Chapter 6 Hypothesis learning theory

1. Introduction

Hypothesis learning theory was studied first, as a language learning theory, by E. Mark Gold (1967) and subsequently, in more mathematical detail by Jerome Feldman (1970). The importance of this theory is that it deals with the progression of a hypothesis discovery method through time. While it appears on the surface to deal exclusively with language guessing, the theory may be adapted for other purposes. Our presentation tries to bring this out by developing part of it on an abstract level. One can continue the abstract theory to cover the rest of Feldman's theory, but enough will be developed to give a good ground for application, criticism and indication of possible future trends.

All this work is essentially an elaboration of a simple but surprising theorem of Gold.

A complete information sequence for a context-free grammar, G, is an infinite list, I of the form $\pm y_1$ $\pm y_2$ .... where:

(1) +y appears in the list iff the string y is in the language determined by G.

(2) -y appears in the list iff the string y is not in the language determined by G.

Suppose that at time t one is presented with the signed string $\pm y_t$. To what extent can one succeed in "eventually" guessing G? It turns out that there is a method, defined independently of G and I, which will identify G in the limit. That is, there is a time τ at which it will guess and ever thereafter continue to guess, a grammar G' whose language is the same as that of G.

Let G_1,.... be an enumeration of all context-free grammars. The method chooses, at time t, the first grammar, G' say, such that if +y has appeared in the list by time t then y is in the language of G', and if -y has appeared in the list, by time t, then y is not in the language of G'. Evidently one can always find such a G', if only because G appears in the list. If G" is a grammar in the list whose language differs from that of G, then either there is a y in G but not in G", or vice versa. At some time, t', say ±y (whichever is appropriate) will appear in the list and thereafter G" will not be chosen by our method. Hence after some time, τ say, every G occurring before the first grammar in the list whose language is the same as that of G, G' say, will never be chosen. At this time G' will be chosen and will continue to be chosen.

This theorem shows that, at least in this case, it is eventually possible to "learn" the truth. However it is never possible to "know" the truth, since any guess G can be forced to change by some ±y or other.

2. Abstract theory

The theory is presented as an informal mathematical theory.

The domains are the hypothesis space, Hyp and the phenomenon space, Phen.

Axiom 1 Hyp is a recursive, infinite set of integers.

The variables ranging over Hyp are h, h' etc.

Examples 1) Hyp is the set of Gödel numbers of first-order lawlike universal sentences. Gödel numbers give a way of coding sentences as integers. This possibility of coding allows us to consider the theory a general one.

2) Hyp is the set of Gödel numbers of context-free grammars.

Hyp is required to be infinite in order to avoid trivial exceptional cases in later theorems. We will asume some fixed enumeration, $h_1, h_2, \ldots$ of Hyp.

Axiom 2 Phen is an infinite recursive set.

The variables ranging over Phen. are f, f_1 etc. We use, F, F^+, F^- etc. to range over finite subsets of Phen. $\mathcal{F}$(Phen) is the set of finite subsets of Phen. The fact that Phen is infinite forces attention on the harder problems. Generally, our theory becomes trivial when Phen is finite.

We will let $\mathcal{S} \subseteq$ Phen be a variable ranging over the recursive subsets of Phen. $\mathcal{S}$ is to be understood as a subject domain separated out from Phen, the class of all possible phenomena.

<u>Examples</u> 1) (Gödel numbers of) strings of letters.

2) (The Gödel numbers of) the set of all ground clauses, C which do not follow from Th, an arbitrary consistent set of sentences, but are consistent with it and Irr, another set of sentences consistent with Th.

The predicates are: Accountsfor is of sort Hyp × Phen,

M_A is of sort Hyp × Phen × Integers,

Consistent is of sort Hyp × $\mathcal{F}$(Phen) × $\mathcal{F}$(Phen)

M_C is of sort Hyp × $\mathcal{F}$(Phen) × $\mathcal{F}$(Phen) × Integers.

<u>Axiom 3</u> Accountsfor is partial recursive, M_A is primitive recursive and Accountsfor(h,f) $\equiv \exists m\ M_A(h,f,m)$.

We extend Accountsfor to a partial recursive predicate of type Hyp × $\mathcal{F}$(Phen) by:

$$\text{Accountsfor}(h,F) \equiv_{\text{def}} \forall f \in F\ \text{Accountsfor}(h,f).$$

Similarly M_A is extended to a primitive recursive predicate of type Hyp × $\mathcal{F}$(Phen) × Integers by:

$$M_A(h,F,m) \equiv_{\text{def}} \forall f \in F\ \exists m' \leq m\ M_A(h,f,m').$$

Evidently, Accountsfor(h,F) $\equiv \exists m\ M_A(h,F,m)$.

Accountsfor is the type of implication being used to explain the phenomena.

Examples 1) $\mathrm{Accountsfor}(h,C) \equiv \vdash_{Th} h \rightarrow C$.

2) $\mathrm{Accountsfor}(h,C) \equiv h \leq \{C\}\ (Th)$.

3) $\mathrm{Accountsfor}(h,f) \equiv$ f is a string in the language given by the grammar h.

