Learning Theory and Descriptive Set Theory

by

Kevin T. Kelly

May 1992

Report CMU-PHIL-27

Carnegie Mellon

Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

Learning Theory and Descriptive Set Theory

Kevin T. Kelly
Department of Philosophy
Carnegie Mellon University

In 1965, E. M. Gold and H. Putnam observed independently that recursive operators can be viewed as mechanical scientists trying to investigate a given hypothesis in the limit [1], [12]. Gold then essentially characterized the hypotheses that mechanical scientists can successfully decide in the limit in terms of arithmetic complexity. These ideas were developed still further by Peter Kugel [4]. In this paper, I extend this approach to obtain characterizations of identification in the limit, identification with bounded mind-changes, and identification in the short run, both for computers and for ideal agents with unbounded computational abilities. The characterization of identification with n mind-changes entails, as a corollary, an exact arithmetic characterization of Putnam's n-trial predicates, which closes a gap of a factor of two in Putnam's original characterization [12].

It will be shown that solvability results concerning identification problems can be viewed as estimations of complexity for second-order relations; arithmetic complexity when the scientist is effective, and Borel complexity otherwise. This general perspective illuminates the relationships between the various learning-theoretic paradigms, since function identification, language identification, and logical theory identification all drop out as special cases.

The hierarchy-theoretic approach of Gold and Putnam has an additional advantage. Traditional negative arguments in learning theory have usually required that the learner succeed regardless of the order in which observations are presented [1]. All of the standard characterization theorems assume this requirement [8] [9] [10] [11]. One technical reason for the assumption is that the negative sides of these characterization theorems are established using variants of a diagonal argument known as the *locking sequence lemma* [11]. By relativizing the recursion-theoretic approach of Gold and Putnam, we can obtain more general characterizations of identifiability that apply no matter what the scientist knows *a priori* about data ordering. The operative notion of relativization is not relativization to an oracle as is usual in recursion theory, but rather relativization to a range of possible inputs representing the agent's background knowledge, as is more common in statistics.

The learning-theoretic interpretation of the apparatus of the theory of recursive operators has some pedagogical advantages. For example, it will be shown how the recursion-theoretic basis theorems answer questions about how the difficulty of empirical science relates to the computational difficulty of the theory under investigation. The learning-theoretic interpretation of recursion theory also raises questions about the status of classical results (such as the basis theorems) under changes in background knowledge.

1. Data Presentations, Hypotheses, and Background Knowledge

I assume that the total evidence received by the scientist at a given time can be encoded effectively by a natural number. The data presentation received by an empirical scientist is potentially infinite, so a ***data presentation*** is as an co-sequence of natural numbers. Hypotheses will also be viewed as discrete objects effectively encoded by natural numbers.

There are many features that scientists would like their hypotheses to have, including simplicity, unity, empirical adequacy, and so forth. To avoid interminable debates about the precise nature of these requirements, we will assume only that there is *some* well-defined relation **A c** co® X co of ***hypothesis adequacy*** holding between data presentations and code numbers of hypotheses. Since the aim of inquiry will be to determine whether a hypothesis is adequate, we may identify a hypothesis with the set H of all infinite data presentations for which it is adequate. So a ***hypothesis*** is just some arbitrary subset of co^{03}. Let $x_H(t)$ be the characteristic function of H.

Background knowledge restricts the scientist's *a priori* uncertainty about the data presentation he will ultimately see in the limit. So we may also think of ***background knowledge*** K as some arbitrary subset of co^{03}.

2. Paradigms of Hypothesis Assessment

A problem of hypothesis assessment is a situation in which a scientist is given a hypothesis and is asked to assess its adequacy on the basis of increasing data fed from some infinite data presentation. Let co* be the set of all finite sequences of natural numbers, which we will think of as finite sequences of observations. An ***assessment method*** is a function that takes a finite data segment to some guess about the adequacy of the hypothesis in question. The guess 1 means that the hypothesis in question is adequate, 0 means that it is not, and # represents a refusal to produce a judgment of adequacy.

$\phi: \omega^* \rightarrow \{1, 0, \#\}$.

Now we will consider a sequence of increasingly lenient notions of reliable success for hypothesis assessment methods.

ϕ ***verifies*** H ***over*** K ***in the limit*** $\Leftrightarrow$
$\forall t \in K\ A(t, i) \Leftrightarrow (\exists n\ \forall m \geq n\ \ \phi(t|m) = 1)$.

ϕ ***refutes*** H ***over*** K ***in the limit*** $\Leftrightarrow$
$\forall t \in K\ \neg A(t, i) \Leftrightarrow (\exists n\ \forall m \geq n\ \ \phi(t|m) = 0)$.

ϕ ***decides*** H ***over*** K ***in the limit*** $\Leftrightarrow$
ϕ both verifies and refutes H over K in the limit.

ϕ ***decides*** H ***over*** K ***in*** n ***mind-changes starting with*** 1 [0, #] $\Leftrightarrow$
ϕ decides H over K in the limit and
$\phi(t|0) = 1$ [0, #] and
$mc(\phi, t) \leq n$;

where

$mc(\phi, t|0) = 0$ and
$mc(\phi, t|n{+}1) = mc(\phi(t|n))$ if $\phi(t|n) = \phi(t|n{+}1)$ and
$mc(\phi, t|n{+}1) = mc(\phi(t|n)) + 1$ otherwise.

ϕ ***verifies*** H ***over*** K ***with certainty*** $\Leftrightarrow$
$\forall t \in K,\ t \in H \ \Leftrightarrow (\exists n$ s.t. $\phi(t|n) = 1$ and $\forall m < n, \phi(t|n) = \#)$.

ϕ ***refutes*** H ***over*** K ***with certainty*** $\Leftrightarrow$
$\forall t \in K,\ t \in \overline{H} \Leftrightarrow (\exists n$ s.t. $\phi(t|n) = 0$ and $\forall m < n, \phi(t|n) = \#)$.

ϕ ***decides*** H ***over*** K ***with certainty*** $\Leftrightarrow$
ϕ both verifies and refutes H over K with certainty.

$$\text{H } \textbf{\textit{is}} \left[\textbf{\textit{effectively}}\right] \begin{bmatrix} \textbf{\textit{decidable}} \\ \textbf{\textit{verifiable}} \\ \textbf{\textit{refutable}} \end{bmatrix} \textbf{\textit{ over }} \text{K} \begin{bmatrix} \textbf{\textit{with certainty}} \\ \textbf{\textit{with n mind-changes}} \\ \quad \textbf{\textit{starting with }} 0\ [1, *] \\ \textbf{\textit{in the limit}} \end{bmatrix} \Leftrightarrow$$

$$\exists \left[\text{total recursive}\right] \phi \text{ s.t. } \phi \begin{bmatrix} \text{decides} \\ \text{verifies} \\ \text{refutes} \end{bmatrix} \text{H over K} \begin{bmatrix} \text{with certainty} \\ \text{with n mind-changes} \\ \quad \text{starting with } 0\ [1, *] \\ \text{in the limit} \end{bmatrix}$$

3. Characterizations of Reliable Hypothesis Assessment

For each $\sigma \in \omega^*$, let $B_\sigma = \{t \in \omega^\omega$: t extends $\sigma\}$. Call B_σ the ***fan*** with ***handle*** σ. R is a ***type*** <k, j> ***relation*** $\Leftrightarrow R \subseteq (\omega^\omega)^k \times \omega^j$. Let R be a type <k, j> relation. R is ***basic open*** $\Leftrightarrow$ $R = \varnothing$ or $\exists$ fans

$F_1, ..., F_k$, $\exists S_1, ..., S_j \subseteq \omega$ s.t. $R = F_1 \times ... \times F_k \times S_1 \times ... \times S_n$. R is ***open*** $\Leftrightarrow$ R is a union of basic open relations of type <k, j>. R is ***closed*** $\Leftrightarrow \overline{R}$ is open. R is ***clopen*** $\Leftrightarrow$ R is closed and open.

