

**N89 - 20068**

**AN EVIDENTIAL APPROACH TO PROBLEM SOLVING WHEN A LARGE NUMBER OF  
KNOWLEDGE SYSTEMS IS AVAILABLE.**

**FINAL REPORT**

**NASA/ASEE Summer Faculty Fellowship Program -- 1988  
Johnson Space Center**

<b>Prepared By:</b>	<b>Andre de Korvin, Ph.D.</b>
<b>Academic Rank:</b>	<b>Professor</b>
<b>University &amp; Department:</b>	<b>University of Houston - Downtown Applied Mathematical Sciences Houston, Tx. 77002</b>
<b>NASA / JSC</b>	
<b>Directorate:</b>	<b>Mission Support</b>
<b>Division:</b>	<b>Mission Planning and Analysis</b>
<b>Branch:</b>	<b>Technology Development and Applications</b>
<b>JSC Colleague:</b>	<b>Robert Lea, Ph.D.</b>
<b>Date Submitted:</b>	<b>August 12, 1988</b>
<b>Contract Number:</b>	<b>NGT 44-005-803</b>

## ABSTRACT

Some recent problems are no longer formulated in terms of imprecise facts, missing data or inadequate measuring devices. Instead, questions pertaining to knowledge and information itself arise and can be phrased independently of any particular area of knowledge. The problem considered in the present work is how to model a problem solver that is trying to find the answer to some query. The problem solver has access to a large number of knowledge systems that specialize in diverse features. In this context, feature means an indicator of what the possibilities for the answer are. The knowledge systems should not be accessed more than once, in order to have truly independent sources of information. Moreover, these systems are allowed to run in parallel. Since access might be expensive, it is necessary to construct a management policy for accessing these knowledge systems. To help in the access policy, some control knowledge systems are available. Control knowledge systems have knowledge about the performance parameters status of the knowledge systems. In order to carry out the double goal of estimating what units to access and to answer the given query, diverse pieces of evidence must be fused. We use the Dempster-Shafer Theory of Evidence to pool the knowledge bases. The present work demonstrates how pooling evidence can determine the necessary strategy to solve the problem as well as solving it. The present work deals with the path the problem solver could follow to answer the query.

## INTRODUCTION

Some recent problems are no longer formulated in terms of imprecise facts, missing data or inadequate measuring devices. Many problems today deal with knowledge and information itself and can be phrased independently of any particular area of knowledge. Problems in such diverse fields as psychology, engineering, artificial intelligence and decision making can exhibit similar formats when it comes to dealing with information and uncertainty. It is desirable to be able to handle information, not in any particular context, but in the form of general principles valid in many situations.

In [1][2] and [8], the authors develop a model for a course of action-feedback loop that is useful when considering learning in repeated environments. This model is not applicable to the case where one-shot decisions are to be made. For such decisions, the decision maker accumulates information and after enough accumulation has taken place, certain alternatives are ruled out. This process is repeated till there is only one remaining alternative. Presumably this last alternative represents the decision to be taken. This model is studied in [9]. In this setting, it is essential to be able to update the general state of information every time an alternative is ruled out.

A powerful method in artificial intelligence is to look at certain features of a problem and try to put together the evidence coming from these features in order to recognize the corresponding pattern. In this context, feature means any indication of what the possibilities of a correct answer are. Special cases of such an approach are, for example, taken in computer vision, see [3], [13], and [14]. It is clear that with such an approach we need to have some kind of mechanism available for evidence pooling. We shall use the D-S theory [12] to carry out this pooling of evidence. In fact, the D-S theory will be used for two purposes: (a) plan the solution to a query by accessing some pertinent knowledge systems (b) define dynamically a policy that would determine what *K*Ss to access at a particular time. The problem we consider here is quite general. Given a query *Q* we look at a set of possible answers to *Q*. We of course would like to determine the precise element of the set that represents the correct answer to *Q*. In order to accomplish this, specialized knowledge systems are consulted. The execution is driven by a table that to each goal associates a sequence of control strategies. A control strategy is a set of simultaneous constraints on the desired performance characteristics of the knowledge systems to be consulted. A number of control strategies weighted by the strength of belief in each constraint is given in the table. The weight of each constraint is the degree of belief that a specific value, for some performance parameter, is optimal. This will be compounded with the uncertainty that some family of knowledge systems has the desired performance parameter value.