M_A corresponds to a program for Accountsfor.

Axiom 4 Consistent is partial recursive, M_C is primitive recursive and $\neg\ \mathrm{Consistent}(h,F^+,F^-) \equiv \exists m\ M_C(h,F^+,F^-,m)$.

$\mathrm{Consistent}(h,F^+,F^-)$ means that h is consistent with the occurrence of the phenomena in F^+ (hence the plus sign) and with the non-occurrence of those in F^-.

If Accountsfor were a logical implication such that a completeness theorem held and one had sufficient logical symbolism, one could define Consistent in terms of Accountsfor by:

$$\mathrm{Consistent}(h,F^+,F^-) \equiv \exists f\ \neg\ \mathrm{Accountsfor}(h \wedge F^+ \wedge \neg F^-, f).$$

However we do not want such a strong implication in general, since it may be easier or more relevant to look for hypotheses bearing implications of a weaker sort, such as generalisation.

M_C corresponds to a program for $\neg$ Consistent.

Examples 1) When $F^+ = \{C_i \mid i=1,n\}$ and $F^- = \{D_j \mid j=1,m\}$, $\mathrm{Consistent}(h,F^+,F^-) \equiv \forall(h \wedge \bigwedge_{i=1}^{n} C_i \wedge \bigwedge_{j=1}^{m} D_j) \wedge Th \wedge Irr$ is consistent.

2) h is a grammar and F^+ and F^- are sets of strings. $\mathrm{Consistent}(h,F^+,F^-) \equiv (F^+ \subseteq L(h)) \wedge (L(h) \cap F^- = \emptyset)$.

The other axioms give some of the logical properties of the predicates. They are meant to be a small set which allows Feldman's theorems to be proved in a general form.

Axiom 5 $\text{Consistent}(h,F^+,F^-) \rightarrow \forall f^- \in F^- \neg \text{Accountsfor}(h,f^-)$.
Accountsfor is a consistent deduction principle.

Axiom 6 $F_1^+ \supseteq F_2^+ \wedge F_1^- \supseteq F_2^- \wedge \text{Consistent}(h,F_1^+,F_1^-) \rightarrow \text{Consistent}(h,F_2^+,F_2^-)$.
Consistency is hereditary downwards (preserved by any operation which produces subsets).

Axiom 7 $\text{Consistent}(h,F^+,F^-) \rightarrow F^+ \cap F^- = \emptyset$.
The phenomena in F^- are effectively negated.

The next axioms use a special hypothesis, T, which functions as a tautology.

Axiom 8 $\forall f\ \text{Consistent}(T,f,\emptyset) \wedge \text{Consistent}(T,\emptyset,f)$.
No phenomenon is necessary but every one is possible.

Axiom 9 $\text{Consistent}(h,F^+,F^-) \rightarrow \text{Consistent}(h,F^+ \cup \{f\},F^-) \vee \text{Consistent}(h,F^+,F^- \cup \{f\})$.
This is a partial version of the law of the excluded middle.

Axiom 10 It is decidable whether or not $\text{Consistent}(T,F^+,F^-)$.

Axiom 11 $\text{Consistent}(T,F^+,F^-) \rightarrow \exists h(\text{Consistent}(h,F^+,F^-) \wedge \text{Accountsfor}(h,F^+$

Every consistent finite set has a consistent explanation.

Information sequences

Information sequences are sequences of observations of the occurrence or non-occurrence of certain phenomena.

Formally, an information sequence is an infinite sequence in $(\{0,1\} \times \text{Phen})^{\infty}$ or else a finite one in $(\{0,1\} \times \text{Phen})^*$. We will display $<1,f>$ as $+f$ and $<0,f>$ as $-f$.

We will use the variables $I, I_1, \ldots\ldots$ to range over information sequences. $|I|$ is the length of I. Note that $0 \leq |I| \leq \infty$. If $0 < t \leq |I|$, $I(t)$ is the tth component of I. If I_1 is finite, $I_1 + I_2$ is that information sequence consisting of I_1 followed by I_2. If I_1 is finite then $n\ I_1 = \underbrace{I_1 + \ldots\ldots + I_1}_{n \text{ times}}$, where $n \geq 0$. More formally,

$0\ I_1$ = the empty information sequence,

$(n+1)I_1 = nI_1 + I_1$.

If $I_1 = I_2 + I_3$, where $|I_3| > 0$ then I_1 extends I_2. This is written as $I_1 > I_2$.

If $t \leq |I|$, then $I^t = <I(1),\ldots\ldots,I(t)>$. If $t > |I|$, then $I^t = I$. The positive information in I is defined to be $S^+(I) = \{f \mid +f$ occurs in $I\}$. Similarly, we define $S^-(I) = \{f \mid -f$ occurs in $I\}$. I_1 agrees with I_2 iff $S^+(I_1) = S^+(I_2)$ and $S^-(I_1) = S^-(I_2)$.

I is complete iff $S^+(I) \cup S^-(I) = \text{Phen}$. If I is complete it is infinite. I is consistent iff for all t, Consistent$(T, S^+(I^t), S^-(I^t)$.