The ***finite Borel hierarchy*** is defined as follows[1], where R is assumed to be of type <k, j>.

$R \in \Sigma_0^B \Leftrightarrow$ R is clopen.

$R \in \Sigma_{n+1}^B \Leftrightarrow \exists P \in \Sigma_n^B$ s.t. (1) P is type <k, j+1> and
(2) $R = \{\langle \mathbf{t}, \mathbf{x} \rangle \in (\omega^\omega)^k \times \omega^j : \exists n \in \omega \text{ s.t. } \neg P(\mathbf{t}, \mathbf{x}, n)\}$,

$R \in \Pi_n^B \Leftrightarrow \overline{R} \in \Sigma_n^B$.

$R \in \Delta_n^B \Leftrightarrow R \in \Sigma_n^B \cap \Pi_n^B$.

Now we proceed to define the arithmetic hierarchy. Let R be a relation of type <k, j>. Turing machine M is a ***decision procedure*** for R over K $\Leftrightarrow \forall \mathbf{t} \in K^k, \forall \mathbf{x} \in \omega^j$, if $R(\mathbf{t}, \mathbf{x})$ then M halts with 1 after receiving **x** as input and after scanning some finite segment of each t occurring in **t**, and M halts with 0 otherwise. R is ***recursive*** $\Leftrightarrow$ R has some decision procedure. The ***arithmetic hierarchy*** starts out with recursive rather than arbitrary, clopen relations:

$R \in \Sigma_0^0 \Leftrightarrow$ R is recursive..

The inductive clause of the definition proceeds just as before, with superscript 0 replacing the superscript B. Observe that in either case,

$$\Sigma_0^0 = \Pi_0^0 = \Delta_0^0 = \Delta_1^0 \qquad \Sigma_0^B = \Pi_0^B = \Delta_0^B = \Delta_1^B$$

We may relativize either hierarchy to background knowledge $K \subseteq \omega^\omega$ and to a range $H \subseteq \omega$ of hypotheses of interest in the following manner. Let Γ be a complexity class in either hierarchy. Let R be of type <k, j>.

$R \in \Gamma^K \Leftrightarrow \exists R' \in \Gamma$ s.t. R' is type <k, j> and $\forall \mathbf{t} \in K^k\ \forall \mathbf{n} \in \omega^j$, $R(\mathbf{t}, \mathbf{n}) \Leftrightarrow R'(\mathbf{t}, \mathbf{n})$.

When Γ already has a superscript, we add K as a second superscript, as follows: $\Pi_2^{B,K} = \left[\Pi_2^B\right]^K$.

---

[1]This definition is a special case of [6], p. 20.

Theorem 3.1 (Gold and Putnam): Let K, H $\subseteq \omega^{\omega}$. Then we have:

(a) H is [effectively] decidable over K with certainty $\Leftrightarrow$ H $\in \Delta_1^{B,K} \left[\Delta_1^{0,K}\right]$.

(b) H is [effectively] verifiable over K with certainty $\Leftrightarrow$ H $\in \Sigma_1^{B,K} \left[\Sigma_1^{0,K}\right]$.

(c) H is [effectively] refutable over K with certainty $\Leftrightarrow$ H $\in \Pi_1^{B,K} \left[\Pi_1^{0,K}\right]$.

(d) H is [effectively] decidable over K in the limit $\Leftrightarrow$ H $\in \Delta_2^{B,K} \left[\Delta_2^{0,K}\right]$.

(e) H is [effectively] verifiable over K in the limit $\Leftrightarrow$ H $\in \Sigma_2^{B,K} \left[\Sigma_2^{0,K}\right]$.

(f) H is [effectively] refutable over K in the limit $\Leftrightarrow$ H $\in \Pi_2^{B,K} \left[\Pi_2^{0,K}\right]$.

Proof: The effective parts of (c) and (d) are established in [1], and the effective part of (c) is shown in [12]. The ineffective cases follow by the same arguments with references to computability suppressed. (a) and (b) are trivial. ■

4. Some Applications

4. 1: The Empirical Irony of Cognitive Science

Cognitive scientists frequently assume that human behavior can be modelled by a computer. Some philosophers have challenged this assumption on *a priori* grounds. Let us approach the question empirically, rather than dogmatically. Imagine a given system emitting a sequence of outputs, and consider the hypothesis H_{rec}, that the system will produce a recursive sequence. H_{rec} = {t: t is a recursive sequence}. Suppose we don't know what to expect out the system, so K = ω^{ω}. Each singleton {t} is closed and since H_{rec} is countable, $H_{rec} \in \Sigma_n^B$, so a non-effective scientist can verify H_{rec} in the limit, by theorem 3.1. But $\overline{H_{rec}}$ is not verifiable in the limit, by a simple diagonal argument (each finite segment of a non-computable sequence can be extended by a computable sequence). So $H_{rec} \in \Sigma_2^B - \Pi_2^B$, by theorem 3.1. A standard fact of recursion theory is that $H_{rec} \in \Sigma_3^0 - \Pi_3^0$. Hence, by theorem 3.1, H_{rec} is not even verifiable in the limit by a computer. The "irony" is this. If human nature is computable, human scientists cannot verify this fact even in the limit because they are computable. But if human nature is not computable, human scientists cannot verify the non-computability of human nature in the limit either, because no system could, ideal or otherwise.

4. 2: Basis Theorems and Hypothesis Complexity

A complete hypothesis specifies everything that will ever be seen, in the correct order. Hence, a complete hypothesis is a singleton {t}. In logic, the arithmetic complexity of {t} is known as the *implicit* complexity of t and the arithmetic complexity of t (viewed as a set of ordered pairs) is known as the *explicit* complexity of t. From a learning theoretic perspective, the implicit complexity of t is just the complexity of investigating the complete hypothesis {t}, whereas the explicit complexity of t is just the computational difficulty of generating the nth prediction specified by t.

Since science is often conceived of as a process of deductively checking the consequences of a theory against the observed data, a natural question is: how much worse than the empirical complexity of a hypothesis can its deductive complexity be if a computable scientist is to determine whether it is correct? More specifically, how impossible can it be to derive predictions from h before it becomes impossible to verify h in the limit, or to refute h with certainty? It turns out that the basis theorems of mathematical logic already provide surprising answers to this question in many cases.

The first result is intuitive: if an effective scientist is to refute a theory making only finitely many different kinds of predictions, then there *must* be a mechanical procedure for deriving each successive prediction entailed by the theory, so that these predictions may be compared against the data.

Theorem 4.2.1: If {t} is effectively refutable with certainty and rng(t) is finite then t is recursive.
Proof: [3, p. 79] and Theorem 3. ^

This intuitive picture may be extended to the case of verifiability in the limit, so long as the theory in question makes only finitely many different kinds of predictions.

Theorem 4.2.2: If {t} is effectively verifiable in the limit and mg(t) £ {1, 0} then t is recursive.
Proof: Kreisel's basis theorem [3, p. 108] and Theorem 3.1. B

The requirement that rng(t) be finite is essential, for there exist theories that predict infinitely many different kinds of events that can be effectively refuted with certainty whose predictions are *infinitely* hard to derive:

Theorem 4.2.3: There is a non-arithmetic t s.t. {t} is effectively refutable.
Proof: [3, p. 107] and Theorem 3.1. •

Theorem 4.2.3 shows that computerized empirical inquiry can do *infinitely better* in some cases if it does *not* follow the intuitive method of checking successive predictions of a theory against the data. This startling result is softened somewhat by the fact that the "magic" method that succeeds will not notice immediately that {t} has been refuted, whereas an ineffective method that can magically deduce predictions from {t} would notice right away. To see this, say that

H *Is* ***consistent*** *with* *a e* ω* over K <=> K n H n B$_a$ * 0.

