Schlag, Karl

Working Paper

Bounded Perception and Learning how to Decide

Discussion Paper, No. 972

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Schlag, Karl (1991) : Bounded Perception and Learning how to Decide, Discussion Paper, No. 972, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/221331

Discussion Paper No. 972

BOUNDED PERCEPTION AND LEARNING HOW TO DECIDE*

by

Karl Schlag

December 1991

* I would like to thank Itzhak Gilboa and Ariel Rubinstein for their useful comments and motivation.

Department of Managerial Economics and Decision Sciences
Northwestern University
2001 Sheridan Road
Evanston, IL 60201

ABSTRACT:

Consider a decision maker who must coordinate his decision with the occurrence of some phenomenon. In order to behave "optimally", the circumstances surrounding the occurrence of the phenomenon must be learned. However, there are natural bounds on the capabilities of perception. More specifically, only a fixed number of attributes may be focused on and observed in each instance.

This paper models this problem in the framework of learning concepts from positive examples involving bounded perception. For clarity and simplicity, it is assumed that for each positive example the decision maker may only observe one of its attributes. The analysis concentrates on finding optimal ways of specifying what attributes should be observed. With certain assumptions of independence we show that a class of local "hillclimbing" algorithms are essentially the only optimal ones. Additionally it is shown that patterns in the observation behavior emerge asymptotically.

The results underscore the importance of diversifying attention when acquiring knowledge.

1. INTRODUCTION:

The profitability of a decision (or action) is often dependent on the occurrence of a phenomenon, say drill for oil only when it is indeed there or invest only when the market is favorable. So the decision maker aims to coordinate his decisions with the occurrence of this phenomenon. However whether or not the phenomenon occurs is usually not revealed until after the decision is made. Hence the outcome of the decision is very much dependent on how well the decision maker can forecast the occurrence of the phenomenon. For this he evaluates the present circumstances and compares this with what he knows about what determines the occurrence of the phenomenon. The process of learning about the circumstances surrounding a phenomenon in preparation of such a (critical) decision is the subject of this paper.

A standard framework of learning when a phenomenon occurs is learning from examples. An example is data recording the circumstances surrounding the phenomenon in some state of the world together with the information whether or not the phenomenon occurred under those circumstances. If the phenomenon occurred (did not occur) then the example is called positive (negative). These circumstances are values of attributes, like the characteristics of the rocks where the oil is suspected.

Learning from examples is a frequent topic in the field of knowledge acquisition, found mostly in the machine learning and artificial intelligence literature. The emphasis is on programming computers to "learn". Some keywords are: "classification", "learning concepts" or just "formal learning theory". One widely adopted definition of learning in this context is due to Valiant [1984] (see Dietterich [1990] for a survey). A closely related term is PAC-learnable (from probably approximately correct, due to Angluin and Laird [1988]). Additional to PAC'ness, Valiant's definition requires learning in polynomial time. The model is "non Bayesian". Other approaches using the standard Bayesian analysis are emerging (see Buntine [1989] for model and survey). Typical results concern the types of information that are "learnable" and "efficient" classification of large data.

In contrast to machine learning, the capability of human perception and processing

is quite limited (see Miller [1956]). These limitations affect both the speed and the complexity of the information observation and processing. Consequently a different sort of analysis than that in the machine learning setup becomes of interest: finding optimal learning algorithms for finite horizons (instead of in the limit) and determining the future focus of attention given past observations. The selection of the focus of attention is strongly influenced by the individual's bias about what causes the phenomenon. And this in turn often causes aspects of the future decision to be predetermined while the information is gathered. In fact, often after the data has been gathered, the analysis about when the phenomenon occurs and how to decide appropriately is quite straightforward.

Therefore, in order to properly capture human decision when learning in prospect of making a decision, it is necessary to impose bounded perception restrictions and incorporate the gathering of information into the optimization. Bounded perception means that only a fixed number of attributes may be observed in each example. Although this number is fixed, the individual may determine which attributes he wants to observe. This endogenizes the individual's bias when gathering the information into the model formulation.

For clarity and simplicity, it is assumed that for each example the decision maker may only observe one of its attributes. This, however, makes it difficult to make inferences about what causes the phenomenon. Very little can be inferred (i.e., with certainty) from negative examples, therefore the individual is only given positive examples to learn from. Additionally, the "complexity" of the possible causes of the phenomenon needs to be restricted: it will be assumed that the fact that the phenomenon occurs in a state of the world (i.e., an example being positive) can be characterized as a Boolean conjunction (i.e., logically connected by "AND") of the circumstances surrounding the phenomenon. (E.g. the market is favorable <u>if and only if</u> the dollar is strong <u>and</u> the economy is growing.) This is the so-called bias, the set of all positive examples is called a concept and the set of all possible concepts is called the hypothesis space.

So the framework of our model can be described as learning concepts from positive examples involving bounded perception.

The following is a more specific description of the progress of the decision making

and how it will be modelled:

An individual wants to coordinate his decision with the occurrence of a phenomenon. He uses a fixed number (r) of binary valued attributes (characteristics) of the state of the world to analyze the circumstances surrounding the occurrence of the phenomenon. These attributes are assumed to be sufficient for determining whether or not the phenomenon occurs.

The model has three phases. In the first phase, the so-called initial phase, nature determines what causes the phenomenon. Then comes the learning phase in which the individual is given a fixed number (n) of rounds to learn under which circumstances the phenomenon takes place, or equivalently what circumstances cause it to occur. In each round, nature presents the individual with a state of the world (independent of the previous rounds) in which the phenomenon occurs (i.e., a positive example). For each state the individual specifies the attribute he wants to observe and then he observes its value in that state. Finally, after the learning phase is over, the individual may make his decision (decision phase). For this, nature presents him with a sequence of random states of the world. For each presented state, he observes the values of all of the attributes but not whether the phenomenon occurs. The individual may make his decision in at most one of these states.

It is assumed that the individual only makes his decision in a state in which the phenomenon occurs and prefers to do this as early as possible. This assumption is closely related to the type I/type II error approach in statistics. A scenario is presented that gives rise to this condition of extreme aversion to making a mistake when coordinating the decision with the occurrence of the phenomenon. Finally it should be noted that this condition also follows from the abstract definition made in Valiant's model (see Shvytser [1988]).

The model will be set in a Bayesian framework: the individual assesses priors to the occurrence of states of the world and to the possibility that an attribute is necessary for the phenomenon to occur.

The main result of this paper is the following: assume that the occurrence of each

attribute and its effect on the occurrence of the phenomenon are independent and independent of the other attributes. Then it is shown that an optimal algorithm (specification of attributes to be observed in the learning phase and when to decide in the decision phase) is essentially a local "hill-climbing" algorithm, i.e. it is constructed as if each round of the learning phase were the last round before the decision has to be made. The decision is made in the first presented state in which the values of all the attributes were observed in the learning phase. If there is no such state, the decision will not be made.

These locally optimal algorithms are easy to implement and are independent of the length of the learning phase. Additionally, after the learning phase has lasted a while, the specification essentially follows a simple pattern: specify each attribute a fixed number of times (dependent on the attribute) and then proceed to the next one. If both values 0 and 1 are observed then stop specifying this attribute.

Thus as long as the learning phase lasts, an attribute will be specified again and again unless all values have been observed and hence no more information can be gained by specifying it.

To summarize, the model of learning presented uses priors to aid in the gathering of information, but avoids "guessing" in the critical decisions. The results underscore the importance of diversifying attention when acquiring knowledge.

Finally, there is a lot of room for future research: a counter-example is presented which demonstrates that without the independence assumption, these locally optimal algorithms are not necessarily (globally) optimal.

2. THE MODEL:

Assume that there are r attributes of the states of the world that the individual uses to classify the causes of the phenomenon. Each attribute can only obtain values 0 or 1. The states of the world will not be explicitly modeled.

Let $\Omega := \{0,1\}^r$ be a set of items.

For all $x \in \Omega$, $1 \leq k \leq r$, denote by x(k) the k-th attribute (component) of the item x, i.e. $x = (x(1),..,x(r))$.

An item can be thought of as a vector of attributes.

We assume that the occurrence of the phenomenon is binary, i.e., either it occurs or it does not. Furthermore, the r attributes are sufficient to classify its occurrence, i.e., whether or not the phenomenon occurs is completely determined by the values of the attributes. The set of all items in which the phenomenon occurs is called a concept.

So a concept (associated to the phenomenon) can be formally defined as a subset $B \subseteq \Omega$. An item $x \in \Omega$ is called positive if the phenomenon occurs when each attribute k obtains the value x(k) for all $1 \leq k \leq r$, i.e. $x \in B$, and it is called negative otherwise.

An item presented to the individual together with the information whether or not the phenomenon occurs, is called an example. It will be assumed that in the learning phase of the model the individual has bounded perception and thus can only observe one attribute of each example. This assumption restricts the class of concepts one can hope to learn. For instance, if $r = 2$ then the concept $B := \{(0,0),(1,1)\}$ cannot be distinguished from the concept $B' := \{(0,1),(1,0)\}$ if, as assumed, only one attribute can be observed per example and the examples are independent of each other. We will restrict our analysis to learnable phenomena.

A cylindrical phenomenon is one whose occurrence depends on some attributes taking fixed values and others remaining arbitrary. Alternatively one might say that its occurrence is a Boolean conjunction (logically connected by "AND") of values of attributes. Therefore learning cylindrical concepts is called learning Boolean conjunctions in the machine learning literature.

Formally define $F' := \{0,1,*\}^r$. Any $f \in F'$ is associated with a concept B_f, where $B_f := \{x \in \Omega$ s.t. $x(i) = f(i)$ whenever $f(i) \neq * \}$. The class of concepts, one of which will be selected to be learned, is $\{B_f$ s.t. $f \in F'\}$ and will be denoted by F.

So formally a cylindrical phenomenon is defined as a phenomenon that is associated with a concept in F.

Abusing notation we will not distinguish between $f \in F'$ the associated $J \in F$ (i.e., $J = B_f$) and will write $J = (J(1),..,J(r))$ where $J(k) := f(k) \in \{0,1,*\}$ for $1 \leq k \leq r$. If J is the concept related to the phenomenon then $J(i) = 0$ ($J(i) = 1$) means that the i-th attribute has to have the value 0 (1) for it to occur and $J(i) = *$ means that the i-th attribute isn't relevant for its occurrence.

It is easy to see that if the concept inferred from the observed data is not cylindrical, then some additional dependencies between the attributes are assumed that cannot be verified if only one attribute per example is observed. Hence we restrict our set of possible concepts to F, i.e, F is the hypothesis space.

There is another restriction caused by assuming that only one attribute can be observed per example and that the examples are independent of each other: "very little" can be inferred from negative examples. Let $B \in F$ be the cylindrical concept associated with the phenomenon. Given any prior history of observations, if the first attribute is specified in a negative example and the value 0 is observed, then the only facts that can be inferred from this (with certainty) is that there exist negative examples, i.e. $B \neq \Omega$, and that the phenomenon cannot occur exactly when the first attribute obtains the value 0, i.e., $B \neq (0,*,..,*)$.

It can be shown that in order to coordinate a decision with the occurrence of a phenomenon, the individual is essentially forced to gather information only from positive examples. (A more precise statement can be made, but the resulting analysis would go beyond the scope of the simple model presented in this paper.) Hence we restrict our analysis from the beginning to learning from positive examples only.