A hypothesis *explains* an information sequence I, for the subject $\mathscr{S}$ iff

$$\forall t \ \mathrm{Accountsfor}(h, S^{+}(I^{t}) \cap \mathscr{S}) \wedge \ \mathrm{Consistent}(h, S^{+}(I^{t}), S^{-}(I^{t})).$$

When $\mathscr{S}$ = Phen we suppress, both here and elsewhere, any reference to it.

Induction machines

An induction machine, $\mathfrak{M}$, is a recursive function from the set of finite information sequences to Hyp.

$\mathfrak{M}$ *identifies* h in the limit on I iff $\exists \tau \forall t \geq \tau \ \mathfrak{M}(I^{t}) = h$.

$\mathfrak{M}$ *matches* an explanation of I for the subject $\mathscr{S}$ in the limit iff $\exists \tau \forall t \geq \tau \ \mathfrak{M}(I^{t})$ explains I for $\mathscr{S}$.

$\mathfrak{M}$ *approaches* an explanation of I for the subject $\mathscr{S}$ iff

1) $\forall f \in S^{+}(I) \cap \mathscr{S} \ \exists \tau \forall t \geq \tau$ Accountsfor$(\mathfrak{M}(I^{t}), f)$.

2) $\forall h$ not explaining I for the subject $\mathscr{S}$, $\exists \tau \forall t \geq \tau \ \mathfrak{M}(I^{t}) \neq h$.

The approach is *strong* iff, in addition:

3) $\exists h$ explaining I for $\mathscr{S}$ such that $\exists \tau \ \forall t \geq \tau \ \exists$ finite I' extending I^{t} and agreeing with I^{t} such that $\mathfrak{M}(I') = h$.

This condition is slightly stronger than Feldman's for a strong approach.

3. Inferring hypotheses

Our first induction machine $\mathcal{M}_1$ is essentially due to Gold. It is defined under the assumption that Accountsfor is recursive, and relative to a subject, $\mathcal{S}$.

$$\mathcal{M}_1(I) = \begin{cases} h_1 \ (\text{if } \neg \ \text{Consistent}(T,S^+(I),S^-(I))) \\ \text{the first } h_i \text{ such that Accountsfor}(h_i,S^+(I) \wedge \mathcal{S}) \text{ and} \\ \quad \forall F^+ \subseteq S^+(I),F^- \subseteq S^-(I) \ \forall m \leq |I| \neg M_C(h_i,F^+,F^-,|I|) \\ \qquad\qquad\qquad\qquad\qquad\qquad (\text{Otherwise}). \end{cases}$$

That $\mathcal{M}_1$ is total recursive follows from the assumption that Accountsfor is recursive and axioms 10 and 11.

Theorem 1 If h explains a consistent, complete I for $\mathcal{S}$ then $\mathcal{M}_1$ identifies a hypothesis, h', in the limit on I which explains I for $\mathcal{S}$.

Proof First we show that if h_t, does not explain I for $\mathcal{S}$ then $\exists \tau \forall t \geq \tau \mathcal{M}_1(I^t) \neq h_{t'}$.

There are two possibilities. Either $\neg$ Accountsfor$(h_{t'},f_{t_1})$ for some $f_{t_1} \in S^+(I) \wedge \mathcal{S}$ or else $\neg$ Consistent$(h_{t'},S^+(I^{t_1}),S^-(I^{t_1}))$ for some t_1. In the first case we can take $\tau = t_1$.

In the second, there is an m such that $M_C(h,S^+(I^{t_1}),S^-(I^{t_1}),m)$. As Phen is infinite and I is complete, there is a τ such that $|I^{\tau}| \geq m$. This τ has the necessary properties in this case.

Suppose that h' is the first hypothesis which explains I for $\mathcal{S}$. Then there is a time τ such that if $t \geq \tau$, $\mathcal{M}_1(I^t)$ is not a hypothesis occurring before h'. Now, $\forall t$ Accountsfor$(h', S^+(I^t) \cap \mathcal{S})$ and since $\forall t$ Consistent$(h', S^+(I^t), S^-(I^t))$, $\forall m\ M_C(h, S^+(I^t), S^-(I^t), m)$ by axioms 4 and 6. Hence $\forall t \geq \tau\ \mathcal{M}_1(I^t) = h'$. This concludes the proof.

If in addition, Consistent is recursive, then we assert, leaving the proof to the reader, that one can choose a simpler machine, $\mathcal{M}_2$.

$$\mathcal{M}_2(I) = \begin{cases} h_1 \text{ (if } \neg \text{ Consistent}(T, S^+(I), S^-(I))) \\ \\ \text{the first } h_i \text{ such that Accountsfor}(h_i, S^+(I) \cap \mathcal{S}) \\ \text{and Consistent}(h_i, S^+(I), S^-(I)). \\ \\ \qquad\qquad\qquad\qquad \text{(Otherwise).} \end{cases}$$

We cannot extend the result to the case where Accountsfor is not recursive. In fact under the assumption:

$$\forall F^+, F^- \text{ Consistent}(T, F^+, F^-) \text{ -> } [\ \exists f \text{ Consistent}(T, F^+ \cup \{f\}, F^-) \wedge \text{Consistent}(T, F^+, F^- \cup \{f\})],$$

we can show that no machine can identify an explanatory hypothesis in the limit when $\mathcal{S}$ = Phen. The assumptions hold when Hyp is the set of general rewriting systems (Feldman, 1970) and Phen the corresponding set of strings.