Φ ***is vigilant about*** H *over* K «
Vt € K Va c t, H is inconsistent with a over K =* 0(a) = 0.

A vigilant method says 0 as soon as the hypothesis under investigation becomes inconsistent with the data. Since the method of Theorem 4.2.3 effectively refutes {t} with certainty and t is not recursive, we have by the following fact that this method is not vigilant.

Fact 4.2.4:
If 0 is recursive and 0 is vigilant about {t} over K and 0 refutes {t} with certainty over K then t is recursive.

Proof: Let 0 be as required. To compute t, we proceed as follows on input n: Stage 0: run 0(<O>), <t>(<1>),... until some <m> is found on which 0 does not return 0. Since 0 is vigilant, t(0) = 1. Inductively, let a[k] be the path constructed by stage k. At stage k+1, run 0(a*<O>), 0(a[k]#<1>),... until some <m> is found such that 0(a*<m>) does not return 0. Again, vigilance guarantees that t(k) = m. When stage n is reached, return ofn]$_n$. H

When background knowledge is increased, it becomes easier for effective science to succeed at investigating deductively intractable theories. In particular, if K = {t}, then no matter how complex t is, {t} can be decided with certainty. It remains an interesting question to determine what background knowledge must be like for therems 4.2.1 and 4.2.2 to hold.

5. Characterizations of Hypothesis Assessment with Bounded Mind-Changes

It remains to characterize empirical decidability with n mind-changes. Putnam [12] discusses a closely related notion under the rubric of ***n-trial predicates***, but his characterization leaves a gap of a factor of two between its upper and lower bounds.[2] I will present an exact characterization in terms of the ***difference hierarchy***[3], a finitary version of the Borel and Arithmetic hierarchies. The topological version will be indexed with MC and the computational version will be indexed with mc, for "mind-changes".

$R \in \Sigma_0^{MC} \Leftrightarrow R = \varnothing$.

$R \in \Sigma_{n+1}^{MC} \Leftrightarrow$ R is of form $\overline{S} \cap O$, where $S \in \Sigma_n^{MC}$ and O is open.

$R \in \Pi_n^{MC} \Leftrightarrow \overline{R} \in \Sigma_n^{MC}$.

$R \in \Delta_n^{MC} \Leftrightarrow R \in \Sigma_n^{MC} \cap \Pi_n^{MC}$.

$MC = \bigcup_{n \in \omega} \Delta_n^{MC}$.

The only difference between the mc hierarchy and the MC hierarchy is that we replace open sets with RE sets in the second clause of the definition:

$R \in \Sigma_{n+1}^{mc} \Leftrightarrow$ R is of form $\overline{S} \cap O$, where $S \in \Sigma_n^{mc}$ and O is RE.

The MC hierarchy can be verified to satisfy the following closure laws:

Proposition 5.1:

Let O be open and let C be closed. Then we have:

(a) For each n: If $S \in \Pi_n^{MC}$ then $\overline{S} \in \Sigma_n^{MC}$. $S \cup C \in \Pi_n^{MC}$; If $S \in \Sigma_n^{MC}$ then $\overline{S} \in \Pi_n^{MC}$; $S \cap O \in \Sigma_n^{MC}$;

---

[2]Putnam did not state the exact characterization as a question, and did not require an exact characterization for the purposes of his paper.

[3]The idea is implicit in [5], p. 96. I am indebted to J. Tappendon for this reference.

$S \cap O \in \Sigma^{MC}_{n+1}$. $S \cup C \in \Pi^{MC}_{n+1}$.

(b) for each odd n, If S e rtf^{c} then If S e I??^{0} then

$S \cap C \in \Pi^{MC}_{n}$; $S \cup O \in \Sigma^{MC}_{n}$;

$S \cup O \in \Pi^{MC}_{n+1}$, $S \cap C \in \Sigma^{MC}_{n+1}$.

(c) for each even n, If S e nir then If S e ST then

SuOenf!^{10}; *SnCe* Σ^{MC}_{n};

$S \cap C$,*en^;*^{MC} SuOeAiS.

and similarly for the me hierarchy when O is RE and C is Co-RE. ▪

Now it is possible to characterize empirical decidability with at most n mind-changes.

Theorem 5.2: For each $n \geq 0$,

(a) H is [effectively] decidable over K in n mind-changes starting with 1 « H e In^{MC} K[*I^Cy* K J.

(b) H is [effectively] decidable over K in n mind-changes starting with 0 ⇔ H € n$_{n}$ L^{n}n J.

(c) H is [effectively] decidable over K in n mind-changes starting with # « H e A$_{n}$ MC K[[A$_{n}$ me KI J.

(d) H is [effectively] decidable over K in n mind-changes » H e In^{MC} K u nJi MC, K [L^{2}^ K u *rff*^{1} K J.

Proof: (a) => Suppose that [recursive] ϕ decides H over K in n mind-changes starting with 0. Define O(t, n) » mc(ϕ, t) $\geq$ n and define C(t, n) ⇔ mc(ϕ, t) $\leq$ n. O(t, n) is K-open [K-RE] and C(t, n) is K-closed [K-Co-RE], First, let's consider the case when n is even. Then since 0 always starts with conjecture 0 and never uses more than n mind-changes over K, we have:

Vt e K, t e H «

0 changes its mind some odd number of times $\leq$ n-1 ⇔

(O(t, 1) & C(t, 1)) v (O(t, 3) & C(t, 3)) v ... v (O(t, n -1) & O(t, n -1)) «

[O(t, 1) & C(t, 1)] v *[Oft, 1)* & Oft 3) & Cft 3)] v *[Oft, 1) & Oft, 3)* & Oft 5) & Cft 5)] v ...

v *[Oft, 1) & Oft, 3) & Oft 5) & ... & Oft, n-3)* & Oft n -1) & Cft n -1)] «

Oft 1) & [Cft 1) v [Oft 3) & [Cft 3) v ... v [Oft n-1) & Cft n-1)]]]] (by factoring).

which is a 2%°K⌈ /^ K⌉ J property of t.

Now for the case in which n is odd. Since 0 starts out with conjecture 0 and never uses more than n mind-changes over K, we have

Vt e K, t e H «

ф does not change its mind some even number of times ≤ n ⇔

-.C(t, 0) & -.(O(t, 2) & C(t, 2)) & ... & -.(O(t, n-1), C(t, n-1)) »

O(t, 1) & [C(t, 1) v O(t, 3)] & [C(t, 3) v O(t, 5)] & [C(t, 5) v O(t, 7)] & ...& [C(t, n-3) v O(t, n-1)] «*

O(t, 1) & [C(t, 1) v O(t. 3)] & *[C(t, 1)* v C(t, 3) v O(t, 5)] & *[C(t, 1) v C(t, 3)* v C(t, 5) v O(t, 7)] & ...&

[C(t, 1) v C(t, 3) v... v C(t, n- 5) v... v C(t, n-3) v O(t, n-1)] ⇔

O(t, 1) & [C(t, 1) v [O(t, 3) & [C(t, 3) v... & [C(t, n-3) v O(t, n-1)]]]] (by factoring)

which is a ^ _.MC, K ⌈ L 2^ _mc, KI ⌉ J property of t.

⇐ Suppose that H e *** ^MC.Kf L 2* _mcKJ J. Then H may be defined over K in the form

Oi n [C2 u [O3 n [C4 u ...[C_n-1 u O_n] if n is odd, or in the form

O1 n [C2 u [O3 n [C4 u ..[C_n-1 n O_n] if n is even.

In either case, define $ to conjecture 0 until Oi is verified by the data, after which ф says 1 until C2 is refuted by the data, after which 0 says 0 until O3 is verified by the data, after which 0 says 1 until.... 0 will succeed with at most n mind-changes.

(b) follows from (a) by duality.