After the completion of each goal, new data will be read in. A control sequence defines a subcycle of the goal execution. After the completion of each subcycle, an update on the information status is performed. In order to obtain information about the knowledge systems, special knowledge systems, called control knowledge systems, are consulted. Based on the pooled evidence provided dynamically by the control knowledge systems, a policy of access is determined.

There are two common ways of representing uncertain information about propositions. The most straightforward (employed in PROSPECTOR) provides a number  $P(A|E)$  which is the probability of proposition A being true, given current evidence E. One problem with this approach is that the precision of  $P(A|E)$  is not known. In fact, Garvey et al note "A likelihood represented by a point probability value is usually an overstatement of what is actually known." Another problem with this approach is that we do not have the amount of evidence for and against A. The other way to proceed is to attach not one but two numbers with each proposition (MYCIN employs this method). The two numbers correspond to measures of belief and disbelief. This approach certainly removes one of the above criticisms. It is still true that the precision relative to the two numbers is not known. Another criticism that may be leveled is that no formal theory indicates how evidence for and against might be combined. Also, it is not clear how to detect conflict. Finally, it has been shown that results degrade quite fast as uncertainty increases. The D-S theory is used here because it avoids, for the most part, the problems outlined above.

We assume that a large number of knowledge sources is available because we do want to access independent sources and therefore do not want to access the same source more than once. In this context, "independent sources" means that the errors in these sources are independent. To approximate independence, we may force a delay before any knowledge source is reaccessed. When all the steps have been taken, a final evaluation is made to determine the answer to the query.

## CONCEPTS AND NOTATIONS

We start by defining the concept of belief structure. For a complete account of general belief structures and related material, the reader is referred to [12]. If  $X$  denotes some set and  $\tau(X)$  denotes the power set of  $X$ , a belief structure is a function  $m$  from  $\tau(X)$  into  $[0,1]$  satisfying

$$(i) \quad m(\emptyset) = 0$$

$$(ii) \quad \sum_{A \subset X} m(A) = 1$$

Elements  $A \in \tau(X)$  for which  $m(A) > 0$  are called focal elements of  $m$ .

Now we define the Belief and Plausability measures on  $\tau(X)$

If  $B \in \tau(X)$

$$Bel(B) = \sum_{A \subset B} m(A); \quad Pls(B) = \sum_{A \cap B \neq \emptyset} m(A).$$

Suppose we have  $n$  experts whose opinions are equally respected. Assume that the opinion of the  $i^{th}$  expert is that a specified object is to be found in some set  $A_i$ . Given these  $n$  opinions we now ask the question: what is the probability that this specified object is in some set  $B$ ? Assume for ease of notation that  $A_1, A_2, \dots, A_K$  are subsets of  $B$ ,  $A_{K+1}, A_{K+2}, \dots, A_{K+\ell}$  intersect  $B$  and its complement and that  $A_{K+\ell+1}, \dots, A_n$  are subsets of the complement of  $B$ . A lower bound on the probability that the object is in  $B$  is  $K/n$ , an upper bound is  $(K+\ell)/n$ . We hence see that a possible interpretation of  $BEL(B)$  and  $Pls(B)$  is a "Lower probability and a higher probability" of  $B$ . Another way to interpret  $BEL$  and  $Pls$  is to identify the set  $B$  with the proposition: "The object is in the set  $B$ ." The truth or falsity of this proposition is determined by the available evidence, in this case, the opinion of the  $n$  experts. With this interpretation  $Bel(B)$  is interpreted to be the degree of support for proposition  $B$ .  $Pls(B)$  is the degree to which one fails to refute proposition  $B$  ( $A_1, A_2, \dots, A_{K+1}, \dots, A_{K+\ell}$  all intersect  $B$  and this constitute evidence that fails to refute  $B$ ). With this interpretation,  $Pls(\neg B)$  is interpreted as the degree to which proposition  $B$  is refuted while  $Pls(B) - Bel(B)$  is the degree of ignorance about proposition  $B$ , indeed:

$$Pls(B) - Bel(B) = \sum_{i=K+1}^{K+\ell} m(A_i)$$

and since the above  $A_i$  do intersect  $B$  and its complement, they fail to refute  $B$  and they fail to refute  $\neg B$ . Ignorance on proposition  $B$  is the degree to which one fails to refute  $B$  and  $\neg B$ . An important property of belief structures is that they allow to fuse evidence from independent sources of information. If  $m_1$  and  $m_2$  are belief structures defined on the same set  $X$  then the combination rule, see [12], states that they combine to yield a belief structure  $m_3$  on  $X$ . The focal element of  $m_3$  are given by taking all possible intersections of focal elements of  $m_1$  with focal elements of  $m_2$  and

$$m_3(C) = \sum_{A \cap B = C} m_1(A) m_2(B) / (1 - K)$$

$K$  is the measure of conflict between the two sources and comes from pairs of disjoint focal elements. In fact

$$K = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$$

If  $m_3$  is given by the formula above, we write  $m_3 = m_1 \oplus m_2$ . The formula above is true if the information sources are independent. For a very readable interpretation of the combination formula, the reader is referred to the article by L.A. Zadeh [16].

In [1], [2], and [8] the authors have defined the concept of an abstract information system. In this model, decisions are made sequentially and information is received as a consequence of the course of action taken. That is, each cycle of the decision-information acquisition loop is initiated by the decision maker ( $DM$ ) choosing a course of action ( $COA$ ) and the data generated by this  $COA$  is fed back to the  $DM$  which completes the cycle. This model is most relevant to situations requiring frequent decisions on a regular basis such as ordering equipment or choosing stocks. The  $DM$  can take the same  $COA$  on successive cycles or change to different  $COA$ , possibly returning to a  $COA$  chosen earlier. Uncertainty has been incorporated into this model. Different policies for selecting a  $COA$  on a given cycle have been discussed in [7]. The model discussed in the works cited above is not relevant to "one-shot decisions" where the  $DM$  uses all available information to pick a single  $COA$  which, once chosen, is irreversible.

In [9] the authors study situations where one-shot decisions are called for. In this context there may be many minor decisions, such as which information source to access and how to interpret information received. The  $DM$  constructs a set of possible  $COAs$  and then, based on the information received, eliminates the less promising  $COAs$  until only one remains. This, of course, is the  $COA$

taken. The main idea in [9] is that for each fixed data vector, the *DM* creates a conditional belief structure and then after  $K$  cycles, combines these masses conditioned by the data vectors into a single mass conditioned by history. From that, the belief and plausability measures conditioned on history are formed and whenever the plausability of a  $COA_i$  falls below the belief of some other  $COA_j$ ,  $COA_i$  is eliminated from consideration. The general problem of gathering evidence through classical statistics is that, as already pointed out in the introduction, an enormous amount of data is required. In addition, with such approaches, results quickly deteriorate if uncertainty is present. Usually some version of Baye's rule of inference is used as a remedy to the large amount of micro-events present, see for example [4] and [5]. In [10] J. Pearl makes the point that a complete probability space is not required. One needs only to estimate likelihood ratios. However a large number of such likelihood ratios would still be required.