<u>Theorem 2</u> Suppose that the assumptions hold. For every machine, $\mathcal{M}$, there is a consistent, complete and recursive information sequence on

which $\mathcal{M}$ does not identify an explanatory hypothesis in the limit.

Proof Let $f_1, \ldots$ be an enumeration of Phen. We will find an I satisfying the conditions of the theorem such that $I(t) = \pm f_t$. Let us say that a sequence, I' is suitable iff when $I'(t)$ is defined, $I'(t) = + f_t$, or else $I'(t) = -f_t$.

Suppose first that:

$$\exists \text{finite suitable } I'(\text{Consistent}(T,S^+(I'),S^-(I')) \wedge$$
$$\forall \text{finite suitable } I'' > I'[\text{ Consistent}(T,S^+(I''),S^-(I'')) \rightarrow (\mathcal{M}(I') = \mathcal{M}(I''))].$$

In this case we choose I' as guaranteed by the supposition. Let f_t be the first phenomenon such that $\text{Consistent}(T,S^+(I') \cup \{f_t\},S^-(I'))$ and $\text{Consistent}(T,S^+(I'), S^-(I') \cup \{f_t\})$. The existence of f_t is guaranteed by the assumption and axiom 7. Let I" be a finite suitable extension of I' such that $I''(t) = + f_t \equiv \neg \text{Accountsfor}(\mathcal{M}(I'),f_t)$. The existence of I" is guaranteed by axiom 9. As I" is finite, it is recursive. Axioms 9 and 10 guarantee that I" has an extension I which is a complete, consistent and suitable information sequence. Now $\mathcal{M}$ identifies $\mathcal{M}(I')$ in the limit, by the supposition. If $\neg \text{Accountsfor}(\mathcal{M}(I'),f_t)$ then $I(t) = + f_t$. If $\text{Accountsfor}(\mathcal{M}(I'),f_t)$ then $I(t) = -f_t$, and so, by axioms 5 and 6, $\neg \text{Consistent}(\mathcal{M}(I'),S^+(I'^t),S^-(I'^t))$.

Therefore under the supposition, $\mathcal{M}$ does not identify, in the limit, a hypothesis explaining I.

Let us assume, to the contrary that:

$$\forall \text{ finite suitable } I'(\text{Consistent}(T,S^{+}(I'),S^{-}(I'))$$
$$\Rightarrow \exists \text{ finite suitable } I'' > I'(\text{Consistent}(T,S^{+}(I''),S^{-}(I''))$$
$$\wedge\ M(I') \neq M(I'')).$$

In this case, I" may be obtained recursively from I' since $\mathcal{M}$ is recursive and by axiom 10, given I" one can tell by a recursive procedure whether or not $\text{Consistent}(T,S^{+}(I''),S^{-}(I''))$. Therefore there is a recursive function g such that $\forall I'(\text{Consistent}(T,S^{+}(I'),S^{-}(I')) \Rightarrow (g(I')$ is a finite suitable sequence extending I' such that $\text{Consistent}(T,S^{+}(I''),S^{-}(I''))\wedge \mathcal{M}(I') \neq \mathcal{M}(I''))$.

Let $I_0 = \langle + f_1\rangle$. By axiom 8, $\text{Consistent}(T,f_1,\emptyset)$. Then $I_0 < g(I_0) < \ldots < g^{t}(I_0) < \ldots$ and so there is a unique, recursive $I > g^{t}(I_0)$ (for all $t \geq 0$) such that for every t' there is a t such that $I^{t'} < g^{t}(I_0)$. Therefore by the properties of g and axiom 6, $\text{Consistent}(h,S_{t}^{+}(I),S_{t}^{-}(I))$. Therefore I is consistent and is certainly complete. Now $\mathcal{M}(g^{t}(I_0)) \neq \mathcal{M}(g^{t+1}(I_0))$ $(t \geq 0)$. Therefore $\mathcal{M}$ cannot identify any hypothesis in the limit.

This eatablishes the theorem.

If Hyp includes every recursive predicate of Phen, then although $\mathcal{M}$ will not identify a hypothesis explaining I in the limit, there is one.

Notice also that the I described in the theorem, although recursive, is not obtained recursively from $\mathcal{M}$. We conjecture that there is no such recursive map when Hyp includes every recursive predicate of Phen. This is not the case if we are guaranteed that for every finite I',

λf Accountsfor($\mathcal{M}(I'),f$) is always a recursive predicate on Phen. There is a unique recursive complete and consistent suitable information sequence recursively specified by:

1) $I(1) = +f_1$

2) $\neg$ Consistent$(T,S_t^+(I),S_t^-(I) \cup \{f_{t+1}\}) \rightarrow I(t+1) = +f_{t+1}$.