(c) => Suppose that 0 decides H over K with n mind-changes starting with #. Define

Vo(a)= { Oif0(a) = # ; 0(a) otherwise }   VI(CT)= { 1 if0(a) = # ; 0(a) otherwise }

YO succeeds in n mind-changes starting with 0 and y1 succeeds in n mind-changes starting with 1. By (a), H e ^n MC.K L^n mc,KI J and by (b), He ^{n} n _MC. K Ln$_n$ [_mc, KI J.

$\Leftarrow$ Suppose $H \in \Delta_n^{MC,K}[\Delta_n^{mc,K}]$. Then by (a) and (b), we have [effective] methods ψ_1, ψ_0 that succeed in n mind-changes starting with 1 and with 0, respectively. Define

$$\phi(\sigma)=\begin{cases}\# \text{ if } \sigma \text{ is empty}\\ \psi_1 \text{ if } \psi_1(\sigma)=\psi_0(\sigma)\\ \phi(\sigma-) \text{ otherwise}\end{cases}$$

which [effectively] decides H with n mind-changes starting with #.

(d) $\Leftarrow$ follows from (a) and (b).

$\Rightarrow$ Let [effective] ϕ decide H over K in n mind-changes. Let σ be empty. If $\phi(\sigma) = 1$ then $H \in \Pi_n^{MC,K}[\Pi_n^{mc,K}]$, by (a). If $\phi(\sigma) = 0$ then $H \in \Sigma_n^{MC,K}[\Sigma_n^{mc,K}]$, by (b). If $\phi(\sigma) = \#$, then $H \in \Delta_n^{MC,K}[\Delta_n^{mc,K}]$, by (c). ■

The following proposition is an illustration of theorem 5.2.

Proposition5.3: Let K be the set of all sequences that converge either to 0 or to 1. Then

(a) $MC \subset \Delta_2^B$ $\qquad mc \subset \Delta_2^0$

(b) $\forall n\ \Sigma_n^{MC} \subset \Sigma_{n+1}^{MC}$ $\qquad \Sigma_n^{mc} \subset \Sigma_{n+1}^{mc}$.

(c) $bc(\Sigma_1^{B,K}) = MC^K$ $\qquad bc(\Sigma_1^{0,K}) = mc^K$,

where $bc(\Gamma)$ stands for the finitary Boolean closure of class Γ.

Proof: (b) To show that the inclusion is proper, define #0(t) = the number of 0's occurring in t.

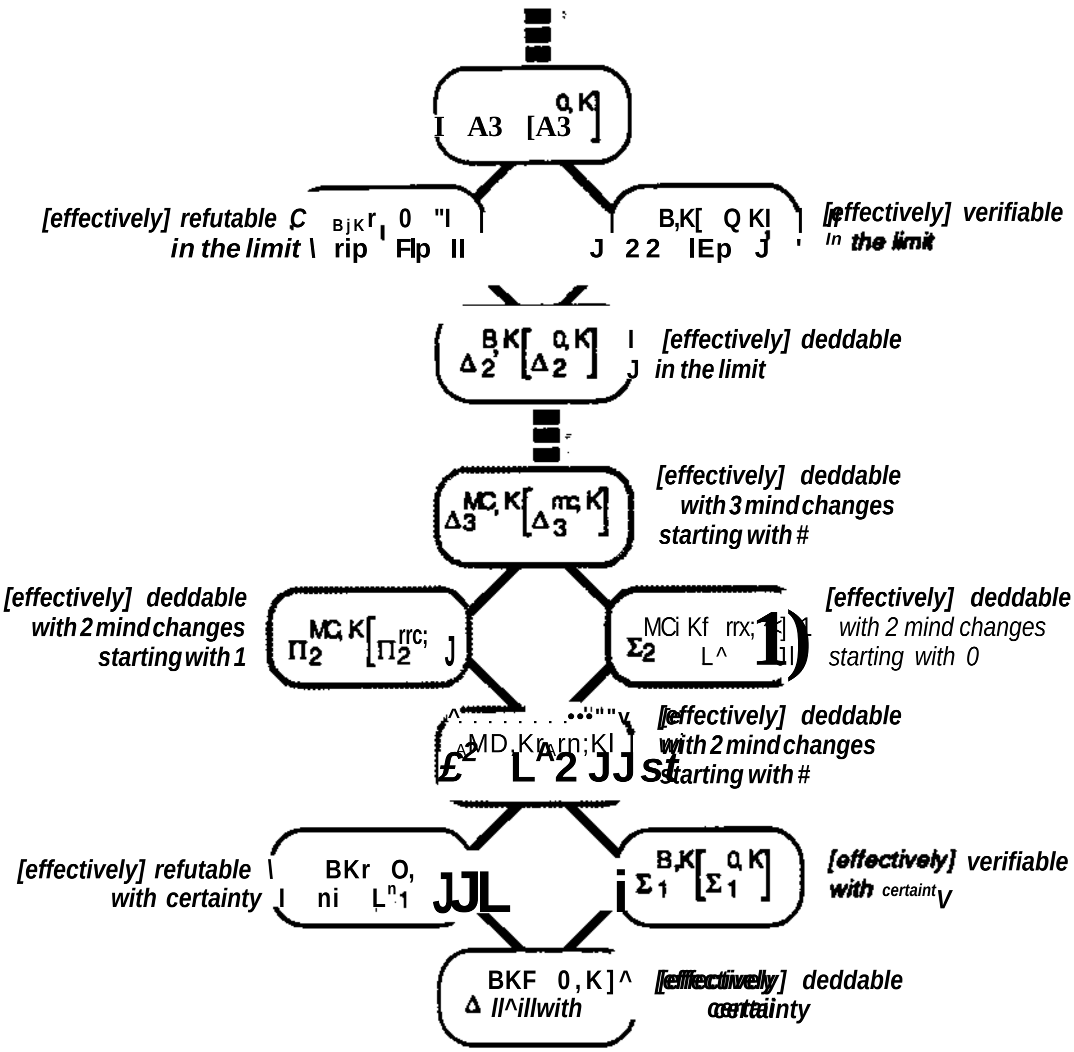
$$P_n(t) = \begin{cases}\left(\bigvee_{k=1}^{n/2} \#0(t) = 2k\right) \text{ if n is even}\\ \left(\left(\bigvee_{k=1}^{(n-1)/2} \#0(t) = 2k\right) \vee \#0(t) \geq n+1\right), \text{ if n is odd}\end{cases}$$

P_n is readily seen to be effectively decidable in n mind-changes starting with 0. A simple diagonal argument shows that P_n cannot be decided with fewer mind-changes starting with 0. Now apply Theorem 5.2.

(a) Let P_n be defined as in the proof of (b). Define $P(t) \Leftrightarrow P_{t_1}(t)$. Let ϕ and n be given. Define t_1 = n+1, and now force ϕ to change its mind concerning P at least n+1 times, as in (b). So by theorem 5.2, P is not in any Σ_n^{MC}, so $P \notin$ MC. But it is easy to decide P in the limit by deciding P_{t_1} with t_1 mind changes. By theorem 5.2, $P \in \Delta_2^B$.

(c) Each finite Boolean combination of open sets may be rewritten in disjunctive normal form. But by the closure of open [closed] sets under finite union and intersection, each disjunct can be rewritten in one of the following forms: $(O \cap C)$, O, or C, where O is open and C is closed. Each disjunct of form $(O \cap C)$ is settled in 2 mind changes (say 0 until O is verified; then say 1 until C is refuted). Each O and each C can be handled with one mind change, so the whole disjunction can be handled in some finite number of mind changes. Apply theorem 5.2. ■

Theorems 3.1, and 5.2, together with proposition 5.3 yield the following, complete characterization of the hypothesis assessment problems according to the standards of success introduced at the beginning of this paper.



Theorem 5.2 also yields a characterization of Putnam's n-trial predicates [12].