Let P be the underlying probability measure and let X denote a random variable

taking values in Ω. $P(X=x)$ is interpreted as the individual's prior probability that a state of the world will occur that has the value x(k) on the k-th attribute, for all $1 \leq k \leq r$. Let Q denote a r.v. taking values in F, where $P(Q=J)$ stands for the individual's prior probability that the concept related to the phenomenon will be J. It reflects his initial knowledge (or a hunch) about the phenomenon.

To simplify the presentation, it is assumed that the random variables X and Q are independent, i.e. $P(X=x \mid Q=J)=P(X=x)$ for $x \in \Omega$ and $J \in F$. Intuitively this means that the values of the attributes occur independently of the phenomenon. All results can easily be extended to the more general case.

The setup for learning a concept from n positive examples by an individual with bounded perception in order to coordinate it with a decision, consists of three phases: the initial phase, the learning phase and the decision phase.

In the initial phase, nature determines what causes the phenomenon. Due to the implications of the individual's prior, we may assume that nature uses the r.v. Q when randomly selecting a concept from the class of concepts F.

The learning phase in which the individual gathers information, comes after the initial phase and consists of n rounds. Before each round, the individual specifies which attribute he wishes to see based on his past observations and beliefs. Then nature randomly selects a positive example and the individual is shown the value of the attribute he had selected. Alternatively the individual is allowed to refrain from specifying an attribute in which case he observes nothing.

The n positive examples are chosen independently and independently of the individual's decisions. The letter n will be reserved to denote the number of rounds in the learning phase.

To formalize the individual's behavior we need to describe how the individual makes his decision about which attribute he specifies next given his past observations:

Denote by H_j the set $\{\{a_1,..,a_r\} \times \{0,1\},\varnothing\}^j$, interpreted as the set of histories of

observations of length j ($H_0 := \varnothing$). So for example $((a_4,0),\varnothing,..)$ means that in the first round, the fourth attribute obtained the value 0, and that the second round was skipped. The notation (a_i,y) is used instead of (i,y) to clarify the difference between the component that denotes the number of the attribute and the component that denotes its value. It follows that the set of all histories of length up to k can be written as $\bigcup_{j=0}^{k} H_j$. Let H denote $\bigcup_{j=0}^{\infty} H_j$, the set of histories of arbitrary length. Let T be a description of how the individual makes his decisions in the learning stage, then T can be written as a function

$$T: \bigcup_{j=0}^{n-1} H_j \rightarrow \{a_1,..,a_r\} \cup \{\varnothing\} \ .$$

For $h \in H_{k-1}$, T(h) represents the attribute specified by the individual to be observed in round k given his previous observations h. As above, $T(h) = \varnothing$ means that no attribute is specified in round k.

T is called a learning procedure, the set of all such T is denoted by $\Gamma(r,n)$, where r is the number of attributes and n is the number of rounds in the learning phase.

For later analysis we need to define the set of histories that can possibly be observed at the end of a learning phase of arbitrary length using the procedure T. This set, denoted by H(T), is defined by the following properties:

For any k, $1 \le k \le n$, let $h \in H_k$ be written as $(w_1,..,w_k)$ where $w_i \in \{a_1,..,a_r\} x \{0,1\}$, $1 \le i \le k$. Then $h \in H(T)$ if and only if $T(\varnothing) = w_1$ and for any $1 \le j \le k-1$,

- $w_{j+1} = \varnothing$ implies that $T((w_1,..,w_j)) = \varnothing$,
- $w_{j+1} = (a_i,y) \in \{a_1,..,a_r\} x \{0,1\}$ implies $T((w_1,..,w_j)) = a_i$.

Notice that the values of T can be arbitrary on the set $H \backslash H(T)$, because they have no influence on the behavior in the learning phase.

After the n rounds of the learning phase are completed, the so-called decision phase takes place. Nature presents the individual with a sequence of items, called decision opportunities. These are assumed (by the individual) to be selected as i.i.d. r.v.s, distributed according to X and independent of the sampling and choices made in the learning phase. For each presented item, the individual observes the values of all of the attributes but not whether or not the phenomenon occurs. The decision maker may choose at most one decision opportunity in which to make his decision. After he makes his decision, the game is over.

We will call a decision or decision opportunity good (bad) if the phenomenon occurs (does not occur).

Let $W := \bigcup_{j=0}^{\infty} \Omega^j$ be the set of all possible histories of decision opportunities in the decision phase. Formally, the individual's behavior in the decision phase can be represented by the function g: H x W x $\Omega \rightarrow$ {'decide','wait'}, where, for $h \in H$, $w \in W$ and $x \in \Omega$, g(h,w,x) = 'decide' means that he makes his decision and g(h,w,x) = 'wait' means that he will wait (i.e. not make his decision) when nature presents him the decision opportunity x given that he has already observed h in the learning phase and w in the decision phase. We will call such a function a reaction function.

Any pair (g,T) satisfying the definitions/conditions above is called a feasible algorithm.

We can summarize the model as a game played as follows:

(INITIAL PHASE)

- Nature chooses a concept, say J, from the class of concepts F according to the r.v. Q.

(LEARNING PHASE)

For k = 1,..,n:

- Before round k the individual decides whether he wants to observe an attribute and if so, which attribute he wants to observe the value of, say i(k), $i(k) \in \{1,..,r\}$.

• Nature then chooses a positive example, say $x \in J$, according to the conditional distribution of X given J - this is done independently of the sampling and of the individual's behavior in the previous rounds.

• The individual observes x(i(k)).

(If the individual decides not to observe an attribute then he does not observe anything.)

(DECISION PHASE)

• Nature chooses a sequence of items from Ω (r.v. X) - independently of each other and of the sampling in the learning phase.

• For each item, the individual observes the values of each attribute (but not whether or not the phenomenon occurs) and can either choose to make his decision or to wait.

• The game is over when the individual makes his decision.

Here a few notes should be made about what can be inferred from the observations in the learning phase about the phenomenon. The specific nature of the class of possible concepts F allows to make inferences about which items must be positive. Consider the individual specifying a_i (i.e. the i-th attribute) in the i-th round and him observing $y_i \in \{0,1\}$ for $i = 1,..,r$. Then after round r he can be sure that $(y_1,..,y_r)$ is a positive item. This follows directly from the properties of F. We will assume that the number of rounds in the learning phase is at least as large as the number of attributes (i.e., $n \geq r$) to allow for such inferences to be made.

On the other hand, it is impossible for the individual to say for certain, based on his observations h, that an item <u>cannot</u> be a positive one if $P(Q = \Omega) > 0$. That means that any decision opportunity might be a "good" one.

We also recall that the individual may only observe one attribute in each round of the learning phase but observes all of the attributes in the decision phase. Imposing a bound on the individual's capability of perception when he is gathering data and learning at the same time but not when he is applying previously accumulated knowledge, is quite natural.

The situation we have in mind is that in the learning phase, the individual is obtaining his data from unknown environments or related problems, observing the actions

of other individuals as a bystander, or forced to make queries. In any case the data is hard to obtain and hence the bound on the information received per state is imposed (only one attribute per example). On the other hand, in the decision phase, the individual applies his accumulated knowledge and as such changes from the observer to the actor. He is in a known environment and is involved in the decision. Therefore, it can be intuitively justified that a bound on his perception capabilities as in the learning phase is not necessary.

Of course there are other scenarios that lead to similar models, for instance assuming that there is a cost to observing an attribute in the learning phase or that the individual must generate the phenomenon in the decision phase by fixing certain values of attributes (at a cost). The analysis of these related models is similar to the present formulation and equivalent results can be derived.

Returning to the analysis, how should feasible algorithms be compared so that a "best" one can be selected amongst them?

Here is a table of the possible constellations that can occur when the individual is confronted with a decision opportunity:

\ outcome behavior \	phenomenon occurs	phenomenon does not occur
makes decision	coordinates correctly	makes decision in wrong state
waits for decision	wastes decision opportunity	makes right choice to wait

We will call the individual's reaction function <u>safe</u> if it is such that he only makes his decision in a state in which the phenomenon occurs. Formally, the reaction function g is called safe if $g(h,w,x) = \text{'decide'}$ implies that x is a positive item ($h \in H, w \in W, x \in \Omega$).

It will be assumed that the individual selects under all feasible algorithms one using the following conditions: he only makes his decision in a state in which the phenomenon occurs with certainty and prefers to do this as early as possible.

Where do these conditions come from, what is the intuition behind them? The individual's preferences are assumed to be as follows: the individual prefers making his decision when the phenomenon occurs to waiting for the next opportunity. But he would rather wait than get the bad outcome of making the decision when the phenomenon does not occur. Finally, given that he waits, he does not care whether the phenomenon occurs or not. The learning is assumed to take place in an unknown environment, i.e., the individual has insufficient experience related to the phenomenon. He does not have enough confidence in his priors to base his decision on them. He is uncertainty averse when it comes to the possibility of getting the bad outcome of making the decision when the phenomenon does not occur. Since it is assumed that there are more rounds in the learning phase than there are attributes (i.e., $n \geq r$), it is possible to make safe decisions. In such a context it is intuitive to assume that the individual chooses a strategy that avoids the bad outcome (see Gilboa [1988] for a applicable framework and a related explanation of the Allais paradox). The preference for making the decision as early as possible comes from discounting future payoffs.

This approach is closely related to the type I/type II error approach in the statistical testing theory and to the PAC model of Valiant [1984]. In the latter case, the learning algorithm should be "optimal" (in an appropriate sense) for arbitrary priors on the examples and on the causes of the phenomenon. Applying this approach to learning from positive examples leads to the "safe" condition demanded here (see Shvytser [1988]).

One more definition will prove useful: (g,T) is called <u>non-wasteful</u> if the individual stops specifying attributes in the learning phase if the algorithm (looking ahead to decision phase) cannot be improved by further search.

This means that the individual does not want to make useless searches. This behavior can be justified by assuming a (small) cost for specifying an attribute. Another way to embed this axiom in the optimality, is to assume a lexicographic utility function that prefers stopping to search all else equal.

Summing up, we can now formulate the optimization problem:

find (g^*,T^*) that are feasible and non-wasteful and with g^* safe so as to minimize the expected time of making the decision.

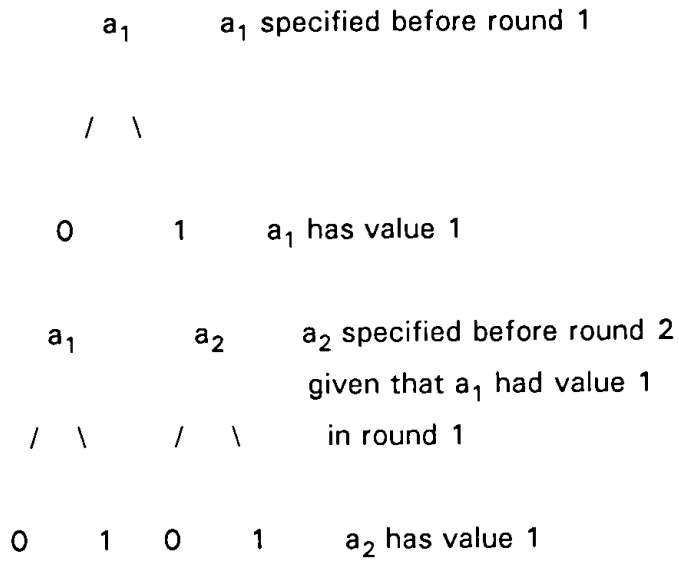
Since the above problem is finite, an algorithm that satisfies the above criterion exists. Such an algorithm will be called globally optimal (or just optimal).