3) $\neg$ Consistent$(T,S_t^+(I) \cup \{f_{t+1}\},S_t^-(I)) \rightarrow I(t+1) = -f_{t+1}$.

4) Consistent$(T,S_t^+(I),S_t^-(I) \cup \{f_{t+1}\}) \wedge$ Consistent$(T,S_t^+(I) \cup \{f_{t+1}\},S_t(I)) \rightarrow (I(t+1) = +f_{t+1} \leftrightarrow \neg$ Accountsfor$(\mathcal{M}(I^t),f_{t+1})$

Of course we are still using the assumption behind theorem 2. (The definition of I is very close to the familiar proof that there can be no recursive enumeration of the recursive functions.) In this case we can actually find an $\mathcal{M}'$ which will do as well as $\mathcal{M}$ on any $I' \neq I$ and will identify a hypothesis, in the limit, which explains I (provided Hyp contains all the recursive functions). Putnam (1967) has made similar observations.

Although it is not possible to devise a machine which will identify an explanation in the limit, it is possible to strongly approach one, using a machine $\mathcal{M}_3$.

To calculate $\mathcal{M}_3(I)$ proceed as follows:

1) If $\neg$ Consistent$(T,S^+(I),S^-(I))$ then $\mathcal{M}_3(I) = h_1$.

2) Otherwise, find, by some fixed effective means, an h and an m such that $M_A(h,S^+(I) \wedge \mathcal{S},m)$ and $\forall m' \leq m + |I| \ \forall F^+ \subseteq S^+(I),F^- \subseteq S^-(I)$ $\neg M_C(h,F^+,F^-,m')$. Then $\mathcal{M}_3(I)$ is the first h_i such that

$M_A(h_i, S^+(I) \cap \mathcal{S}, m + |I|)$ and $\forall F^+ \subseteq S^+(I), F^- \subseteq S^-(I)\ \forall m' \leq m + |I|$ $\neg M_C(h, F^+, F^-, m')$.

Theorem 3 $\mathcal{M}_3$ strongly approaches an explanation in the limit on I for $\mathcal{S}$, if there is one.

Proof Suppose that there is an explanation of I for $\mathcal{S}$. Then I is consistent and so Accountsfor($\mathcal{M}_3(I^t), S^+(I^t) \cap \mathcal{S}$) and so condition 1 for approaching a strong explanation is verified. Suppose h_i does not explain I. Then either $\neg$ Accountsfor(h_i, f_{t_1}) for some f_{t_1} in $S^+(I)$ or else $\neg$ Consistent($h_i, S^+(I^{t_1}), S^-(I^{t_1})$) for some t_1. In the first case, if $\tau \geq t_1$, $\mathcal{M}_3(I^\tau) \neq h_1$. In the second case there is a t_2 such that $M_C(h_i, S^+(I^{t_1}), S^-(I^{t_1}))$. Therefore if $t \geq \max(t_1, t_2)$, $\mathcal{M}_3(I^t) \neq h_i$. This verifies condition 2.

Let h_i be the first hypothesis explaining I for $\mathcal{S}$. For some τ, $\mathcal{M}_3(I^\tau)$ is not any h occurring before h_i. Suppose $t \geq \tau$ and choose an n such that $M_A(h_i, S^+(I^t) \wedge \mathcal{S}, n)$. If $I' = I^t + nI(t)$ then $\mathcal{M}_3(I') = h_i$. This verifies condition 3 and concludes the proof.

These theorems have all been concerned with good behaviour in the limit. It is worth noting their local behaviour.

Suppose I is consistent. The Accountsfor($\mathcal{M}_i(I^t), S^+(I^t) \cap \mathcal{S}$) for i=1,3 and all t. Further Consistent($\mathcal{M}_1(I^t), S^+(I^t), S^-(I^t)$) for all t, although we only have, for i=2,3, $M_C(\mathcal{M}_i(I^t), S^+(I^t), S^-(I^t), m)$ where m depends on t and $m \to \infty$ as $t \to \infty$.

4. Inferring good hypotheses

By incorporating a complexity measure one can obtain better standards of good local behaviour.

We require that the standard ordering of the hypotheses, $h_1, h_2, \ldots$ is according to their simplicity. For example if the hypotheses are context-free grammars then, perhaps, if $j>i$, h_j has no less symbols than h_i. The number of symbols will also order sets of clauses in this linear way.

The derivational complexity $d(F^+,h)$ of F^+ from h is defined when Accountsfor(h,F^+) and then,

$d(F^+,h)$ = the smallest integer m such that $M_A(h,F^+,m)$.

In other words, using a standard notation, d is that partial function defined by $d(F^+,h) = \mu m\ M_A(h,F^+,m)$.

The complexity function γ is a partial recursive function from $\mathcal{F}(\text{Phen}) \times \text{Hyp}$ to $\mathbb{R}$, the set of the rationals. It combines the simplicity of a hypothesis with the derivational complexity.