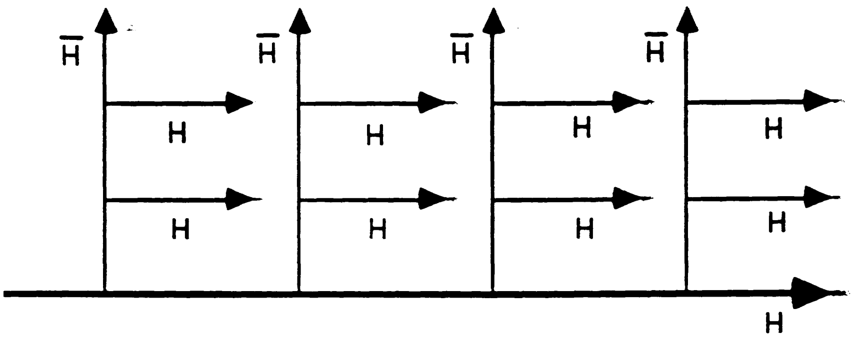
> $S \subseteq \omega$ is an ***n-trial predicate*** ⇔
> there is some total recursive f such that for each $n \in \omega$, $\lim_{k\to\omega} f(n, k)$ converges to $\chi_S(n)$ with at most n mind-changes.

Putnam's n-trial predicates can be viewed as a special kind of empirical hypothesis whose adequacy depends only on the first datum observed. Define $H_S = \{t: t_0 \in S\}$. By theorem 5.2, we have:

Proposition 5.4: S is an n-trial predicate ⇔ $H_S \in \Sigma_2^{mc} \cup \Pi_2^{mc}$. ■

6. Feathers and Demons

The proof of theorem 5.2 hinged on finding a Σ_n^{MC} definition of H in terms of the method ϕ that succeeds in n mind-changes starting with 0. A complementary characterization can be given, that draws into relief the structure underlying diagonal arguments. Suppose that hypothesis H is distributed in K in the following manner.



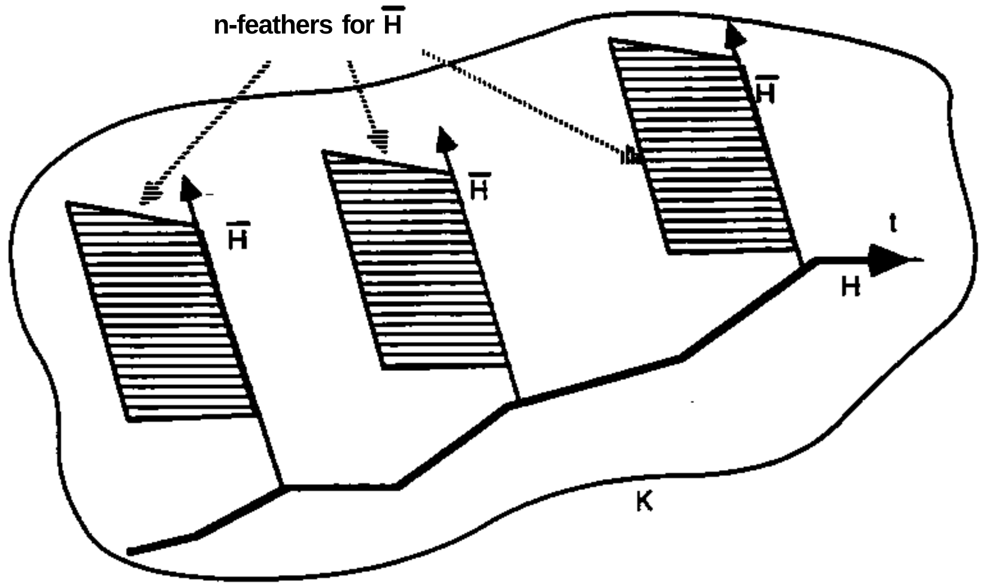
Given the depicted situation, a demon can easily force an arbitrary scientist who starts with 0 to change his mind three times, and can force a scientist who starts with 1 to change his mind twice. The demon leads the scientist down the bold path until ϕ says 1 (which must happen, else the demon stays with the bold path and ϕ fails in the limit). As soon as ϕ says 1, the demon proceeds up the next available path for $\bar{H}$. Now ϕ must eventually say 0, at which time the demon veers to the right down the next available path for H.

K may be thought of as an infinite "feather" whose "shaft" is the bold path, and whose alternating "barbs" are the other paths. We may define feathers more generally as follows:

K ***is a*** 1***-feather for*** H ***with shaft*** t $\Leftrightarrow$ t $\in$ K $\cap$ H.

K ***is an*** n+1***-feather for*** H ***with shaft*** t $\Leftrightarrow$
t $\in$ K $\cap$ H and
$\forall$n $\exists$t $\in$ K s.t.
t|n = t'|n and
K is an n-feather with shaft t for $\bar{H}$

K ***is an*** n***-feather for*** H $\Leftrightarrow$ $\exists$t s.t. K is an n-feather for H with shaft t.



We may now define the ***feather dimension*** of K for H:

$D_{H}(K) = n \Leftrightarrow$ K is an n-feather for H and K is not an n+1-feather for H.

Theorem 6.1:

(a) H is decidable over K in n mind-changes starting with 0

$\Leftrightarrow$ K is not an n+1-feather for H

$\Leftrightarrow H \in \Sigma_n^{MC,K}$.

(b) H is decidable over K in n mind-changes starting with 1

$\Leftrightarrow$ K is not an n+1-feather for $\overline{H}$

$\Leftrightarrow H \in \Pi_n^{MC,K}$.

(c) H is decidable over K in n mind-changes starting with #

$\Leftrightarrow$ K is not an n+1-feather for H and K is not an n+1-feather for $\overline{H}$

$\Leftrightarrow H \in \Delta_n^{MC,K}$.

(d) H is decidable over K in n mind-changes

$\Leftrightarrow$ K is not an n+1-feather for H or K is not an n+1-feather for $\overline{H}$

$\Leftrightarrow H \in \Sigma_n^{MC,K} \cup \Pi_n^{MC,K}$.

Proof: (a) & (b) $\Rightarrow$ The contrapositive can be established by means of theorem 6.1 and the usual demonic argument.

$\Leftarrow$ Base case for (a): Suppose that K is not a1-feather for H. Then $H = \emptyset$. Let $t \in K$. Then $t \in \overline{H}$. So the trivial method $\phi_0(\sigma) = 0$ succeeds in 0 mind changes. The base case for (b) is similar.

Now suppose (a) and (b) for each $n' \leq n$. Hence, if K is not an n'+1 feather for $\overline{H}$ then there is a method $\Psi^1_{H,K}$ that decides H over K in n' mind-changes starting with 1, and if K is not an n'+1 feather for H then there is a method $\Psi^0_{H,K}$ that decides H over K in n' mind-changes starting with 0. Let $K_\sigma = \{t \in K: t \text{ extends } \sigma\}$. Then define:

$$\text{trunc}(H, K, n, \sigma) = \begin{cases} \text{the shortest } \tau \subseteq \sigma \text{ s.t. } Dim_H(K_\tau) \leq n, \text{ if there is one} \\ \sigma \text{ otherwise} \end{cases}$$

$$\phi^0_{n+1}(\sigma) = \begin{cases} 0 \text{ if } \sigma \text{ is empty} \\ 0 \text{ if } Dim_H(K_\sigma) > n \text{ and } Dim_{\overline{H}}(K_\sigma) > n \\ \psi^1_{H, K_{trunc(H,K,n,\sigma)}}(\sigma) \text{ if } Dim_{\overline{H}}(K_\sigma) \leq n \\ \psi^0_{H, K_{trunc(H,K,n,\sigma)}}(\sigma) \text{ otherwise (i.e. if } Dim_H(K_\sigma) \leq n) \end{cases}$$

Suppose that K is not an n+2-feather for H. Let $t \in K$. Since dimension never increases on evidence, there are two cases to consider.