Before we continue with the analysis, a note should be made on randomization: in the present setup, randomizing between strategies is not allowed. This was done to keep things simple, and nothing is lost since randomization cannot improve the learning algorithm w.r.t. the way the optimality criterion is presently formulated. However one might imagine to invoke some randomization whenever there are equally good strategies to avoid bias in selecting attributes.

It should be mentioned that since the sampling is done independently and the objective function does not depend on the round in which the data was obtained, there is no incentive to skip a round if the individual still wants to observe attributes. In view of the non-wastefulness condition it will be assumed that once the individual decides not to observe an attribute, he will stop specifying attributes in the rest of the learning phase. Hence w.l.o.g. we may exclude the empty set from being a part of any history (except of course the initial vacuous history), so redefine H_j for $j \geq 1$ as $\{\{a_1,..,a_r\} \times \{0,1\}\}^j$. That means if h is the current history and $\varnothing$ is specified in round k, then the history up to and including round k is also h.

How do learning procedures look? The simplest way of looking at a procedure T is to consider its graphical representation (see the following example). Because of the obvious features, we will also refer to learning procedures as (decision) trees. With this representation in mind, a history in H_j is said to have the length j $(0 \leq j \leq n)$, and a path is a history $h \in H(T)$ of maximum length. Formally, $h \in H(T)$ is called a path if $T(h) = \varnothing$ or $h \in H_n$. A tree is said to have depth k if k is the length of the longest path in H(T).

EXAMPLE: The following is a learning procedure $T \in \Gamma(2,2)$



After, say, $a_i = 0$ was observed (which means that the value of the i-th attribute of the presented positive example was 0), the hypothesis that a_i must be 0 suggests itself, but is contradicted once $a_i = 1$ is observed. Hence if both $a_i = 0$ and $a_i = 1$ appear on a path, we say that a_i was contradicted.

3. OPTIMAL DECISIONS:

We will assume that the priors on the occurrence of the items and on the possible concepts have full support, i.e., for every $J \in F$: $P(Q=J)>0$ and for every $x \in \Omega$: $P(X=x) > 0$.

For all $1 \le i \le n$, $y \in \{0,1\}$ and $h \in H$ we say that $(a_i,y) \in h$ if there exists $1 \le j \le k$ s.t. $h(j)=(a_i,y)$ where $h=(h(1),..,h(k))$ and $h \in H_k$,i.e., the value y has been observed for the i-th attribute in the history h.

Further denote by S(h) the set of attributes specified in the history h and by Z(h) the set of attributes that were contradicted in h, formally
$S(h):=\{i$ s.t. $(a_i,y) \in h$ for some $y \in \{0,1\}\}$ and $Z(h):=\{i$ s.t. $(a_i,0) \in h$, $(a_i,1) \in h\}$.

For every attribute that appeared in a history h but was not contradicted, denote by $s_h(j)$ the value that was observed and by $v_h(j)$ the value that was not. Formally, for every $h \in H$, $j \in S(h) \backslash Z(h)$, define $s_h(j) \in \{0,1\}$ s.t. $(a_j,s_h(j)) \in h$ and $v_h(j):=\{0,1\} \backslash s_h(j)$ (so $(a_j,v_h(j)) \notin h$).

Because an optimal algorithm has to have a safe reaction function, i.e., never makes a bad decision, and the concept is assumed to be "cylindrical", it is easy to characterize when the individual makes his decision:

We will call a reaction function g stationary if the choice to decide or to wait is independent of the decision opportunities that have already been passed up, i.e., $h \in H$, $u,w \in W$, $x \in \Omega \Rightarrow g(h,u,x)=g(h,w,x)$.

For a stationary reaction function g and a history $h \in H$ let B(g,h) be the set of decision opportunities in which the decision is made, formally,
$B(g,h):=\{x \in \Omega$ s.t. $g(h,x)=$'decide'$\}$.

We will also refer to B(g,h) as a stationary reaction function.

Furthermore we will make the following definitions:

• define $A:\bigcup_{j=0}^{\infty} H_j \rightarrow F \cup \{\varnothing\}$ so that

if $S(h) \neq \{1,..,r\}$ then $A(h) = \varnothing$

if $S(h) = \{1,..,r\}$ then $A(h) \in F$ and $A(h)(i) := \begin{cases} s_h(i) \text{ if } i \notin Z(h) \\ * \text{ if } i \in Z(h) \end{cases}$

• define for every history $h \in H$: $EV(h) := \sum_{J \in F} P(X \in A(h))\ P(h \mid Q = J)\ P(Q = J)$ and define

$EV(T) := \sum_{h \in H(T)} EV(h)$.

EV(T) is the a priori probability of making a decision in a decision opportunity using the stationary reaction function A(h) and the learning procedure T.

THEOREM 1:

A feasible algorithm (g^*, T^*) is globally optimal if and only if the following two conditions hold

(i) g^* is stationary and A(h) is the corresponding stationary reaction function, i.e. for every $h \in H(T)$, $B(g^*, h) = A(h)$.

(ii) $T^* \in \text{argmax}\ \{EV(T) \text{ s.t. } T \in \Gamma(r,n) \text{ and } T \text{ non-wasteful}\}$.

So there is a unique reaction function for optimal algorithms and it is stationary. The decision is made in the first decision opportunity in which the values of the attributes were all observed in the learning phase. This behavior can be interpreted as: "seeing it is believing it".

Our further analysis concentrates on finding the best learning procedures, keeping the reaction function fixed $A(\bullet)$. A learning procedure $T \in \Gamma(r,n)$ will be called (globally) optimal if the corresponding algorithm using the optimal reaction function $A(\bullet)$ is a globally optimal. The set of all such trees will be denoted by $\Gamma^*(r,n)$. Interpreting theorem 1, a learning procedure is globally optimal if it maximizes the probability of making a decision in the first decision opportunity using the rule "seeing it is believing it" and implementing the non-wasteful condition.

Also notice that the optimality condition can now be stated solely in the context of learning concepts (i.e when the phenomenon occurs): maximize the probability of being able to forecast the phenomenon with certainty (implementing the non-wasteful condition).

A simple conclusion about optimal algorithms can now be made: if $S(h) \neq \{1,..,r\}$ then $A(h) = \varnothing$ and therefore the decision will not be made in the decision phase and $EV(h) = 0$. So a path does not contribute to the value of the algorithm unless each attribute has been specified on it, because otherwise the decision is not made in the decision phase.

Before presenting the proof, let us illustrate the optimal reaction function by looking at a particular example when $r = 2$: $A((a_1,0),(a_2,1),(a_1,1)) = (*,1)$.

To simplify notation we will write $P(A \mid J)$ instead of $P(X \in A \mid Q = J)$ from now on $(A \subseteq \Omega,\ J \in F)$.

PROOF of theorem 1:

Since the occurrence of the phenomenon is not revealed in the decision phase, the individual cannot discover new positive examples that he is certain about as the decision phase progresses. Additionally the individual prefers to make his decision as early as possible. Hence, the individual makes his specifications in the learning phase to maximize the probability of making a good decision in the first decision opportunity and then stays with the same reaction function in the following decision opportunities.

It is easy to see that a reaction function g is safe if and only if for every $h \in H$,

$B(g,h) \subseteq A(h)$: (i) if the i-th attribute wasn't specified in h ($i \notin S(h)$) then the phenomenon could depend on a_i being 0 or it being 1 (or either). Together with the assumption that Q has full support, the only way to avoid making a bad decision is to wait.

(ii) if the i-th attribute obtained the value 0 but was not contradicted in h, then making the decision independently of the i-th attribute, i.e. $B(g,h)(i) = *$, might cause a bad decision when the i-th attribute of the decision opportunity is 1 and the concept determines that i-th attribute must be 0 for the phenomenon to occur.

Now if we can show that the probability of making a good decision is strongly monotonic in B(g,h) w.r.t. the strict inclusion ($\subset$), then the proof will be complete: let

$EV(h,C) := \sum_{J \in F} P(C)\ P(h \mid J)\ P(Q = J)$, then the probability of making a good decision is

$\sum_{h \in H(T)} EV(h,B(g,h))$ and will be denoted by EV(g,T). Because of the additivity of EV(g,T) w.r.t. histories, it is enough to show that EV(h,C) is strictly monotone in C, but this is quite straightforward: $C \subset C' \subseteq \Omega$, $C \neq C' \Rightarrow EV(h,C) < EV(h,C')$. Notice that the strict inequality follows directly from $P(Q = \Omega) > 0$. This completes the proof of theorem 1.
* * * *

In view of the theorem 1, our further analysis will concentrate on the various learning procedures, keeping the optimal reaction function fixed. In this context, we will sometimes refer to EV(T) as the <u>value</u> of the algorithm.

We need to adjust the definition of the set of all learning procedures (trees) involving r attributes and n rounds $\Gamma(r,n)$ to include fixed histories: let $\Gamma(r,n,h)$ be the set of procedures of depth at most n that have the fixed history h in the first k rounds, where $h \in H_k$ and $k \leq n$. So $T \in \Gamma(r,n,h)$ is a tree of depth at most n-k attached to the fixed history h of length k. Especially in the case of $k = n$, $\Gamma(r,n,h)$ is just the fixed history h (degenerate tree).

Formally, if $h = \varnothing$, i.e., there is no fixed history, then $\Gamma(r,n,\varnothing) := \Gamma(r,n)$. If $h \in H_k$, $1 \leq k \leq n$ then

$\Gamma(r,n,h) := \{T \text{ s.t. } T: \bigcup_{i=0}^{n-k-1} (h,H_i) \rightarrow \{a_1,..,a_r\} \cup \{\varnothing\}\}$, where $(h,H_i) \subseteq H$ defined s.t.

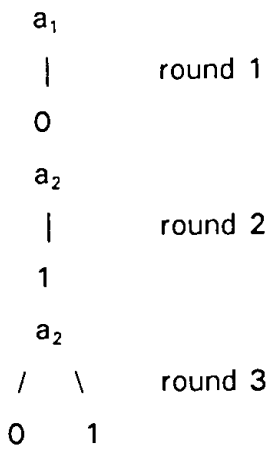
$h' \in (h,H_i)$ if $h' \in H_{k+i}$ and $h'(j) = h(j)$ for $1 \leq j \leq k$.

All definitions w.r.t. $\Gamma(r,n)$, such as H(T) and EV(T), will be extended to $\Gamma(r,n,h)$ in the natural way.

For any tree $T \in \Gamma(r,n)$ and history $h \in H(T)$ define the subtree of T with initial history h, denoted by T/h, by taking the tree T and leaving away all paths that don't follow h in the first rounds. Essentially, T/h is just T with a restricted domain. Formally, for any $T \in \Gamma(r,n)$ and $h \in H(T)$ define $T/h \in \Gamma(r,n,h)$ s.t. for every $h_1 = (h,h') \in H(T)$, $T/h(h_1) := T(h_1)$.

The following class of trees with fixed histories will be important for the further analysis: for any history h and any attribute a_i, let $h \cup a_i$ be the tree of depth $k+1$ that has the fixed history h in the first k rounds, and specifies a_i in the $k+1$-st round, where k is the length of history h. Formally: $h \cup a_i \in \Gamma(r,k+1,h)$ s.t. $(h \cup a_i)(h) := i$ $(h \in H_k)$.

EXAMPLE: The following is the graphic representation of $h \cup a_2 \in \Gamma(2,3,h)$ where $h = ((a_1,0),(a_2,1))$.