There is a total recursive function $\gamma' : N^2 \rightarrow \mathbb{R}$ increasing unboundedly with each of its arguments such that:

$$\gamma(F^+,h_i) = \gamma'(i,d(F^+,h_i)).$$

The machines $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ could all be specified, using a recursive (under the appropriate conditions) predicate,

$P_j(1\leq j\leq 3)$ on Hyp × ({0,1} × Phen)*, by:

$\mathcal{M}_j(I)$ = the first h_i such that $P_j(h_i,I)$

For example:

$$\begin{aligned} P_4(h_i,I) \equiv \; & (\neg \; \text{Consistent}(T,S^+(I),S^-(I)) \to h_i = h_1) \\ & (\text{Consistent}(T,S^+(I),S^-(I)) \\ & \to (M_A(h_i,S^+(I) \wedge \mathcal{S}, m + |I|) \\ & \wedge \; \forall F^+ \subseteq S^+(I),\; F^- \subseteq S^-(I) \; \neg \; M_C(h_i,F^+,F^-,m + |I|))), \end{aligned}$$

where m is obtained recursively from I.

In each case for every finite information sequence I there is an h such that $P_j(h,I)$. Further, if I is consistent then $P_j(h,I)$ implies that Accountsfor$(h,S^+(I) \wedge \mathcal{S})$.

We will define corresponding machines $\mathcal{M}'_j(1\leq j\leq 3)$ with the properties:

1) $P_j(\mathcal{M}'_j(I),I)$
2) Suppose that I is consistent. If $P_j(h,I)$ then
   $\gamma(S^+(I) \wedge \mathcal{S},h) \geq \gamma(S^+(I) \wedge \mathcal{S}, \mathcal{M}'_j(I))$.

That is $\mathcal{M}'_j$ will choose a <u>best</u> machine rather than a first one. To compute $\mathcal{M}'_j(I)$ one proceeds as follows:

1) If I is not consistent, then $\mathcal{M}'_j(I) = \mathcal{M}_j(I)$.
2) Otherwise, compute $\mathcal{M}_j(I)$. Let k be the least integer such that $\gamma'(k,0) \geq \gamma(S^+(I) \wedge \mathcal{S}, \mathcal{M}_j(I))$.

$\mathcal{M}_j(I)$ is that first hypothesis minimising $\gamma(S^+(I)\cap\mathcal{S},h)$ amongst those $h_l(1\leq l\leq\max(k,j))$ such that $P_j(h_l,I)$.

To see that $\mathcal{M}'_j(I)$ is well-defined and effectively obtainable, note that γ' is computable and increases unboundedly with its first argument, thus ensuring the existence and computability of k, that P_j is recursive and that if $P_j(h_l,I)$ and I is consistent then Accountsfor$(h_l,S^+(I)\cap\mathcal{S})$ and so $\gamma(S^+(I)\cap\mathcal{S},h_l)$ is defined.

Evidently $\mathcal{M}'_j(I)$ has property one. Suppose I is consistent and that $P_j(h_i,I)$. If $i\leq k$ then $\gamma(S^+(I)\cap\mathcal{S},h_i)\geq\gamma(S^+(I)\cap\mathcal{S},\mathcal{M}'_j(I))$. Otherwise $\gamma(S^+(I)\cap\mathcal{S},h_i)\geq\gamma'(i,0)\geq\gamma'(k,0)\geq\gamma(S^+(I)\cap\mathcal{S},\mathcal{M}_j(I))$ $\geq\gamma(S^+(I)\cap\mathcal{S},\mathcal{M}'_j(I))$. Therefore $\mathcal{M}'_j(I)$ also has property two.

<u>Theorem 1</u> Suppose that Accountsfor is recursive. $\mathcal{M}'_2$ matches an explanation of I for $\mathcal{S}$ in the limit, if I has one. If $\gamma(S^+(I^t)\cap\mathcal{S},h)$ converges for all h explaining I for $\mathcal{S}$, then $\mathcal{M}'_2$ will eventually guess only hypotheses, h, which minimise $\lim_{t\to\infty}\gamma(S^+(I^t)\cap\mathcal{S},h)$.

<u>Proof</u> From theorem 3.1, $\mathcal{M}_2$ identifies some explanatory hypothesis h in the limit. So there is a τ such that if $t\geq\tau$, $\mathcal{M}_2(I^t)=h$. Hence, when $t\geq\tau$, the k calculated by $\mathcal{M}'_2$ will be independent of t and only a finite number of hypotheses will be considered by $\mathcal{M}'_2$. From the properties of P_2 developed in the proof of theorem 3.1, if h_i does not explain I for $\mathcal{S}$ then eventually $P_2(h_i,I^t)$ will always be false. Therefore in the limit $\mathcal{M}'_2$ will choose only hypotheses explaining I for $\mathcal{S}$. That is, $\mathcal{M}'_2$ matches an explanation of I for $\mathcal{S}$ in the limit. The

The second part of the theorem is obvious. This concludes the proof.

When both accountsfor and consistent are recursive, $\mathcal{M}_1'$ has the same limiting behaviour as $\mathcal{M}_2'$ although, of course, its local behaviour is better.