(α) $\forall k\ Dim_H(K_{t|k}) = n+1$ and $Dim_{\overline{H}}(K_{t|k}) \geq n+1$, or
(β) $\exists k\ Dim_H(K_{t|k}) \leq n$ or $Dim_{\overline{H}}(K_{t|k}) \leq n$.

Lemma 1: if (α) then $t \notin H$ and ϕ^0_{n+1} converges correctly to 0 on t with no mind-changes.

For suppose that $t \in H$. Then since $\forall k$, $Dim_{\overline{H}}(K_{t|k'}) \geq n+1$, we have that $\forall k, \exists t' \exists k' > k$ s.t. $K_{t|k'}$ is an n + 1 feather for $\overline{H}$ with shaft t' and $t'|k' = t|k'$. Hence, K is an n + 2 feather for H with shaft t, contrary to assumption. Finally, observe that $\forall k$, the second clause of ϕ^0_{n+1} is satisfied on t|k, so $\forall k, \phi^0_{n+1}(t|k) = 0$.

Lemma 2: if (β) then ϕ^0_{n+1} converges to the truth in n mind-changes.

Let m be the least k such that DimH(K$_t$|k) £ n or Dim$\overline{H}$(Kt|k) ^ n. Suppose that Dim$\overline{H}$(Kt|$_m$) ^ n. Then $0\overset{0}{n}$ «. ifflm^1) = 0, for all m^1 < m and V m^1 > m, $\phi^0_{n+1}(t|m') = \psi^1_{H, K_{trunc(H,K,n,\sigma)}}(t|m')$. Since $\psi^1_{H, K_{trunc(H,K,n,\sigma)}}$ decides H over K$_t$|$_m$ in n-1 mind-changes starting with 1, ϕ^0_{n+1} succeeds in n mind-changes starting with 0. In case Dim^(Kt|m) < n, we have a similar situation, except that VHK<w^^K-n-«» starts with 0, so ϕ^0_n • 1 succeeds in n -1 mind-changes.

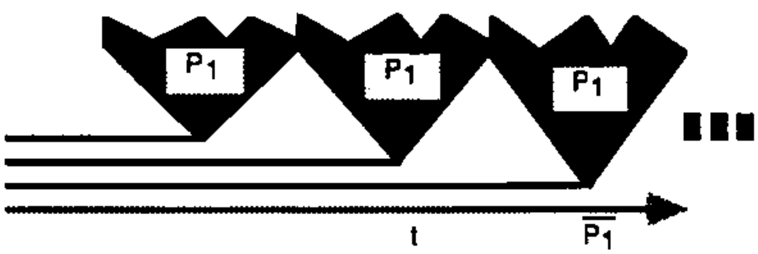
The induction for (b) is similar, except that the method employed is:

$$\phi^1_{n+1}(\sigma) = \begin{cases} 1 \text{ if } \sigma \text{ is empty} \\ 1 \text{ if } Dim_H(K_\sigma) > n \text{ and } Dim_{\overline{H}}(K_\sigma) > n \\ \psi^1_{H, K_{trunc(H,K,n,\sigma)}}(\sigma) \text{ if } Dim_{\overline{H}}(K_\sigma) \leq n \\ \psi^0_{H, K_{trunc(H,K,n,\sigma)}}(\sigma) \text{ otherwise (i.e. if } Dim_H(K_\sigma) \leq n) \end{cases}$$

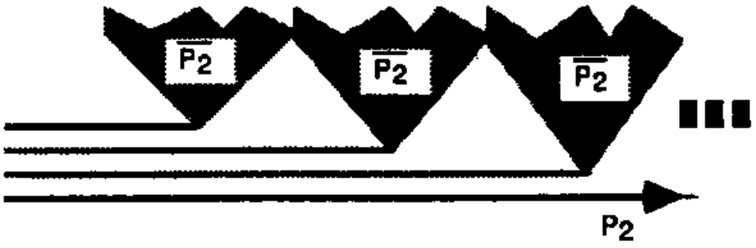
(c) and (d) may be obtained just as in Theorem 5.1. •-

By theorem 6.1, feather dimension and MC-compiexity coincide exactly. It is interesting to see how the correspondence works by constructing feathers outof intersections and unions of open and closed sets. To start, choose some data presentation t, and let

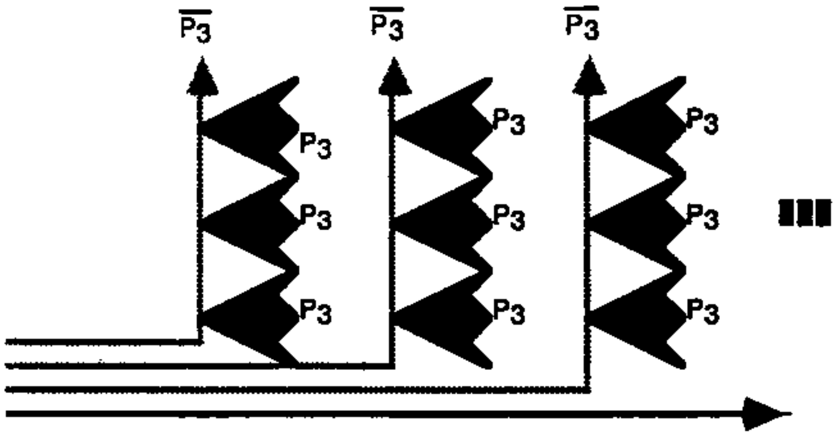
$$P_1 = \bigcup_{t \text{ does not extend } \sigma} K_\sigma$$



K is clearly a 2-feather for $\overline{Pi}$ and Pi € I_1^{MC} since Pi is open. Next, let P2 = $\overline{Pi}$. Now we have a 2-featherfor P2, and P2 e Π_1^{MC}.

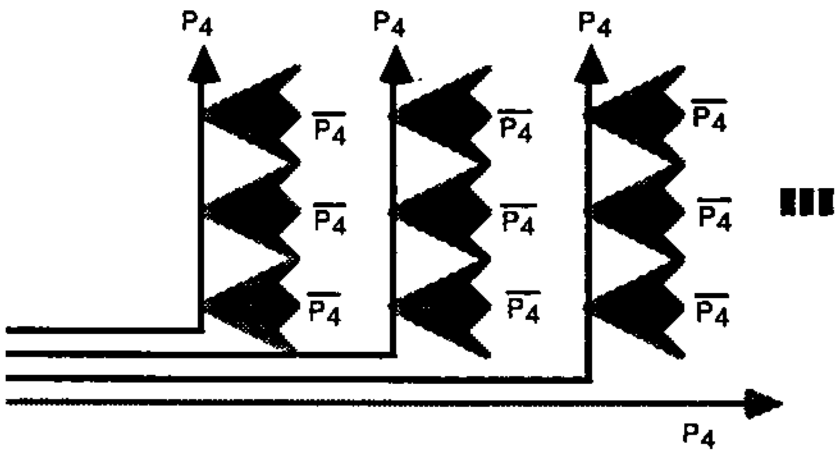


We know from theorem 6.1 that to build a **more complicated feather, we add some open set to P2** to **form** P3.



P3

P3 - P2 is clearly open, since it is depicted as a union of fans. So P3 is of the form **O u C ,** and **is** therefore in Π_2^{MC}. Let P4 be the complement of P3.



Now we are again free to add a dimension to P4 by augmenting it with an open set. By successive complementations and open set additions, we can build feathers of arbitrary, finite dimension.

7. Paradigms of Discovery

In problems of hypothesis assessment, the scientist is assigned some hypothesis whose adequacy is to be investigated on the basis of increasing data. In discovery problems, the scientist is required to invent an adequate hypothesis on the basis of increasing data. Most results in learning theory concern discovery rather than assessment. Interest has centered on grammatical inference, recursive function identification, and the induction of first-order theories from presentations of structure diagrams. Each of these applications is a special case of the following setting.