4. LOCALLY OPTIMAL TREES:

Let us now restrict our analysis to a certain class of trees that are constructed optimally in each round as if that round were the last round before the decision phase. In other words, in each round of the learning phase, the individual specifies the attribute that maximizes the increase in the probability of making a safe decision in the first decision opportunity. He disregards any influence this specification may have on future specifications in the learning phase. Additionally, because of the non-wastefulness condition, the individual only specifies this attribute if it improves the present tree. This one round horizon optimization is just the standard "hill-climbing" algorithm for this setup and will be referred to as "locally optimal".

The importance of these algorithms arises quite naturally from the intuition behind the model that led to the assumption of bounded perception. The lack of time, the complexity of the problem and the resulting bounded perception make such an easily implementable algorithm very desirable.

The concept of "locally optimal" only influences the construction of the algorithm in the learning phase, so the "locally optimal" reaction function is the same as the globally optimal one (see theorem 1).

It follows directly from theorem 1 that the concept of "locally optimal" as stated above, is too strict: if there is more than one attribute ($r > 1$), then local optimization together with the non-wastefulness condition will imply that no attribute will be specified in the learning phase. This is because every possible specification of an attribute in the first round will not change the evaluation of the tree if that round would be the last round (see theorem 1). Requiring that the tree is non-wasteful therefore implies that nothing is specified. Even without the assumption of non-wastefulness, the tree does not improve if the same strategy is always specified, which is a locally optimal strategy if $r > 2$.

So strictly enforcing the idea of local optimality does not lead to any desired outcomes. However, if the individual is to be able to calculate the safe decisions, then he needs to have every characteristic specified on every path at least once. Why not specify each characteristic in the first r rounds? Note that this modified strategy is also easily implemented and intuitive.

Therefore a slight modification of the local one-round optimization has to be made: the individual specifies all attributes in the first r rounds on every path, and from round $r+1$ on, he chooses a locally optimal attribute in every round if it improves the algorithm (i.e. he implements the non-wastefulness condition).

Formally, a learning procedure T ($T \in \Gamma(r,n)$) is called locally optimal (denoted by $T \in \Gamma^L(r,n)$) if for every history $h \in H(T) \cap H_k$,

- if $0 \leq k < r$ then $T(h) \in \{1,..,r\} \setminus S(h)$
- if $r \leq k < n$ then $T(h) \in \underset{i}{\text{argmax}} \{ EV(h \cup a_i) \text{ s.t. } EV(h \cup a_i) > EV(h) \}$.

Similarly, a tree in $\Gamma(r,n,h)$ is called locally optimal given h if first each attribute that was not specified in h is specified, and from there on optimizing locally and implementing the non-wastefulness condition.

Formally, a learning procedure T ($T \in \Gamma(r,n,h)$) is called locally optimal given h ($T \in \Gamma^L(r,n,h)$) if for every history $h' \in (h,H)$,

- if $S(h') \neq \{1,..,r\}$ then $T(h') \in \{1,..,r\} \setminus S(h')$
- if $S(h') = \{1,..,r\}$ then $T(h') \in \underset{i}{\text{argmax}} \{EV(h' \cup a_i) \text{ s.t. } EV(h' \cup a_i) > EV(h')\}$,

where $h' \in (h,H)$ means that there exists $h^o \in H$ so that $h' = (h,h^o)$.

It follows directly that the property of local optimality is independent of the length of the learning phase (n). If the learning stage happens to last longer and the individual is implementing a locally optimal tree, then he does not have to rethink his whole strategy, he can just extend his present tree. Vice versa, if the learning stage happens to last shorter than expected, the locally optimal tree still remains locally optimal. So locally optimal algorithms can also be implemented in a more natural setup where the individual is uncertain about the number of rounds in the learning phase.

Let us extend locally optimal trees to arbitrary depth. Denote the set of trees of

arbitrary depth by $\Gamma(r,\bullet)$, formally $\Gamma(r,\bullet) := \{T: \bigcup_{k=0}^{\infty} H_k \rightarrow \{1,..,r\} \cup \{\varnothing\}\}$.

Denote by $\Gamma^L(r,\bullet)$, the set of locally optimal trees of arbitrary depth. This set is well defined and non-empty by definition of local optimality. The definition of a path now has to be slightly altered. A path is a sequence of histories, each of which is a prefix of the next, formally $(h_i)_{i \in \mathbf{N}}$ is a path if $h_1 = \varnothing$ and for every j either $T(h_j) = \varnothing$ or $[T(h_j) \neq \varnothing$ and there exists $y \in \{0,1\}$ s.t. $h_{j+1} = (h_j,(T(h_j),y))]$.

However, we do not yet know how locally optimal trees compare to each other, and how they compare to the "best" (i.e., globally optimal) trees. This will be the topic of the later section on globally optimal trees.

It is intuitive that there is no gain in specifying an attribute in the learning phase that has been previously contradicted, as the value of that attribute has been determined as irrelevant to the phenomenon's occurrence. This can be easily verified for the one round horizon case using theorem 1 and the definition of EV(•). We therefore state without proof the following lemma:

LEMMA 1: $EV(h \cup a_i) = EV(h)$ if $i \in Z(h)$.

Next, denote by M(h), the set of indices of those attributes that are locally optimal after history h. Formally define $M(h) := \{i \text{ s.t. } h \cup a_i \in \Gamma^L(r,k+1,h)\}$ where $h \in H_k$.

We present a theorem that characterizes the set of locally optimal attributes M(h) given the history h:

For any $B \in F$, $i \in \{1,..,r\}$ and $y \in \{0,1,*\}$, define $B \oplus (i,y) \in F$ by $[B \oplus (i,y)](i) := y$ and $[B \oplus (i,y)](j) := B(j)$ for all $j \neq i$. So $B \oplus (i,y)$ is derived from B by substituting y for B(i) in the i-th component of B. As a special case, $A(h) \oplus (i,y)$ is well defined for any history $h \in H$ since $A(h) \in F$.

THEOREM 2:

For any history $h \in H$:

$$M(h) = \begin{cases} \{1,..,r\} \setminus S(h) & \text{if } S(h) \neq \{1,..,r\} \\ \underset{i \notin Z(h)}{\operatorname{argmax}} \; P(X \in A(h) \oplus (i, v_h(i))) \sum_{J(i)=*} P(X(i)=v_h(i) \mid J) P(h \mid J) P(Q=J) & \text{if } S(h)=\{1,..,r\} \text{ and } Z(h) \neq \{1,..,r\} \\ \varnothing & \text{if } Z(h)=\{1,..,r\} \end{cases}$$

where, as above, $v_h(i)$ denotes the value of a_i that did not appear in h.

Notice that $M(h)=\varnothing$ means that the path stops after h, so in a locally optimal tree of arbitrary depth (i.e. $T \in \Gamma^L(r,\bullet)$), a path stops if and only if all attributes have been contradicted (i.e. $Z(h)=\{1,..,r\}$). Thus there is always something to learn unless we determined that the phenomenon always occurs.

PROOF:

W.l.o.g. assume that if $i \notin Z(h)$ then $a_i=0$ was observed. Then

$$EV(h \cup a_i) = \sum_{J \in F} P(A(h)) \; P(X(i)=0 \mid J) \; P(h \mid J) \; P(Q=J) \; +$$

$$\sum_{J \in F} P(A(h)(i,*)) \; P(X(i)=1 \mid J) \; P(h \mid J) \; P(Q=J) \; =$$

$$= EV(h) + \sum_{J \in F} [P(A(h)(i,*)) - P(A(h) \cap J)] \; P(X(i)=1 \mid J) \; P(h \mid J) \; P(Q=J) \; =$$

$$= EV(h) + \sum_{J \in F} P(A(h)(i,1)) \; P(X(i)=1 \mid J) \; P(h \mid J) \; P(Q=J)$$

We used the fact that $P(h \mid J)>0$ implies $A(h) \subseteq J$. This completes the proof.

****

Next we will analyze how well locally optimal algorithms perform in the limit, i.e., when the learning phase lasts sufficiently long.

We call a tree $T \in \Gamma(r,\bullet)$ "persistent" if each attribute appears infinitely often on every path unless it is contradicted. Formally, a tree is called persistent if for any path $h = (h_i)_{i \in \mathbb{N}} \in H(T)$, any attribute $j \in \{1,..,r\}$ and any $m \in \mathbb{N}$ either $j \in Z(h_m)$ or there exists $m' > m$ s.t. $T(h_{m'}) = j$.

The following lemma states that in a persistent tree, the number of times an attribute is specified is uniformly unbounded on all paths on which it has not been contradicted.

LEMMA 2:

If $T \in \Gamma(r,\bullet)$ is persistent, then for any $m \in \mathbb{N}$ there exists $k(m) \in \mathbb{N}$ s.t. for any $j \in \{1,..,r\}$ and $h = (h_i)_{i \in \mathbb{N}} \in H(T)$, either $j \in Z(h_{k(m)})$ or $| \{h_i \text{ s.t. } T(h_i) = j,\ i < k(m)\} | \geq m$.

PROOF:

I apologize for only sketching the proof since notations are cumbersome.
Since T is persistent, each attribute appears infinitely often on every path unless it is contradicted. The idea of the proof is the following. There is a finite number of paths without contradiction (2'), so given m there is a round k such that each attribute occurs at least m times on these paths before round k. Now there are only finitely many paths with exactly one contradiction before round k, so there is a number $k_1(m)$ so that the lemma holds for all paths with zero or one contradiction. Continue this argument and the lemma is proven.

****

Now we can state a result about the class of persistent trees. Take any algorithm that uses the specification rules given by a persistent tree in the learning phase and uses the safe reaction function A(•) in the decision phase. The probability that this algorithm fails to recognize a good decision opportunity and waits for a "better" one, tends to zero

as the length n of the learning phase goes to infinity. This follows directly from the fact that since the tree is persistent, the probability that the concept is not A(h) tends to zero as n gets large.

This result can be directly applied to locally optimal trees since they are persistent:

THEOREM 3:

Any locally optimal tree $T \in \Gamma^L(r,\bullet)$ is persistent.

PROOF:

For every attribute $i \notin Z(h)$, let G(h,i) denote the incremental benefit of specifying the attribute a_i after observing h. Formally,

$$G(h,i) := EV(h \cup a_i) - EV(h) = \sum_{J(i)=*} P(A(h)(i,v_h(i))) \; P(X(i)=v_h(i) \mid J) \; P(h \mid J) \; P(Q=J)$$

(see proof of theorem 2).

G(h,i) depends on the probability that a contradiction is achieved after specifying a_i, on the fact that this contradiction is useful for the decision phase and on the probability that the history h occurs.

Fix a path and consider the choices for attributes to be specified after the last contradiction. Notice that the attributes contradicted only appear a fixed number of times. W.l.o.g. assume that each of $a_1,..a_j$ appear a bounded number of times, and that $a_{j+1},..,a_k$ are specified infinitely often, $s_h(i)=0$ for $j+1 \le i \le k$ and $a_{k+1},..,a_r$ are contradicted on the path. Then G(h,i) is bounded below, for all $1 \le i \le j$, since the phenomenon might be determined by $B=(*,...,*,\underset{j+1}{0},...,\underset{k}{0},*,...,*)$ with positive probability because of the assumption of full support. On the other hand, G(h,i), $j+1 \le i \le k$, goes to 0 as a_i is specified more often. It is easy to see that this leads to a contradiction and hence $a_1,..,a_j$ are not specified a bounded number of times.