<u>Theorem 2</u> Suppose I has an explanation, h_i for $\mathcal{S}$. $\mathcal{M}_3'$ approaches an explanation. If $\gamma(S^+(I^t) \cap \mathcal{S}, h_i)$ is bounded as $t \to \infty$, then $\mathcal{M}_3'$ will only consider finitely many hypotheses and will match an explanation of I for $\mathcal{S}$ in the limit. If $\gamma(S^+(I^t) \cap \mathcal{S}, h')$ converges for all h' explaining I for $\mathcal{S}$, then $\mathcal{M}_3'$ will eventually guess only hypotheses, h', which minimise $\lim_{t \to \infty} \gamma(S^+(I^t) \cap \mathcal{S}, h')$.

<u>Proof</u> The properties of P_3 developed in the proof of theorem 3.3 show at once that $\mathcal{M}_3'$ approaches an explanation. If $\gamma(S^+(I^t) \cap \mathcal{S}, h_i)$ is bounded, so is $d(S^+(I^t) \cap \mathcal{S}, h_i)$. Suppose $t_1 \geq \max_{0<t<\infty} d(S^+(I^t) \cap \mathcal{S}, h_i)$. Then $M_A(h_i, S^+(I^t) \cap \mathcal{S}, t_1)$. Therefore, if $t \geq t_1$, $P_4(h_i, I^t)$. From the properties of γ we see that for some k, $l \geq k$ implies that $\gamma(S^+(I^t) \cap \mathcal{S}, h_l) \geq \gamma(S^+(I^t) \cap \mathcal{S}, h_i)$. Consequently $\mathcal{M}_3'$ only considers finitely many hypotheses. Since in general, it approaches an explanation, it must, in this case match one. The last part of the theorem is obvious and this concludes the proof.

5. Generalisation and hypothesis learning theory

The algorithms and theory developed in the previous chapters provide a class of induction machines each of which chooses a nicest explanatory hypothesis generalising a given H_0 and consistent with its knowledge. We will take a brief look at the behaviour in the limit of one such machine in a decidable case.

The hypothesis space, Hyp, is the set of finite sets of clauses containing no function symbols, other than constants.

Phen is the set of ground non-tautologous clauses containing no function symbols other than constants.

Accountsfor(H,H_0) iff $H \leq H_0$.
Consistent(H,H_0^+,H_0^-) iff $\forall H \wedge \bigwedge_{C \in H_0^+} C \wedge \bigwedge_{D \in H_0^-} D$ is consistent.

Since Accountsfor and Consistent are both recursive there is an algorithm which identifies an explanation in the limit. However, we will see that if we take $\precsim = \precsim_{cpg}$, any algorithm which chooses a nicest explanation need not match an explanation in the limit even on natural information sequences which arise from repeated presentations of the formal problem.

Suppose that f_i $(i \geq 0)$ is a sequence of ground literals and Ev is a map from $\{f_i \mid i \geq 0\}$ to conjunctions of ground literals such that $\overline{Ev(f_i)} \cup \{f_i\}$ is in Phen and $\{f_i \wedge Ev(f_i) \parallel i=1,n\}$ is consistent. Let $Ev(f_i) = e_{i1} \wedge \ldots\ldots \wedge e_{ij(i)}$ where the e_{ij} are ground literals and we let $\mathcal{E} = \{\overline{\mathbf{Ev}(f_i)} \cup \{f_i\} \mid i \geq 0\}$; a <u>natural</u> complete and consistent information sequence for the f_i and $\mathbf{Ev}$ is one such that $S^+(I) \supseteq \mathcal{E} \cup \{e_{ij} \mid i \geq 0,\ 1 \leq j \leq j(i)\}$ and if $i' > i$, $+(\overline{\mathbf{Ev}(f_i)} \cup \{f_i\})$ occurs, doing so before $+(\overline{\mathbf{Ev}(f_{i'})} \cup \{f_{i'}\})$, in I.

Recollect the $D_2 j$ of chapter 3, section 3.3.2 which provided an example of an infinite strictly decreasing chain and the γ^i_j of the representation theorem, theorem 3.3.3.2.4 and their properties.

One can find f_i and an $\mathbf{Ev}$ satisfying the necessary conditions outlined above such that:

$$f_{2i} = Q(x_{[1,2^i]})\ \gamma^1_{n(i)}$$
$$f_{2i-1} = Q(x_{[1,2^i]})\ \gamma^2_{n(i)}$$
$$\overline{\mathbf{Ev}(f_{2i})} = D_2{}^i\ \gamma^1_{n(i)}$$
$$\overline{\mathbf{Ev}(f_{2i-1})} = D_2 i\ \gamma^2_{n(i)}$$

for suitably large $n(i)$ and all $i \geq 1$.

Now such an f_i and $\mathbf{Ev}$ has a natural information sequence explained by $\{Q(x)\}$. Yet by the properties of the $D_2 j$ and the representation theorem 3.3.3.2.4 the nicest hypothesis, in the sense of $\prec_{cpg}$, explaining $\{\overline{\mathbf{Ev}(f_i)} \cup \{f_i\} \mid 1 \leq i \leq 2n\}$ and consistent with $S^+(I)$ and the set of negations of members of $S^-(I)$, is $D_2 n \cup \{Q(x_{[1,2^n]})\} = E_n$ say. Thus at time t_n, E_n will be chosen. As this is a strictly decreasing sequence no member of which explains I, our machine will not even match an explanation in the limit.