In the next section we will outline an approach that uses the DS theory to collect evidence from different cycles as in [9]. We will collect evidence by analyzing different features of the problem. Breaking a problem down to subsets of features has already been tried in diverse fields of artificial intelligence. For example for computer vision, see [3], [13], and [14]. In most of the existing approaches, information on specific features is relatively static and no update of that information is incorporated into the DS approach. In this work, an update mechanism will be built into the decision making process. Also the *DS* theory will not only be used to identify a solution of the problem but also will determine how to access the knowledge sources (*K*Ss) Since access to some *K*Ss might be difficult and since we will at the very least delay reaccessing a *K*S, it is particularly important to determine a viable policy for access. We will also allow different *K*Ss to be multi-tasked on particular cycles or sub-cycles. Finally it is important to have a model that is potentially applicable to many situations and for this reason we have made the discussion as general as possible.

## RESULTS

The setting is as follows: Let  $Q$  denote some specific query and let  $H_Q$  be the corresponding frame of discernment, i.e.  $H_Q$  is a set of possible answers to  $Q$  and one of the elements of  $H_Q$  is the correct answer. Of course we do not know a priori which element is the correct answer. The data vector  $x$  is of the form

$$x = \left( f_1^{k_1}, f_2^{k_2}, \dots, f_s^{k_s} \right)$$

where  $s$  is large and where  $f_i^{ki} \in F_i (1 \leq i \leq s)$ .  $F_i$  is a set of (not necessarily numerical) values pertaining to the  $i^{th}$  feature. For example for some  $i$ ,  $F_i$  could be different values of light intensity. These values could range in the set {low, average, fairly high, high}. For another  $i$ ,  $F_i$  could be color values such as {red, blue, green} etc... Associated with each set  $F_i$  we have a knowledge system  $KS_i$  that only reports on values in  $F_i$ . Thus for some values of  $i$ ,  $KS_i$  may report only on the value of the color. Since we would like in principle, not to access any  $KS_i$  more than once (in order to truly have independent sources of information), we assume that  $s$  is large enough to yield appropriate information to obtain an answer to  $Q$ . Given a feature value  $f_i^{ki} \in F_i$  we define a corresponding subset of  $H_Q$  by setting

$$A(f_i^{ki}) = \left\{ g \in H_Q / i^{th} \text{ feature of } g \text{ has value } f_i^{ki} \right\}$$

$A(f_i^{ki})$  can be identified with the proposition: "The  $i^{th}$  feature has value  $f_i^{ki}$ ". In fact if  $G \subset H_Q$  then  $G$  can be identified with the proposition: "The answer is in  $G$ ". Thus  $\tau(H_Q)$  can be identified with the set of relevant propositions.

As already mentioned earlier, the value of the  $i^{th}$  feature will be reported by  $KS_i$ . However, there is some uncertainty built into the reporting of  $KS_i$ . Thus for the corresponding  $i$ ,  $KS_i$  might report that the intensity is low .6, is average .3 and is fairly high .1

In general, let  $a_i^k$  be the degree of belief that the  $i^{th}$  feature value is  $f_i^k$ . Clearly

$$\sum_k a_i^k = 1, 1 \leq i \leq s.$$

We now define a belief structure  $m_i$ :

$$m_i: \tau(H_Q) \rightarrow [0,1], m_i(A(f_i^k)) = a_i^k$$

Another way to state this is to say that  $a_i^k$  is the degree of belief that any element of  $A(f_i^k)$  is possibly the correct answer. Thus to sum up: in order to get the answer to  $Q$ , we look at the values of some features. These values are reported by specialized  $KS$ s with some degree of uncertainty built into the report. It is also clear that  $m_i, 1 \leq i \leq s$  constitute independent sources of information since they deal with distinct features.