Hypothesis assessment methods do not have to read or to produce hypotheses, so hypotheses could be viewed abstractly as uncountable sets of infinite sequences. This will not do when discovery procedures are computers. Instead, we will assume simply that hypotheses are stated in a discrete, finitary language with a decidable syntax. Hence, hypotheses, like data sentences, may encoded by natural numbers. As before, we will assume that the goal of inquiry is some relation of ***adequacy*** $A \subseteq \omega^{\omega} \times \omega$ holding between infinite data presentations and hypotheses. A may entail consistency with the total data, explanatory completeness over the total data, simplicity, unity, or any other desideratum that depends only upon the hypothesis and the total data. We will let $A_i = \{t: A(t, i)\}$.

A discovery method will be a map from finite segments of data presentations to hypotheses, i.e. $\phi: \omega^* \rightarrow \omega$. We will consider the following concepts of successful discovery:

ϕ ***identifies*** A-***adequate hypotheses over*** K ***with certainty*** $\Leftrightarrow$
$\forall t \in K\ \exists n$ s.t. $A(t, \phi(t|n))$ and $\forall m < n, \phi(t|n) = \#$.

ϕ ***identifies*** A-***adequate hypotheses in*** n ***mind-changes*** $\Leftrightarrow$
$\forall t \in K\ \ \exists n \in \omega\ \forall m > n, \phi(t|m) = \chi_H(t)$ and
$mc(\phi, t) \leq n$.

ϕ ***identifies*** A-***adequate hypotheses over*** K ***in the limit*** $\Leftrightarrow$
$\forall t \in K\ \exists n\ \forall m \geq n\ \phi(t|m) = \phi(t|n)\ \&\ A(t, \phi(t|n))$.

A-***adequate hypotheses are*** [***effectively***] ***identifiable over*** K
[***with certainty*** / ***with n mind changes*** / ***in the limit***] $\Leftrightarrow$

$\exists$[total recursive] ϕ s.t. ϕ identifies A-adequate hypotheses over K
$\begin{bmatrix} \text{with certainty} \\ \text{with n mind changes} \\ \text{in the limit} \end{bmatrix}$

8. Characterizations of Reliable Discovery

Each of these senses of success requires that an adequate hypotheses be found for each data presentation in K. It is therefore trivial that A must cover K in the following sense if success is to be possible.

A ***covers*** K $\Leftrightarrow \forall t \in K \, \exists i \in \omega$ s.t. A(t, i).

Now we may characterize identification with certainty and identification in the limit.

Theorem 8.1:

(a) A-adequate hypotheses are [effectively] identifiable over K with certainty

$\Leftrightarrow \exists A' \subseteq A$ s.t. A' covers K and $A' \in \Sigma_1^{B,K}\left[\Sigma_1^{0,K}\right]$
$\Leftrightarrow \exists A' \subseteq A$ s.t. A' covers K and $A' \in \Delta_1^{B,K}\left[\Delta_1^{0,K}\right]$.

(b) A-adequate hypotheses are [effectively] identifiable over K in the limit

$\Leftrightarrow \exists A' \subseteq A$ s.t. A' covers K and $A' \in \Sigma_2^{B,K}\left[\Sigma_2^{0,K}\right]$
$\Leftrightarrow \exists A' \subseteq A$ s.t. A' covers K and $A' \in \Delta_2^{B,K}\left[\Delta_2^{0,K}\right]$.

Proof: (a) $\Rightarrow$ Let effective discovery method ϕ identify A-adequate hypotheses over K. Define

$A'(t, i) \Leftrightarrow \exists n$ s.t. $\phi(t|n) = i$ and $\forall m < n, \phi(t|n) = \#$.

A' covers K and $A' \subseteq A$ since ϕ identifies A-adequate hypotheses over K with certainty. By definition, $A' \in \Sigma_1^{B,K}\left[\Sigma_1^{0,K}\right]$. But we also have

$\forall t \in K, \neg A'(t, i) \Leftrightarrow \exists k \neq i$ s.t. $\phi(t|n) = k$ and $\forall m < n, \phi(t|n) = \#$

so $A' \in \Delta_1^{B,K}\left[\Delta_1^{0,K}\right]$.

⇐ Suppose that $A' \subseteq A$ covers K and $A' \in \Sigma_1^{B,K}\left[\Sigma_1^{0,K}\right]$. Let VERIFY[t, i] be an [effective] positive test for A'. Now we can define the following [effective] discovery method, which identifies A-adequate hypotheses over K with certainty.

> DISCOVER(σ):
> set n := length(σ)
> if $\forall i \leq n$, VERIFY(σ_1, i), VERIFY(σ_2, i), ..., VERIFY(σ_n, i) all fail to produce 1 in n computational steps, return #
> else, return the first i such that the first non-# output of TEST on σ is 1.

(b) ⇒ Let effective discovery method ϕ identify A-adequate hypotheses over K in the limit. Define

$$A'(t, i) \Leftrightarrow \exists n\, \forall m \geq n\; \phi(t|m) = i.$$

A' covers K and $A' \subseteq A$ since ϕ identifies A-adequate hypotheses over K. By definition, $A' \in \Sigma_2^{B,K}\left[\Sigma_2^{0,K}\right]$. Since ϕ converges to some i on each $t \in K$, we also have that

$$\forall t \in K,\ \neg A'(t, i) \Leftrightarrow \exists n\, \forall m \geq n\; \phi(t|m) \neq i.$$

so $A' \in \Delta_2^{B,K}\left[\Delta_2^{0,K}\right]$.

⇒ (effective case) Suppose that $A' \subseteq A$ covers K and $A' \in \Sigma_2^{0,K}$. Then there is a recursive relation S such that for each $i \in \omega$, $t \in K$, $A'(t, i) \Leftrightarrow \exists n\, \forall m\, S(t, i, n, m)$. Let P_k be a decision procedure for S. Let ***code***: $\omega^2 \to \omega$ be a fixed, recursive bijection from pairs of numbers to numbers, and let $(.)_1$, $(.)_2$ be recursive functions such that $((x)_1, (x)_2)$ is the inverse of code. Define

LIMDISCOVER(σ) =

> ($\mu x \leq$ length(σ): $\forall m \leq$ length(σ), $P_k(\sigma, (x)_1, (x)_2, m)$) does not return 0 in length(σ) steps)$_1$
> if there is such an x.
> # otherwise.

LIMDISCOVER is computable because the minimization and quantifiers are all bounded. Let $t \in K$. Then since A' covers K, there is some $i \in \omega$ s.t. A'(t, i). Thus, $\exists n\, \forall m\, S(t, i, n, m)$. Pick $x = \mu y$: $\forall m$, $S(t, (y)_1, (y)_2, m)$. For each $y < x$, $\neg S(t, i, n, m)$ so there is some k_y, and some $w \leq k$, such that

—iS(t|k_y, i, n, w). Let k* = sup{k_y: y < x}. Thus, Vk £ k*, the value of the minimization expression in the definition of LIMDISCOVER on t|k is at least x. Since 3n Vm S(t_f $(x)_2$, n, m), the value of the minimization never exceeds x after $k^{\#}$, so Vk ≥ k*. LIMDISCOVER(t|k) - $(x)_2$. Since 3n Vm S(t, $(x)_2$, n, m), we also have that A'(t, (x)2). and hence that A(t, (x)2), as required. The ineffective case is similar, with references to computation omitted. ▪

9. **Learning Theory** Results as **Relative Complexity Classifications**

The following examples illustrate how the standard paradigms of language leamability and function identification drop out as special cases of the approach adopted here. From our perspective, standard results in learning theory may be thought of as strong relative complexity classifications for relations of type co" X co.

