*