****

Summarizing theorem 3 and the result on how well persistent trees learn, we can now make a statement as to how well locally optimal algorithms learn to make the decision independently of how "well" the individual "assessed" his priors X and Q:

Assume that nature samples the examples using X' and the phenomenon is fixed. Then any locally optimal algorithm based on the priors X and Q (with full support) makes an arbitrary small mistake of failing to recognize a good decision opportunity if the learning phase lasts long enough. So one could say that locally optimal algorithms can asymptotically learn Boolean conjunctions.

In fact it follows directly from the results without restricted perception (see e.g. Dietterich [1990]) that persistent trees using the safe decision can PAC-learn Boolean conjunctions. However a persistent tree can easily be constructed that specifies an attribute only in exponential rounds. Such a tree cannot learn in polynomial time - which is required for Valiant's definition of learning. For a persistent tree to learn Boolean conjunctions in Valiant's sense, it is sufficient to prove the following: for any attribute, the number of the round in which it is specified, is bounded by a polynomial function of the number of times the attribute has been specified up to then. The analysis of whether or not locally optimal algorithms satisfy this condition will be postponed to a later section.

5. GLOBALLY OPTIMAL TREES:

In this section we continue the analysis of the set of globally optimal trees, i.e., the subset of all learning procedures that maximize the probability of making a safe decision in each decision opportunity satisfying the non-wastefulness condition. It turns out that there is a very close connection between these globally optimal trees and the locally optimal trees discussed in the last section.

The first simple result concerns the criteria for when a path stops: in a globally optimal tree of depth n (i.e., $T \in \Gamma^*(r,n)$), a path stops before round n if and only if all attributes have been contradicted (i.e., $Z(h) = \{1,..,r\}$) on that path. This had been shown for locally optimal trees in theorem 2.

LEMMA 3:

If $T \in \Gamma^*(r,n)$ $(n \geq r)$, $h \in H(T) \backslash H_n$ then $T(h) = \varnothing \Leftrightarrow A(h) = \Omega \Leftrightarrow Z(h) = \{1,..,r\}$.

The proof follows from theorem 2 and the fact that if the tree cannot be improved globally, then it cannot be improved locally either. And if the tree cannot be improved locally, then A(h) is Ω, the decision will be made in the first decision opportunity and hence the evaluation $EV(\bullet)$ of the tree will not improve anymore on that path.

The following assumption on the distributions of X and Q will turn out to be essential for our analysis:

It is assumed that for any $x \in \Omega$, $P(X = x) = \prod_{i=1}^{r} P(X(i) = x(i))$ and for any $B \in F$,

$P(Q = B) = \prod_{i=1}^{r} P(Q(i) = B(i))$, where X(i), and Q(j) denote the marginal distributions of X on $\{0,1\}$ and Q on $\{0,1,*\}$.

This assumption will be referred to later as independence (of Q and X) over attributes.

The assumption of independence over attributes simplifies the calculation of the set of locally optimal attributes M(h) considerably, since the increment of the value EV(•) now becomes separable. The condition for local optimality can now be interpreted quite intuitively: the i-th attribute is locally optimal if it maximizes the probability of being contradicted (i.e., getting a "*") and of improving the time the decision is made in the decision phase by this contradiction. This result is stated formally in the next theorem.

Define a function obs(h) from $\{(a_i,y)$ s.t. $1 \le i \le r$, $y \in \{0,1\}\}$ to $\mathbb{N}$, counting how many times each attribute (including its value) was observed in h. Formally, $obs(h)((a_i,y)) := |\{h_0 \in H$ s.t. $h = (h_0,(a_i,y),h_1)$ for some $h_1 \in H\}|$, where $h \in H$, $i \in \{1,..,r\}$ and $y \in \{0,1\}$.

THEOREM 2':

If Q and X are independent over attributes then for every history $h \in H$, the set of locally optimal attributes M(h) is

$$M(h) = \begin{cases} \{1,..,r\} \setminus S(h) & \text{if } S(h) \neq \{1,..,r\} \\ \underset{i \notin Z(h)}{\operatorname{argmax}} \left\{ \dfrac{P(X(i)=v_h(i))^2 P(X(i)=s_h(i))^{t(h,i)-1} P(Q(i)=*)}{P(Q(i)=s_h(i)) + P(X(i)=s_h(i))^{t(h,i)} P(Q(i)=*)} \right\} & \text{if } S(h)=\{1,..,r\} \text{ and } Z(h) \neq \{1,..,r\} \\ \varnothing & \text{if } Z(h)=\{1,..,r\} \end{cases}$$

where $t(h,i) := obs(h)((a_i,s_h(i)))$, the number of times $a_i = s_h(i)$ is observed in h for $i \in S(h) \setminus Z(h)$.

Notice that the expression in the argmax is decreasing in t.

We first observe the following property of M(•) that follows immediately from theorem 2' and will be the basis for the results on the relationship between locally and globally optimal trees:

If the assumptions of theorem 2' hold then any locally optimal attribute that is not

specified, remains locally optimal in the next round. Additionally, no new attributes become locally optimal unless all locally optimal attributes of previous rounds have been specified. This is stated formally in the following lemma.

LEMMA 4:

Under the assumptions of theorem 2', for $i \in \{0,1\}$,

$$M(h,(a_p,i)) = \begin{cases} M(h) \setminus \{p\} & \text{if } p \in M(h) \\ M(h) & \text{if } p \notin M(h) \end{cases}$$

We will refer to this property as property (P).

PROOF of theorem 2':

We only need to prove the statement for $S(h) = \{1,..,r\}$ and $Z(h) \neq \{1,..,r\}$, since the rest has already been shown:

for any $h \in H$, compare specifying a_j to specifying a_i $(i,j \notin Z(h), i \neq j)$:

w.l.o.g. assume $s_h(i) = s_h(j) = 0$.

$$EV(h \cup a_i) = [P(X(i)=0)\ P(Q(i)=0) + P(X(i)=0)\ P(X(i)=0)^{t(i)+1}\ P(Q(i)=*) +$$
$$P(X(i)=1)\ P(X(i)=0)^{t(i)}\ P(Q(i)=*)\ ] \bullet EV(h)\ /$$
$$[P(X(i)=0)\ P(Q(i)=0) + P(X(i)=0)\ P(X(i)=0)^{t(i)} P(Q(i)=*)] =$$
$$= EV(h) + EV(h)\ [\ P(X(i)=0)^{t(i)+1}\ [P(X(i)=0)-1] + P(X(i)=1)\ P(X(i)=0)^{t(i)}]\ P(Q(i)=*)\ /$$
$$[P(X(i)=0)\ [P(Q(i)=0) + P(X(i)=0)^{t(i)} P(Q(i)=*)]] =$$
$$= EV(h) + EV(h) \bullet P(X(i)=1)^2\ P(X(i)=0)^{t(i)}\ P(Q(i)=*)\ /$$
$$[P(X(i)=0)\ [P(Q(i)=0) + P(X(i)=0)^{t(i)}\ P(Q(i)=*)]].$$

This completes the proof of theorem 2'.

****

The proof of theorem 2' gives us an additional result regarding how much the value of the tree EV(T) improves from round to round:

COROLLARY 2':

If $S(h) = \{1,..,r\}$ and $i \in \{1,..,r\} \setminus Z(h)$ then

$$EV(h \bigcup a_i) = \left[1 + \frac{P(X(i)=v_h(i))^2 P(X(i)=s_h(i))^{t(h,i)-1} P(Q(i)=*)}{P(Q(i)=s_h(i)) + P(X(i)=s_h(i))^{t(h,i)} P(Q(i)=*)} \right] EV(h)$$

Now we will turn to the task of characterizing the set of globally optimal algorithms. As mentioned in theorem 1, the optimal reaction function is uniquely defined by A(•), so we only need to analyze the globally optimal trees. It should also be mentioned here that the results will rely solely on the fact that M(•) (depending on EV) satisfies the property (P). So although in this paper EV(•) is defined as P(make the decision), the following results can easily be generalized for any function EV(•) as long as the associated set of locally optimal attributes M(•) satisfies property (P).

Notice that because $EV(T) = \sum_{h \in H(T)} EV(h)$, a tree is globally optimal if and only if it is globally optimal on any subtree, formally $T \in \Gamma^*(r,n)$ if and only if for every $h \in H(T)$: $T/h \in \Gamma^*(r,n,h)$. So the attributes that are specified in the future (in an optimal algorithm) do not depend on the histories that did not occur (or equivalently only depend on the particular history that did occur).

Next, we need to formally describe the notion of "two trees have the same paths up to the order in which the observations were made". Let O be the set of all functions from $\{(a_i,y)$ s.t. $1 \le i \le r$, $y \in \{0,1\}\}$ to $\mathbb{N}$. Define a <u>profile</u> of a tree $T \in \Gamma(r,n)$ to be a function pro(T) from O to $\mathbb{N}$, counting how many paths have the same sequence of observations (independently of the order in which they were made), formally, $pro(T)(u) := |\{h \in H(T)$ and $[h \in H_n$ or $T(h) = \varnothing]$ s.t. $obs(h) = u\}|$ for $u \in O$. In the sequel we will use the equivalence relation induced by the function pro.

Notice that $EV(T) = EV(T')$ if $pro(T) = pro(T')$ since EV(h) only depends on the

attributes specified and the values observed (i.e. obs(h)), not on the order of specification.

Now we come to the main theorem of the paper. It states that every globally optimal algorithm is a locally optimal one up to the order in which the attributes are specified in the learning phase. Note that this order does not affect the value EV(•). So the locally optimal trees together with the optimal reaction function A(•) are sufficient for characterizing the globally optimal algorithms.

THEOREM 4:

A feasible tree (i.e. $T \in \Gamma(r,n)$) is globally optimal (i.e. $T \in \Gamma^*(r,n)$) if and only if it is a reordered locally optimal tree, i.e. if there exists a locally optimal tree $T' \in \Gamma^L(r,n)$ s.t. $pro(T) = pro(T')$.

As a corollary we immediately obtain that every locally optimal algorithm is also globally optimal and in particular the individual is indifferent between any two locally optimal algorithms. Formally,

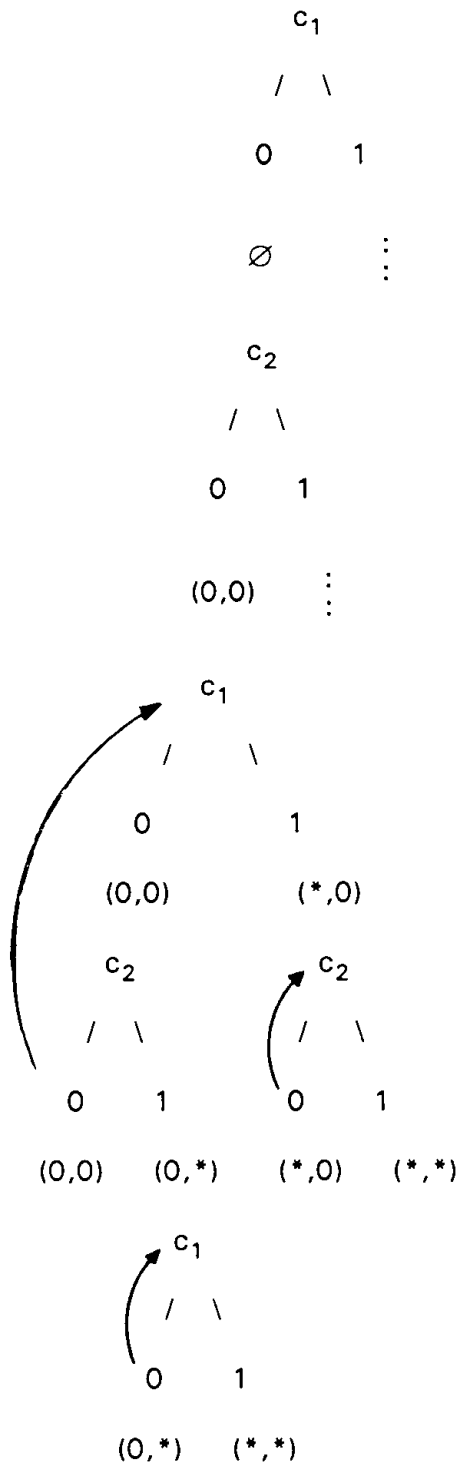
COROLLARY 4:

$\Gamma^L(r,n) \subseteq \Gamma^*(r,n)$

As a consequence, globally optimal trees can be easily and intuitively constructed in a way that is independent of the length of the learning phase: simply make the optimal one stage decisions following the r initial searches.