One might object that $\mathcal{S}$ does not contain enough instances of the explanation. If it contains all instances, we will see that choosing the nicest will match an explanation in the limit.

For, suppose H is an explanation of a natural, complete and consistent information sequence I and that $\mathcal{S} \geq H_0 = \bigcup \{H \ \gamma^i_j \mid i,j \geq 1\}$. Then for some t_0, H is equivalent to a subset of $\mathcal{G}_\phi(S^+(I^{t_0}) \cap \mathcal{S})$ by the representation theorem. Therefore if $t \geq t_0$, the nicest explanatory hypothesis will have complexity less than or equal to that of H. Let C be in H. The set $\{C \ \gamma^i_j \mid i,j \geq 1$, $C \ \gamma^i_j$ is ground and if $1 \leq j' \leq j$, $a_{ij'}$ does not occur in $C\}$ is infinite. Therefore it must eventually be the case since the complexity of the nicest explanatory hypothesis is bounded that any nicest explanatory hypothesis must contain a clause subsuming at least two members of this set, and so subsuming C itself, by the representation theorem. Therefore after some time $t_1 \geq t_0$ any nicest explanatory hypothesis must generalise H. Further, after t_1 no clause can occur in such a hypothesis which does not subsume some clause of H, as H is an explanation. Therefore by the minimality, with respect to $\geq_{cpg}$ requirement, such a hypothesis will be equivalent to a subset of $\mathcal{G}_\phi(H)$, generalising H. There is a fixed collection of such subsets of $\mathcal{G}_\phi(H)$ of equal cardinality any member of which is consistent with $S^+(I)$ and the set of negations of members of $S^-(I)$ such that eventually the nicest explanation will always be equivalent to one of this set, the choice being determined solely by power. Which has the greatest power depends on, amongst other things, the order of occurrence of the $C \gamma^{i'}_j (C \in H)$ and so, in general, any machine which chooses a nicest explanation will match rather than identify an explanation in the limit.

These results are critically dependent on $\prec_{cpg}$. Let us look instead at $\prec_{s'g}$. H $\prec_{s'g}$H' iff H has a smaller number of symbol occurrences than H or, if they have the same number of symbols, then H' $\leq$ H.

There is now a machine which will identify a nicest explanation in the limit. Let $x_1,\ldots,x_i,\ldots$ be an infinite list of distinct variables. Let $\mathcal{H} = \{H \mid$ If H contains n variables these are precisely $x_1,\ldots,x_n\}$. Notice that any H has an alphabetic variant in $\mathcal{H}$. Now there are only finitely many members of $\mathcal{H}$ with a fixed number of symbols, and $\leq$ is a quasi-ordering. Hence $\mathcal{H}$ can be enumerated as $H_1,H_2,\ldots$ where $H_1,\ldots,H_{i_1}$ have one symbol, $H_{i_1},\ldots,H_{i_2}$ have two symbols and so on, and where if H_j and $H_{j'}$ have the same number of symbols then if $H_j \leq H_{j'}$ either $j \leq j'$ or $H_{j'} \leq H_j$. This follows easily from the fact that any finite partial ordering can be enlarged to a linear one. Consequently, if $j < j'$, $H_j \prec_{s'g} H_{j'}$. Since every H has an alphabetic variant in $\mathcal{H}$, if there is an explanation of some given I there will be one in the enumeration. Consequently that machine which picks the first explanation in the enumeration which consistently explains the phenomena will identify an explanation in the limit and will always pick the nicest.

These examples show that much work remains to be done in picking niceness relations.

5. Conclusions

We have presented a generalisation of Feldman's (1970) work. Feldman remarks that one of the more interesting theoretical problems was the inference of systems with semantics. In so far as our general theory covers systems using the predicate calculus which has a semantics, we have covered this problem. Here however we see that the notion of complexity seems inadequate to apply to some of the niceness relations developed earlier.

The various machines used and developed do not behave at all in accordance with any hypothesis discovery procedure employed by practising scientists. One could look for reasons in two general directions. A better description of normal scientific practice, including the discovery methods used would lead to more realistic machines. For example one might study how old theories are modified to obtain new ones. This is a descriptive approach.

On the other hand, it may be that the machines do not behave in the way they ought to. There is no formulation of any notions of justification of criticism of hypotheses. This is a normative approach.

Leaving these general points aside, the machines all have one deficiency, they are extremely inefficient. Each one would take so long to operate that the process of hypothesis discovery would lag irretrievably far behind the process of information acquisition. We believe therefore that it would be illuminating to formalise and prove the conjecture:

Suppose Accountsfor and Consistent are recursive. Then there is a "natural" model of the axioms and choice of $\mathcal{S}$ such that no machine can efficiently identify an explanation in the limit on every (or almost every) explainable information sequence I.

Such a proof would show that it is necessary to consider special cases and methods, even from this simple point of view.