Since there is a large number of *KSs* to access, it is important to formulate a policy on how to access *KSs*. We base this policy on performance characteristics. Let  $P_1, P_2, \dots, P_\ell$  be sets of performance values e.g.  $P_1$  could be a set of values corresponding to the cost of access,  $P_2$  could correspond to response delays etc... To each  $P_i$  corresponds a special *KS*, distinct from the ones mentioned above, which knows about performance (as opposed to features). To distinguish these *KSs*, we call them control *KSs* and denote them by *CKSs*. Thus  $CKS_i$  has knowledge of values in  $P_i$ . This knowledge is much more dynamic than the knowledge possessed by *KSs*, since performance of the *KSs* strongly depends on time. We would like to have some of the  $CKS_i$  run in parallel. To this end we set

$$H_p \subset \tau \{KS_1, \dots, KS_s\}$$

to be all possible combinations of *KSs* that may run in parallel. Some of the combinations are of course ruled out as some of the *KSs* might not be compatible or may have been already used. As earlier, for  $p_i^{ki} \in P_i$  define

$$A \left( p_i^{ki} \right) \in H_p$$

where each element of  $A(p_i^{ki})$  is such that its  $i$ th performance characteristic value is  $p_i^{ki}$ . For example if  $p_i^{ki}$  means the cheapest access and if  $A(p_i^{ki}) = \{KS_{12}, KS_{19}\}$  then  $KS_{12}$  or  $KS_{19}$  might be the cheapest *KS* to access.

In order to find an answer to  $Q$ , we will have a sequence of goals to satisfy. The process will be table-driven in the sense that each intermediate goal will correspond to a sequence of control strategies. A control strategy will be a set of performance constraints to be simultaneously satisfied. Ultimate control strategies with their respective degrees of belief may be given. For example a control strategy could be: get the cheapest access *KSs* (.8) and also the smallest response delays (.7). By a subcycle we mean the execution of a particular control strategy. By a cycle we mean the execution of one goal. We use the symbol  $j_k$  to refer to the  $k^{th}$  subcycle of cycle  $j$ .

On each subcycle,  $CKS_i$  assigns a mass to  $A(p_i^k)$ . Thus

$$n_i^{jk} \left( A \left( p_i^k \right) \right) = b_i^k c_i^k$$

where  $b_i^k$  marks the degree of belief that for the  $i^{th}$  performance it's the performance value  $p_i^k$  that is needed and  $c_i^k$  is the degree of belief that any element in  $A(p_i^k)$  has  $i^{th}$  performance value  $p_i^k$ .

We have  $\sum_k b_i^k = 1$  for all  $i$

$$0 \leq c_i^k \leq 1 \text{ for all } i \text{ and } k$$

and  $n_i^{j,k} (KS_1, KS_2, \dots, KS_s) = 1 - \sum_k b_i^k c_i^k$ . It is assumed that we always have

$$A(p_i^k) \neq \{KS_1, KS_2, \dots, KS_s\}$$

That is, every  $p_i^k$  determines a proper subset of  $\{KS_1, \dots, KS_s\}$ . Note that  $n_i^{j,k}$  do satisfy the properties of a belief structure that the excess mass, after  $b_i^k c_i^k$  has been assigned, is distributed over all  $KS$ s. Assume that a particular control strategy dictates that we look at particular values in performance spaces  $P_1, P_2$ , etc. ... Let  $X_1 = \{KS_{i_1}, \dots, KS_{i_p}\}$  be the set of  $KS$ s whose first performance characteristic matches  $p_1^{k_1}$ , similarly let  $X_2, X_3, \dots$  be sets of  $KS$ s whose performance characteristic matches  $p_2^{k_2}, p_3^{k_3}, \dots$ . Of course this information is given by  $CKS_1, CKS_2, \dots$  at the time the driving table is consulted for the appropriate control strategy to be followed. Thus the focal elements of  $n_i^{j,k}$  are of the form  $\tau(X_1) \times H_p \times H_p \times \dots$ . The focal elements of  $n_i^{j,k}$  are of the form  $H_p \times \tau(X_2) \times H_p \times \dots$  etc...