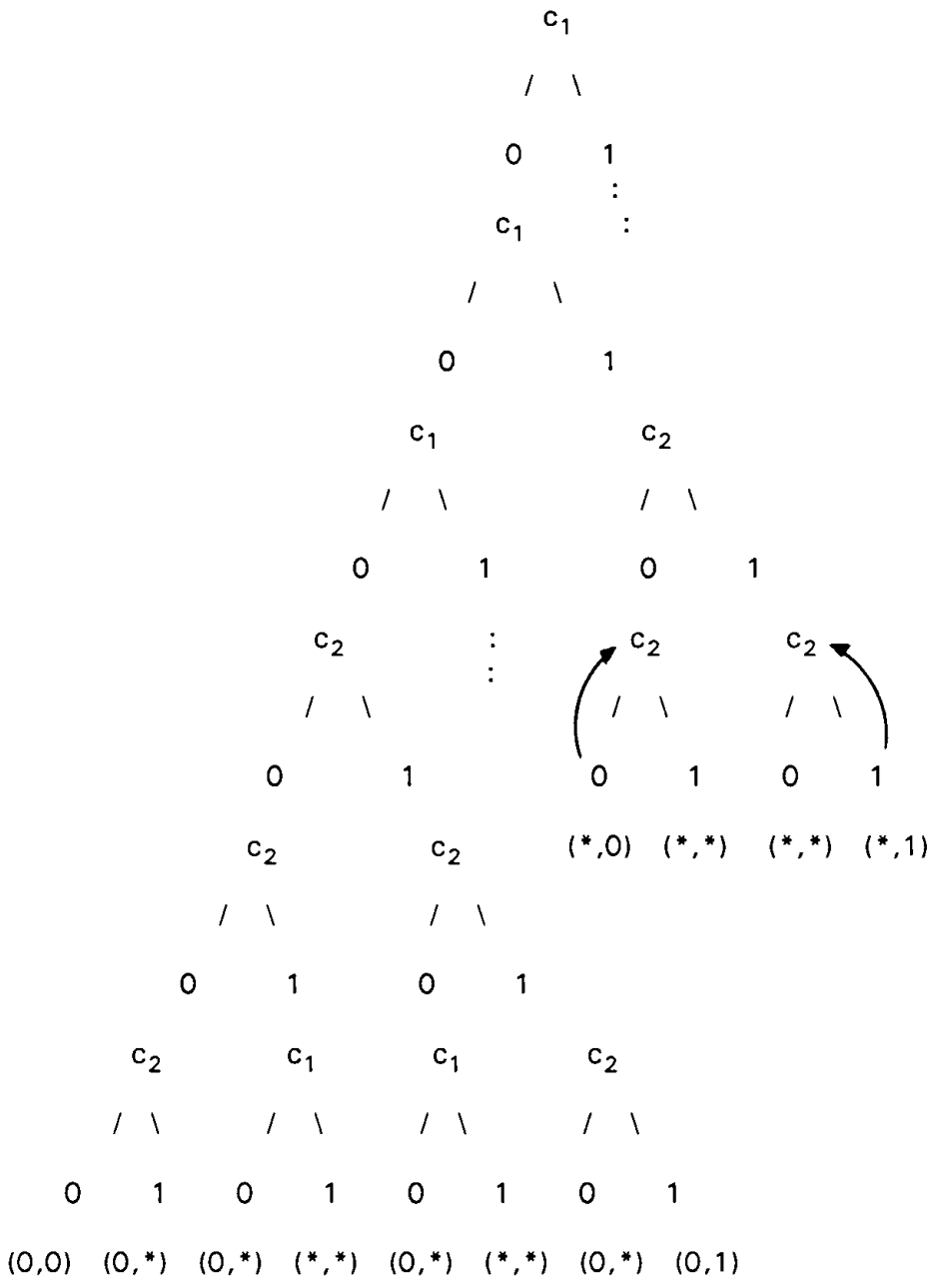
EXAMPLES: $r = 2$, fix $c \in (0,1/2)$, $P(Q(i) = k) := c$, $P(X(i) = k) := 1/2$, $i = 1,2$, $k = 0,1$

(1) This is the graphic illustration of a globally optimal algorithm that is locally optimal:



Notation: The algorithm follows the arrows as long as the number of observations is $\leq n$, then it stops with the reaction function written at that node as the output.

(2) An example of a globally optimal algorithm that is not locally optimal, $n=6$:



PROOF of theorem 4:

We will prove by induction on k that after any round k and any history $h \in H_k$ every globally optimal tree is a reordered locally optimal tree and that all locally optimal trees have the same value EV(•).

First consider $k = n-1$:

In the last round n, the attribute will be chosen in the locally optimal way because the definitions of locally and globally coincide in this case. So given $h \in H_{n-1}$, $\Gamma^L(r,n,h) = \Gamma^*(r,n,h)$.

Next assume correctness for $l > k$ and consider k:
For every $h \in H_k$ and $T \in \Gamma^*(r,n,h)$:

case #1: all attributes are contradicted in h, i.e. $Z(h) = \{1,..,r\}$.
Then $A(h) = \Omega$ and the path stops for locally and globally optimal trees (by theorem 2 and lemma 3), i.e. $M(h) = \varnothing$ and $T(h) = \varnothing$ so $\Gamma^L(r,n,h) = \Gamma^*(r,n,h)$.

case #2: $Z(h) \neq \{1,..,r\}$

First show that T is just a reordered locally optimal tree.

By lemma 3, the path does not stop after h. Let a_q be the attribute specified in round $k+1$, i.e., $q := T(h)$ and let T_i be the subtree of T with initial history $(h,(a_q,i))$, i.e. $T_i := T/(h,(a_q,i))$ for $i = 0,1$.

Then the by induction hypothesis, T_i is a reordered locally optimal tree, i.e., there are trees $T'_i \in \Gamma^L(r,n,(h,(a_q,i)))$ s.t. $pro(T'_i) = pro(T_i)$, $i = 0,1$, so w.l.o.g. assume that $T_i \in \Gamma^L(r,n,(h,(a_q,i)))$ for $i = 0,1$.

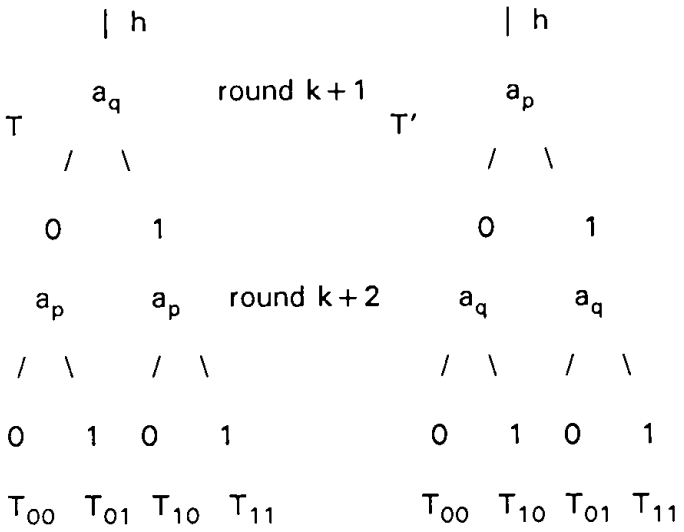
If the attribute specified in round $k+1$ is locally optimal, i.e., $q \in M(h)$ then T is locally optimal, i.e. $T \in \Gamma^L(r,n,h)$.

So assume not, i.e. $q \notin M(h)$:

The tree T does not stop after round $k+1$ following history h since there are still

attributes to be contradicted. So let a_p denote the attribute specified after observing history $(h,(a_q,0))$. By the assumption made above for T_i, $i=0,1$, the attributes specified in round $k+2$ are locally optimal, so in particular $p \in M(h,(a_q,0))$. By property (P) and the fact that $q \notin M(h)$, we obtain $M(h,(a_q,0)) = M(h) = M(h,(a_q,1))$.

If a_p is specified in both rounds $k+2$ of T then rounds $k+1$ and $k+2$ can be switched.



For the resulting tree T' the following holds: it is a reordering of T and hence has the same value $(EV(T) = EV(T'))$, it has a locally optimal attribute specified in round $k+1$ since $p \in M(h)$, and by the induction hypothesis the subtrees with initial history $(h,(a_p,i))$ are reordered locally optimal trees given $(h,(a_p,i))$, $i=0,1$. Therefore T' is a reordered locally optimal tree and hence so is T.

Next consider the case in which a_p is not specified in round $k+2$ of T_1

First we will show that a_p is specified on every path in T after history $(h,(a_q,1))$ $(\equiv T_1)$.

For this we need to show that n is large enough so that each locally optimal attribute in M(h) can be specified before the learning phase is over. Exchange the subtree of T with initial history $(h,(a_q,1))$ for a locally optimal tree given history $(h,(a_q,1))$ with the attribute a_p specified in round $k+2$. By the induction hypothesis this does not change the value EV(•). Hence the resulting tree is globally optimal given history $(h,(a_q,1))$. Now

switch rounds $k+1$ and $k+2$ and we get a reordered locally optimal tree given history h utilizing the induction hypothesis and the fact that a_p is locally optimal after h. Additionally, the attribute a_q is specified in both rounds $k+2$. If we would reorder this tree to be locally optimal then a_q would be specified at least once on every path after round $k+\mid M(h)\mid$ since all locally optimal attributes must be specified before attribute a_q because $q \notin M(h)$.

Hence T_1 is a locally optimal tree of depth at least $k+\mid M(h)\mid +1$. With $p \in M((h,(aq,1)))$ and using property (P), it follows that a_p must be specified on every path in T_1.

Using this result, it can easily be seen that T_1 can be reordered so that a_p is specified in round $k+2$.

Now a_p is specified in both rounds $k+2$ of T given history h, so as explained above it follows that T is a reordered locally optimal tree.

Summary: we began with a globally optimal tree and reordered it until the attributes were specified locally optimal.

To show that every locally optimal tree given history $h \in H_k$ has the same value EV(•) for case #2, follow same construction as above, only this time comparing two different locally optimal trees. The result follows from similar arguments.

This ends the induction step and therefore the proof of theorem 4 is completed.