Function Identification:

The problem of identifying set Rec of total recursive functions:

Adequacy relation:	Af_{un}(t, i) » <+j = t
Background knowledge:	K c Rec

One of the first negative results about function identification is that the collection of all recursive functions is identifiable in the limit, but not effectively so. The positive result follows from the fact that Af_{un}(t, i) <=> ft = t <=> Vn **0j(n)** = t_n. Since the relation ^(n) = t_n is $^A i^{B, Rec}$, Af_{un} € $^n i^{B, Rec}$. The situation is different in the computable case: Af_{un}(t, i) «=> <t-i = t <=> Vn <f>j(n) = t_n » Vn 3k <t>j(n)it_n. Gold's negative result together with Theorem 10 tells us that this characterization is optimal, i.e. that Afun ^ $^n2^{0, Rec}$ - $fz^{0, Rec}$. Indeed, Gold's result tells us in light of theorem 10 that there is no A c Af_Un covering Rec such that **A G ^{z} $2^{0, Rec}$** -

Language Identification by RE index:

The problem of identifying language class L e RE:

Adequacy relation:	ARE(t, i) » Wj = rng(t)
Background knowledge:	K_L = {t: 3S e L s.t. rng(t) = S}.

Here, the basic theorem is that no collection of languages L' containing all finite languages and one infinite language is identifiable in the limit, even by an ineffective learner [11]. In our generaOzed notation, this is the claim that ARE hypotheses are not identifiable over Kf, which together with Theorem 10 implies that there is no A^1 **c** ARE covering $K\backslash j$ such that A^1 € *I?* B.Kf A general upper bound meeting Golds lower bound is easy to calculate.

$W_i = rng(t) \Leftrightarrow$

$\forall n[\phi_i(n)\downarrow \Leftrightarrow$ 3k s.t. n = t|<] $\Leftrightarrow$

$\forall n\ \forall k\ \exists k'\ [...] \in \Pi_2^{0,\ K_{RE}} \subseteq \Pi_2^{0,\ K_L}$.

Hence, AREO, i) G $n_2^{0,\ K_{RE}}$ -/? B. KRE.

Another standard example is the collection Lfj_n of all finite languages.

$\forall t \in K_{Lfin},\ W_i = rng(t) \Leftrightarrow$

3k Vk' ≥ k Vn fo(n)!I« n e mg(t|k')J e $zj^{Ku..}$.

Language Identification by recursive **index:**

The problem of identifying language class L Q RE:

Adequacy relation:	A_R(t, i) <* ft « Xrng(t)
Background knowledge:	K_L = {t: 3S G L s.t. rng(t) = S}.

Let L* be as in the last example. Gold showed that L' is not identifiable by an effective learner even when the data presentations are assumed to be primitive recursive. Let Prim be the set of all primitive recursive sequences. Let K^1 = K(j n Prim. Then Gold's result shows that there is no A c AR covering KL* n Prim such that A e $^{0,\ KL^*\ n\ Prim}$. Once again it is easy to compute an upper bound that matches Gold's lower bound:

ft - Xrng(t) »

Vn[[(3m n = t_m)=>3kft(n)Ii] and $[\forall m'\ n \neq t_{m'} \Rightarrow (\exists k'\ \phi i(n)\downarrow^{k'} 0)] \Leftrightarrow$

Vn Vm 3m' 3k 3k' [...] e $n_2^{0,\ KRE}$ c $n_2^{0.\ KL\text{-}\ ^\wedge\ Prim}$.

9. Characterizations of Reliable Discovery with **Bounded Mind-Changes**

Given the results so far, one might guess that A-adequate hypotheses are identifiable over K in n mind-changes just in case $\exists A' \subseteq A$ s.t. A' covers K and $A' \in \Pi_n^{MC, K} \cup \Sigma_n^{MC, K}$. But in fact:

Proposition 11: $\exists A$, K s.t. $A \in \Sigma_1^{mc, K}$ but $\forall n$, A-adequate hypotheses are not identifiable over K in n mind-changes even by a non-computable method.

Proof: Define A(t, i) $\Leftrightarrow$ 0 occurs in t at least i times, and let $K = \omega^{\omega}$. Evidently, $A \in \Sigma_1^0 = \Sigma_1^{mc}$. But a simple diagonal argument permits us to fool an arbitrary discovery method an arbitrary number of times. ■

The problem is that discovery depends not only on the topology of each hypothesis, but also on how the data presentations for different hypotheses are interleaved together. This interleaved structure of the adequacy relation can be captured exactly if we generalize the notion of n-feathers slightly.

K ***is a*** 1***-feather for*** i ***mod*** A ***with shaft*** t $\Leftrightarrow$ $t \in K \cap A_i$.

K ***is an*** n+1***-feather for*** i ***mod*** A ***with shaft*** t $\Leftrightarrow$
$t \in K \cap A_i$ and
$\forall n \exists t' \in K \exists k \in \omega$ s.t.
$t|n = t'|n$ and
K is an n-feather with shaft t for K mod A with shaft t', and $t' \notin A_k$.

K ***is an*** n***-feather for*** i ***mod*** A $\Leftrightarrow$ $\exists t$ s.t. K is an n-feather for i mod A with shaft t.

K ***is an exact*** n***-feather for*** i ***mod*** A $\Leftrightarrow$
K is an n-feather for i mod A and $\forall m > n$, K is not an m-feather for i mod A.

Theorem 9.1:
A-adequate hypotheses are identifiable over K in n mind-changes starting with # $\Leftrightarrow$
$\forall i$, K is not an n-feather for i.

Proof: Analogous to the proof of Theorem 6.1. ■

Example: Recall the case of learning finite languages by RE index. It is easy to see that $K_{L_{fin}}$ is an n-feather for A_{RE} for each n, so the finite languages are not identifiable under any bounded number of mind changes.

10. Conclusion

Complete characterizations have been presented for effective and ineffective hypothesis assessment, in the short run, in the long run, and with bounded mind changes. Complete characterizations have also been presented for effective and ineffective discovery in the limit, and for non-effective discovery with bounded mind-changes. It remains to characterize effective discovery with bounded mind changes.

Acknowledgments

This research was supported, in part, by a generous grant from the Buhl foundation. I would like to thank Clark Glymour, Dan Osherson, Scott Weinstein, Cory Juhl, Peter Spirtes and Teddy Seidenfeld for useful comments and discussions. The anonymous referee report was especially useful.

References

[1] Gold, E. M., "Language Identification in the Limit", *Information and Control* 10 (1967) 447-474.

[2] Gold, E. M., "Limiting Recursion", *Journal of Symbolic Logic,* 30:1 (1965). pp. 27-48.

[3] Hinman, P. G., *Recursion-Theoretic Hierarchies,* New York: Springer (1978).

[4] Kugel, P., "Induction, Pure and Simple", *Information and Control,* 33 (1977), pp. 276-336.

[5] Kuratowski, K., *Topology,* Vol. 1, New York: Academic Press (1966).

[6] Levy, A., *Basic Set Theory,* New York: Springer (1979).

[7] Moschovakis, Y., N., Descriptive Set Theory, New York: North Holland (1980).

[8] Osherson, D., and S. Weinstein , "A Universal Inductive Inference Machine", Unpublished manuscript (1989).

[9] Osherson, D., and S. Weinstein, "Identification in the Limit of First Order Structures", *Journal of Philosophical Logic,* 15, (1986), pp. 55-81 .

[10] Osherson, 0., and S. Weinstein, "Paradigms of Truth Detection", *Journal of Philosophical Logic,* 18 (1989), pp. 1-42 .

[11] Osherson, 0., M. Stob, and S. Weinstein, *Systems that Learn,* Cambridge: MIT Press (1986).

[12] Putnam, H. Trial and Error Predicates and a Solution to a Problem of Mostowski". *Journal of Symbolic Logic,* 30:1 (1965), pp. 49-57.