We now use the  $D-S$  theory to formulate a policy of selecting a set of  $KS$ s. Of course each  $X_i$  reported by  $CKS_i$  is assumed to be compatible with  $H_p$  i.e., if  $X_i = \{KS_2, KS_5, KS_{10}\}$  and  $\{KS_2, KS_5\} \in H_p$  but  $X_i \notin H_p$  then  $KS_{10}$  will not be considered, but running  $KS_2$  and  $KS_5$  in parallel will be considered. For example

$$n_i^{j,k} \left( \left\{ KS_2, KS_5 \right\} \times H_p \times H_p \times \dots \right)$$

will denote the degree of belief that for some value  $p_i^{j_1}$  for access cost (assuming access cost is  $P_1$ ) that value  $p_i^{j_1}$  is the best to consider compounded with the degree of belief that  $KS_2$  and  $KS_5$  have an access cost value of  $p_i^{j_1}$ . Now on subcycle  $j_k$  if we consider, for example, three performance characteristics which, for ease of notation, we assume to be  $P_1, P_2, P_3$  then we will consider the performance space

$$H_p^{j,k} = \epsilon(X_1) \times \epsilon(X_2) \times \epsilon(X_3)$$

where  $\in (X_i) (1 \leq i \leq 3)$  are the possible values of  $X_i$  (there may be several possible values of  $X_i$  due to uncertainty wherever  $p_i^{j_k}$  is the best value to consider in  $P_i$ ). If  $Y_1, Y_2, Y_3$  are possible values of  $X_1, X_2,$  and  $X_3$  then  $n_1^{j_k}$ , for example, will be induced on the set  $\{Y_1\} \times \in (X_2) \times \in (X_3)$  by

$$n_1^{j_k} \left( \left\{ Y_1 \right\} \times \in \left\{ X_2 \right\} \times \in \left\{ X_3 \right\} \right) = n_1^{j_k} \left( \left\{ Y_1 \right\} \times H_p \times H_p \times \dots \right)$$

thus, in contrast to  $H_Q, H_p^{j_k}$  (which is the set of answers of what  $KS$ s to use on subcycle  $j_k$ ) is highly dynamic and is generated by the appropriate control sequence. In general  $H_p^{j_k}$  is a set of elements of the form  $(A_1, A_2, \dots, A_{t_k})$  where  $t_k$  is the number of performance characteristics considered on a particular control strategy on cycle  $j$  and each  $A$  is an appropriate set of  $KS$ s. We now combine the masses on  $H_p^{j_k}$  by setting

$$n^{j_k} = \bigoplus_{i=1}^{t_k} n_i^{j_k}$$

The problem is to select the "best combination"  $V = (B_1, B_2, \dots, B_{t_k})$ . Note again that independence of information sources is justified as different  $CKS$ s are pooled. To get the "best combination"  $U$  we maximize

$$Bel(V) - Bel(\neg V)$$

over  $V \in H_p^{j_k}$ , and relative to the pooled belief structure  $n^{j_k}$ . The rationale for the optimization described above is to maximize the support of  $V$  over the support of its competitors. If several elements are tied for this optimization, we select one of the elements at random. If  $V$  is an optimal element we make an arbitrary selection  $b_i \in B_i, (1 \leq i \leq t_k)$ . Each  $b_i$  denotes some  $KS \in B_i$ . Now we follow the policy of running the  $b_i$  in parallel. We of course keep track of which  $b_i$  have been run and update  $H_p$  appropriately after each subcycle, so no  $KS$  will be accessed twice.