****

Theorem 4 gives us the connection between locally optimal and globally optimal algorithms. However, if we have to determine whether an arbitrary tree is globally optimal or not, it might be quite difficult to use theorem 4 by considering all possible reordering of the tree. The following theorem gives a criterion for a tree being optimal without having to reorder it. It states that a feasible tree is globally optimal if and only if on each path each attribute is specified, the number of the attributes that are not contradicted on a path is determined independently of their order (according to local optimality), and any path stops after all attributes have been contradicted.

THEOREM 5:

Any feasible tree $T \in \Gamma(r,n)$ $(n \geq r)$ is globally optimal (i.e., $T \in \Gamma^*(r,n)$) if and only if for every history $h \in H(T)$

- $S(h) = \{1,..,r\}$,
- if $i,j \in S(h) \backslash Z(h)$ and $t(h,i) \geq 2$ then $G(j,t(h,j)) \leq G(i,t(h,i)-1)$ and
- $Z(h) = \{1,..,r\}$ if and only if $T(h) = \varnothing$.

where $t(h,i) := obs(h)((a_i, s_h(i)))$, the number of times $a_i = s_h(i)$ is observed in h for $i \in S(h) \backslash Z(h)$ and

$$G(i,t) := \frac{P(X(i)=v_h(i))^2 P(X(i)=s_h(i))^{t-1} P(Q(i)=*)}{P(Q(i)=s_h(i)) + P(X(i)=s_h(i))^t P(Q(i)=*)}$$ is the increment of the value function

$EV(\bullet)$ when the $(t+1)$-st a_i is specified.

This theorem points out that the number of attributes that are not contradicted, not the order in which they appear, is relevant for global optimality. In the set of globally optimal trees, the locally optimal trees have two great advantages:

- theorem 2' gives us a way to construct them,
- their construction is independent of n.

Note that in the case of $n < r$, every tree has the same value 0 (see theorem 1).

PROOF:

"only if":

Let T be a globally optimal tree. By theorem 4 we may assume w.l.o.g. that T is locally optimal. Assume further that $h \in H(T)$, $i,j \in \{1,..,r\} \backslash Z(h)$, $t(h,i) \geq 2$ and $G(j,t(h,j)) > G(i,t(h,i)-1)$ holds. This however means that a_j is locally "better" than a_i when a_i is selected the last time in h (need here that the increment G(j,t) is decreasing in t). This contradicts the fact that T is locally optimal.

The rest of the statement follows directly from theorem 4.

"if":

This case follows directly from the proof of theorem 4. The only part where the conditions of the theorem enter is in the initial phase of the induction, the proof for paths of length n-1. In this case the conditions stated in the above theorem are identical to the condition of local optimality. The rest of the proof involves only switching arguments that work similarly for this proof.

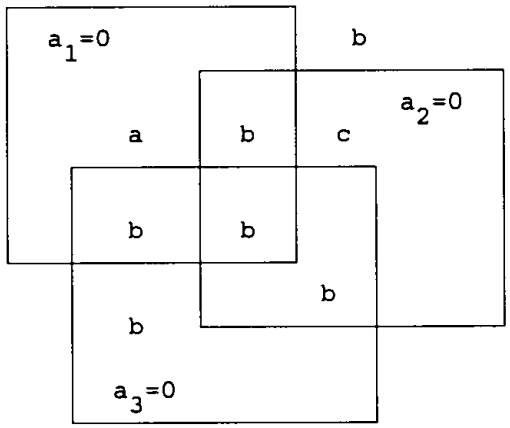
****

6. COUNTEREXAMPLE to theorem 5 when X is dependent over attributes:

The assumption that Q and X are independent over attributes is essential for theorems 4 and 5 to hold; here is a counter-example in which a globally optimal algorithm fails to be locally optimal in the absence of this assumption.

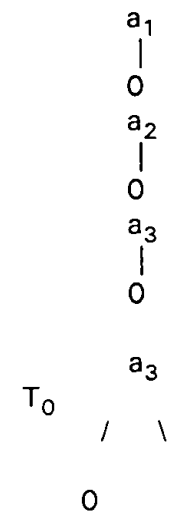
$r=3$, $n=5$, Q is uniformly distributed, i.e, for every $B \in F$: $P(Q=B)=1/27$, $P(X=(0,1,1))=0.1=:a$, $P(X=(1,0,1))=0.06=:c$, and $P(X=x)=0.14=:b$ for $x \notin \{(0,1,1),(1,0,1)\}$.

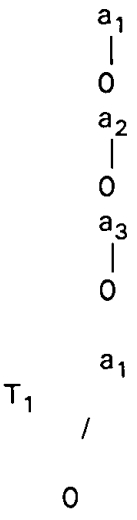
Diagram of the probabilities:



T_i, $i=0,1,2,3$ defined as in the graphs below
T_0 locally and globally optimal tree of length 4.
T_1, T_2 and T_3 are locally optimal in round 5.
Only T_3 is locally optimal in every round.
$\Gamma^*(3,5,h) = T_1$ but $1 \notin M(h)$ so $T_1 \notin \Gamma^L(3,5,h)$,
where $h=((a_1,0),(a_2,0),(a_3,0))$.

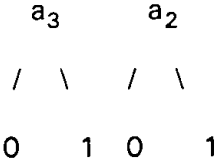
Diagram of the trees:





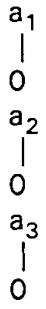
round 4

round 5

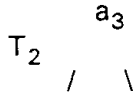


Unique globally optimal tree for n = 4 (given h)

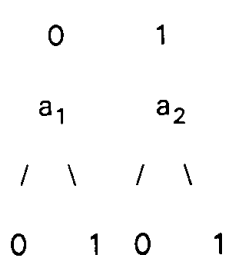
Unique globally optimal tree for n = 5 (given h)



unique locally optimal tree for n = 5 (given h)



round 4



round 5

7. ASYMPTOTIC BEHAVIOR of locally optimal trees:

Despite the nice properties of locally optimal trees, the individual still must make calculations before every round to find out which is a locally best attribute. What happens when the learning phase lasts a long time, i.e., when the locally optimal trees get large? Will patterns for selecting the attributes emerge? What kind of calculations are needed?

To achieve some regularity in constructing the locally optimal tree we need some assumptions about which attribute is chosen if the set of locally optimal attributes M(h) is not a singleton: under all attributes in M(h) we will choose the one with the lowest index.

We will analyze the locally optimal tree defined on all possible lengths of histories that is constructed using the above tie breaking rule.

Define $T_L(r) \in \Gamma^L(r,\bullet)$ s.t. for every $h \in H(T_L(r))$,

$$T_L(r)(h) = \begin{cases} \min\{1,..,r\} \setminus S(h) \text{ if } S(h) \neq \{1,..,r\} \\ \min M(h) \text{ if } EV(h \bigcup a_i) > EV(h) \text{ for some } i \\ \varnothing \text{ otherwise} \end{cases}$$

Now we come to a theorem that characterizes the asymptotic properties of the locally optimal tree $T_L(r)$. The theorem characterizes the relationship between two arbitrary attributes on any path on which they are not contradicted. The general pattern of all attributes can be easily deduced from this. The statement is that the number of times that an attribute a_j is specified between two attributes a_i "converges" (up to integrability problems) to a limit that is independent of the prior on the possible concepts Q.

Notation:

N:= set of strictly positive integers,

$\lfloor w \rfloor := \max \{z \text{ s.t. } z \in I, z \leq w\}$, and we get $\lfloor w \rfloor \leq w < \lfloor w \rfloor + 1$,

$\lceil w \rceil := \min \{z \text{ s.t. } z \in I, z \geq w\}$, and we get $w \leq \lceil w \rceil$.

THEOREM 6:

For every path in the locally optimal tree $T_L(r)$ and every $i,j \in \{1,..,r\}$ not contradicted on that path, let $p := \frac{\ln P(X(i)=s(i))}{\ln P(X(j)=s(j))}$ and N(t) denote the # of times a_j is specified between the t-th a_i and the $(t+1)$-st a_i $(t \geq 2)$ on the path, then

- there exists $t^\wedge$ s.t. for every $t \geq t^\wedge$, $N(t) \in \{ \lfloor p \rfloor , \lceil p \rceil \}$ and
- $\lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} N(k) = p$.

Thus, if $P(X(j)=s_h(j)) \geq P(X(i)=s_h(i))$ then eventually each a_i will be followed by approximately p specifications of a_j, where $p \geq 1$. This is quite intuitive since $P(X(j)=s_h(j))^p = P(X(i)=s_h(i))$, the probability of observing a contradiction in the decision phase after specifying the j-th attribute p times is the same as after specifying the i-th one once.

Notice also that if p is an integer, then N(t) is equal to p once t is large enough.

PROOF of theorem 6:

The idea behind the proof is quite simple. Consider the reciprocal of the increment G(i,t) of the value EV(•):

$$R(i,t) := \frac{1}{G(i,t)} = \frac{P(Q(i)=s_h(i)) + P(X(i)=s_h(i))^t P(Q(i)=*)}{P(X(i)=v_h(i))^2 P(X(i)=s_h(i))^{t-1} P(Q(i)=*)} \quad \text{for } t \geq 1.$$

For every history $h \in H$ s.t. $S(h) = \{1,..,r\}$ and $Z(h) \neq \{1,..,r\}$, $M(h) = \text{argmin} \{R(i,t(h,i))$ s.t. $i \notin Z(h)\}$ where t(h,i) denotes the # of times a_i specified in h.

R(i,t) can be written as $a_0 [(1/x)^{t-1} + a_1 x]$, where $x := P(X(i)=s_h(i))$, $a_0 := P(Q(i)=s_h(i))/[P(X(i)=v_h(i))^2 \ P(Q(i)=*)]$ and $a_1 := P(Q(i)=*)/P(Q(i)=s_h(i))$. We observe

that $x \in (0,1)$, $a_0, a_1 > 0$ and R(i,t) is increasing and unbounded in t. Similarly R(j,s) can be written as

$b_0 [(1/y)^{s-1} + b_1 y]$.

In the following we will compare the sequences $(R(i,t))_{i \in N}$ and $(R(j,s))_{s \in N}$. Write both sequences on the real line. Since $1/x > 1$ and $1/y > 1$, as s and t get large, we are essentially comparing the sequences $(a_0(1/x)^t)_{t \in N}$ and $(b_0(1/y)^s)_{s \in N}$. It can be shown that once t is large enough, there are either $\lfloor p \rfloor$ or $\lceil p \rceil$ and on average p elements of the second sequence separating two elements of the first, where $p := \ln(x)/\ln(y)$.

Finally we note that if $R(i,t-1) < R(j,s+1) < .. < R(j,s+k) < R(i,t)$ then a_j will be specified k times between the t-th a_i and the $(t+1)$-st a_i.

* * * *

Theorem 3 stated that locally optimal trees are persistent. Using theorem 6 we can now be more specific about the uniform bound on the specification of the attributes given in lemma 2 for locally optimal trees and denoted by k(•).

COROLLARY 6:

For any locally optimal tree T there exists $c \in \mathbb{N}$ s.t. $k(m) \leq c*m$, where $k(m) = \max\{\text{length}(h) \text{ s.t. } \text{obs}(h) \leq m \text{ and } h \in H(T)\}$

PROOF: follows directly from theorem 6.

The linear form of k(•) reflects the regularity of the specification of the attributes on every path. This corollary together with the results stated after theorem 3 now shows that locally optimal algorithms can learn Boolean conjunctions in Valiant's sense, i.e., probably approximately correct (PAC) in polynomial time.

8. DISCUSSION:

In this paper we consider an individual learning about a phenomenon from examples in prospect of coordinating a decision with its occurrence. To emphasize the aspect of human decision making, we assume that the individual either explicitly has bounded perception or implicitly has restricted perception because he obtains his information by making queries. This assumption of bounded perception has a major influence on the model and its analysis and thereby gives the model a unique flavor. Our model utilizes the basic framework of formal learning theory (Valiant [1984]) and applies it to a different field, namely human learning and decision making under restrictions on perception.

The current approaches in the concept learning literature assume that all attributes of an example are observed or that at least the relevant attributes are presented (Blum [1990]). They do not impose any bounds, partly because their research is devoted to programming machines to learn.

In the present paper, the individual must select which attributes he wants to observe because the total number of attributes he may observe per example is limited. Thus the individual is active in the process of gathering information which is modelled as part of the optimization.

Restricting the perception makes information harder to gather. For instance, limiting the observation to one attribute per example essentially forces the individual to gather his information only from positive examples. Also the set of possible causes of the phenomenon that the individual can distinguish is limited by the bounded perception. In the case of limiting the perception to one attribute per example as done in the paper, the set of causes that can be distinguished are essentially the set of causes that can be written as a Boolean conjunction (i.e., logically connected by "AND") of the values of the attributes.

One of the major issues in learning theory is how to appropriately define what it means to "learn" some information. Such a definition enables us to compare the effectiveness of different algorithms or to determine what can be learned. In our model, the learning is related to a decision. Given that we know when the individual will make his decision, the individual "learns" how to decide by analyzing the effect the information obtained has on the decision making. However, the resulting improvement of the decision

making is really an expected improvement because the impact of the information gathered is not apparent until after the decision has been made and the outcome of the phenomenon is revealed. Additionally we need to make assumptions and analyze when the individual makes the decision. So a central issue is to determine how prior knowledge and hunches (or beliefs) influence when the decision is made and how they are used to calculate the expected effect of information gathered on the decision.

In the following I will relate the approach taken in this paper to two major frameworks, Valiant's formal model of concept learning (see Valiant [1984]) and subjective expected utility theory (Savage [1954]).

Learning concepts from examples is the basis for Valiant's framework, a widely adopted model in the machine learning literature. It can also be described as PAC-learning (from: probably approximately correct, due to Angluin and Laird [1988]) in polynomial time. The idea is that the learning algorithm must "learn" (with high probability) for arbitrary distributions that nature chooses from. Especially the machine does not have any priors. This reflects an attempt to learn with as few assumptions on prior knowledge as possible, incorporating knowledge into the model only through the bias (i.e., the set of possible concepts). The goal of the machine is to learn the concept in this abstract sense.

In our model, learning is associated with making a future decision. In order to apply Valiant's framework to our model, we must assume that the individual only wants to learn when the phenomenon occurs (independently of the decision). It turns out that Valiant's definition of learning is too general to generate conclusive results for "optimal" behavior in the learning phase: it does not lead to a specific rule dictating which attribute to specify next. It only requires that the algorithm must be persistent (each attribute is specified infinitely often unless contradicted) and the number of the round in which an attribute is specified must be bounded by a polynomial function of the number of times the attribute has been specified up to then. Essentially the PAC-model applies a worst case analysis, collecting data like a machine. In a context without bounded perception, this has been criticized by Buntine [1989] (see also Amsterdam [1988]) who suggests to use the classical model with priors (see subjective expected utility theory developed by Savage [1954]).

However, when we apply the notion of PAC-learning to the behavior in the decision

phase (unlike the learning phase), we get a well specified result, namely the condition of making a safe decision (compare to Shvaytser [1988]).

The second framework related to the model is subjective expected utility theory, developed by Savage [1954]. This is a general framework for making optimal decisions and learning through updating of prior beliefs. It is intuitive to assume that the individual uses his knowledge or hunches when he determines which attribute to specify next in the learning phase - and the notion of subjective priors provides a means to incorporate the individual's initial knowledge. However, it also implies that he makes his decision maximizing his utility with respect to his priors. These priors are more than just hunches: it is implicitly assumed by the definition of optimality that nature randomly selects states using the priors (just like the game is set up in this paper). In other words, the optimality is merely subjective, due to the subjectivity of the prior. In our model, such priors would be based on insufficient experience (unknown environment). Additionally, we are concerned with optimal behavior for a finite horizon (In the uninteresting case of the limit, a very broad class of algorithms can learn when the phenomenon occurs). So the outcome of the decision would be very sensitive to the specification of the priors if we were to use expected utility theory when making the decision. Because the decision is critical, the individual is uncertainty averse when it comes to the possibility of getting the bad outcome of making the decision when the phenomenon does not occur. So the individual wants to avoid this sensitivity due to the insufficient experience. Subjective expected utility theory has the disadvantage of being too specified when the agent (or individual) has insufficient experience because the results rely substantially on the priors and does not incorporate such uncertainty aversion. Such criticism seems to be growing and alternative models have been developed (see Bewley [1986], Schmeidler [1989] and others).

Prior knowledge is not needed in the decision phase if, as assumed, the number of rounds in the learning phase is greater than the number of attributes. Decision opportunities in which the phenomenon occurs with certainty, can be found using the observations in the learning phase, disregarding the priors (provided that each attribute was observed). With this in mind, it can be argued that the individual only makes safe decisions. However, conditioned on making a safe decision, priors can be incorporated in the learning phase to reflect the initial hunches. Therefore we assume that the individual

maximizes his expected payoff given the safe decision. The individual incorporates his (limited) knowledge where it is needed, but does not rely on it when making the critical decision.

Of course, subjective expected utility theory can be directly applied to our model to get the same conditions as required in this paper: assume a cost of infinity for making a mistake. However, the author does not think that this correctly reflects the intuition behind choosing to avoid the bad outcome because the environment is unknown. Two similar approaches to ours that also try to avoid the bad outcome, are the type I/type II error approach in statistics and Gilboa's [1988] framework explaining the Allais Paradox.

Note that both of the priors X and Q do not need to be based on insufficient experience to fit in the setup of this paper: the decision problem might also be in a partially unknown environment where the individual is not aware of the existence of the phenomenon until right before the beginning of the game. Due to previous experience, the individual has a well defined prior X on the occurrence of the items. The prior Q on the possible concept is established when he finds out about the existence of the phenomenon. In this setup it is natural to assume that the prior X (established over time) and the prior Q are independent (since the phenomenon was not known when establishing X). The individual's aversion towards making a mistake (because of the lack of experience w.r.t. the phenomenon) again leads to making a safe decision.

We characterize the set of optimal algorithms and in so doing, find a class of optimal algorithms that additionally have appealing properties, namely that they are easy to implement, are constructed independently of the length of the learning phase and asymptotically follow a simple pattern. This class of algorithms, which we call locally optimal, are constructed as follows: for each learning example, observe the attribute that maximizes the expected improvement of the decision, disregarding possible effects this specification has on future rounds. After the learning phase is over, make the decision in the first decision opportunity in which all its attributes' values were observed in the learning phase.

As the length of the learning phase grows, the specification procedure approximates a fixed pattern that is quite simple to characterize: following a fixed order,

each attribute is specified a fixed number of times (dependent on the attribute) if it has not yet been contradicted. Otherwise it is not specified any more. Moreover, this order does not depend on the prior on the possible concepts.

In view of the emphasis on human decision making, the discovery of a class of optimal algorithms that is easy to implement is very appealing. Furthermore, the asymptotic pattern of specification gives a nice rule-of-thumb for learning in such a setup once the learning phase has lasted long enough. It especially points out that only at the beginning of the learning, hunches might cause a "concentrated" search. Eventually, each attribute will be specified evenly (according to its probability of occurrence).

Additionally, it is mentioned that the locally optimal algorithms are "optimal" w.r.t. Valiant's model too: locally optimal algorithms (using priors with full support) can learn Boolean conjunctions in his sense.

The seemingly related learning processes in the two-armed-bandit (TAB) problems (see DeGroot [1970]) turn out under more careful examination to have quite different assumptions. In the problem of the TAB, the actions in each round determines some payoff and hence experimenting can be costly. In models of learning in prospect of a future decision, the payoff is not determined until after the learning is over. In these models, there are decreasing returns to specifying an attribute repeatedly because the expected payoff depends on the knowledge of each attribute. This difference is also reflected in the results: in the TAB problem, after finitely many rounds, it is optimal for the individual to keep on using the same slot machine (compare to results in this paper).

Of course the presented model has very specific assumptions. However, it should be noted that the employed switching technique and consequent results derived in this paper can very easily be generalized to similar setups as long as property (P) holds. For instance, the results can easily be generalized to include attributes with finitely many values (instead of just obtaining the values 0 and 1). They also can be extended to setups where the probability of an item occurring is conditional on the concept (i.e., $P(X=x \mid Q=I)$). In this case the independence assumption for X must be replaced by

$$P(X=x \mid Q=l) = \prod_{i=1}^{r} P(X(i)=x(i) \mid Q(i)=l(i)) \text{ for any } x \in \Omega \text{ and } l \in F.$$

The results heavily depend on the independence assumption on the probability measures. An example is given that shows that the locally optimal algorithms are not necessarily (globally) optimal. However, as a first approach this assumption is quite straightforward and more importantly, very intuitive for many examples.

It is possible to view the model presented here as a symbiosis of Valiant's model and the standard Bayesian learning model to tackle a specific problem of human (in contrast to machine) learning with bounded perception in an unknown environment. As the counterpart to the issues of complexity studied in Valiant's model, we study finite horizon algorithms and implementability. Instead of making the decision that maximizes expected utility, we maximize expected utility subject to a condition equivalent to demanding that the conjectures on which the decision is based must be probably approximately correct.

Assuming certain independence assumptions, the result of the analysis is a characterization of the best algorithms for the individual to use and a rule of thumb for when the learning has already lasted a while.

REFERENCES to related papers:

Amsterdam Jonathan, "Some Philosophical Problems with Formal Learning Theory", AAAI-88, 580-584 (1988).

Angluin D., Laird P., "Learning from noisy examples", Machine Learning, 2(4): 343-370, 1988.

Bewley Truman, "Knightian Decision Theory: Part 1", Mimeo, Yale University (1986).

Blum Avrim, "Learning boolean functions in an infinite attribute space", 22 ACM Symp. on Theory of Comp., 64-72 (1990).

Buntine Wray, "Learning Classification Rules Using Bayes", 6th Int. Work. Mach. Learn., 94-98 (1989).

Buntine Wray, "A Critique of the Valiant Model", IJCAI-89, 837-842 (1989).

DeGroot Morris H., "Optimal Statistical Decisions", 394-405 (1970), Mc Graw-Hill Book Co., Inc., New York.

Dietterich Thomas G., "Machine Learning", Annu. Rev. Comput. Sci., 4:255-306, 1990.

Gilboa Itzhak, "A Combination of Expected Utility and Maxmin Decision Criteria", J. Math. Psych. 32 (4): 405-420 (1988).

Miller George, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information", Psych. Rev. 63, 81-97 (1956).

Savage Leonard J., "The Foundations of Statistics", John Wiley and Sons (1954).

Schmeidler David, "Subjective Probability and Expected Utility Without Additivity", Econometerica 57(3), 571-587 (1989).

Shvaytser Haim, "Representing Knowledge in Learning Systems by Pseudo Boolean Functions", Proc. of the Sec. Conf. on Theor. Aspects of Reasoning about Knowledge, 245-259 (1988).

Valiant Leslie G., "A Theory of the Learnable", Proc. 16th Ann. ACM Symp. on Theory of Comp., 436-445 (1984).