Now at the beginning of cycle  $j$  assume that the belief structure relative to subsets of  $H_Q$  is given by  $m^j$ . Of course  $m^j$  itself was obtained by pooling together the belief structures  $m_i$  where  $i$  ranges over all features considered prior to cycle  $j$ . On the first subcycle of cycle  $j$ , assume that the policy described above indicates that features  $i_1, i_2, \dots, i_t$  should be looked at. Pooling the corresponding  $KS$ s we obtain the belief structures  $m_{i_v}^{j_1} (1 \leq v \leq t)$  and we form

$$\Delta m^{j_1} = \bigoplus_{v=1}^t m_{i_v}^{j_1}$$

At the end of cycle  $j$ ,  $m^j$  is updated by

$$m^{j+1} \leftarrow m^j \oplus \Delta m^{j_1} \oplus \Delta m^{j_2} \dots \oplus \Delta m^{j_v}$$

where  $v$  denotes the number of control sequences in cycle  $j$ .

At the completion of each subcycle we check if

$$Pls(A(f_i^k)) < Bel(A(f_i^j)) \text{ for } j \neq k$$

We use the current value of  $m_j$  to do this. Of course that current value is given by

$$m^j \leftarrow m^{j-1} \oplus \Delta m^{j_1} \oplus \dots \oplus \Delta m^{j_k}$$

if we have just completed subcycle  $j_k$ .

If the inequality above is satisfied we rule out the proposition "the value of the  $i^{th}$  feature is  $f_i^k$ ". The rationale for doing this is if, for example, the upper probability for value green falls below the lower probability of some color value, say blue, then we should rule out that the color is green. Of course we could embellish this rule and require for instance that  $Bel(A(f_i^j))$ , at least initially, exceeds  $Pls(A(f_i^k))$  by some margin which could be relaxed as data accumulates. We then should perform a post-elimination update. There are many ways to do this. A straightforward method would be to distribute the mass of  $A(f_i^k)$  uniformly over the remaining  $A(f_i^j)$ . Of course if we have some additional information, the distribution may be chosen not to be uniform. We also could use the plausability of the remaining values of  $F_i$  to guide our redistribution of mass.

At the end of the cycle we have reduced the possible number of values in some of the sets  $F_i$ . We have at this point followed the actions dictated by the current goal. Now we take a new reading of the data and proceed to work in a similar manner on the next goal. A special note should be made when the control strategy is to reduce ignorance, which was defined in the previous section. Say the largest ignorance is for  $\{q_1, q_5, q_7\} \subset H_Q$ . The CKS in charge of ignorance (which is treated here as a particular feature) has access to all  $F_i (1 \leq i \leq s)$  and the corresponding  $A(f_i^k)$ . To reduce the ignorance on  $\{q_1, q_5, q_7\}$  one must get hold of those KSs that do not deal with  $q_1, q_5, q_7$  simultaneously (otherwise the ignorance may continue to propagate). Thus if KSs has information about a set containing say  $q_1$  but not  $q_5$  or  $q_7$ , then KSs is a viable candidate to reduce ignorance.

Finally when all the actions corresponding to all of the listed goals have been executed we end up with some of the sets  $F_i$  reduced (e.g. value, green and red may have been eliminated from  $F_3$ , if the third feature is color value) Now to get the answer to query  $Q$  we maximize

$$Bel(q) - Bel(\neg q) , q \in H_Q$$

relative to the final value of  $m^i$ .

The rationale for the optimization defined above is of course the same as earlier. We want to maximize the support for a particular answer over its competitors.

To sum up we have used the *D-S* theory to pool the evidence given by the *CKSs* to select a policy by which we would use the appropriate *Ks* to find information relative to query  $Q$ . Then we pooled the evidence collected from the *Ks* to rule out certain propositions about feature values. Each time we ruled out a proposition we had to redistribute its mass over remaining propositions pertaining to a specific feature. When all the steps are completed we pick the answer that maximizes support over its competitors. Of course, there are other ways to pool beliefs from different sources and for a sample of these methods we refer the reactor to [6], [11], [15], and [17].

Finally, if it is desired to access *Ks* more than once, some delay must be built into reaccessing any *KS* in order to approximate the independence of information sources. Recently, theories more general than the *D-S* approach have been studied. More flexible rules than the *D-S* rule of combination have been developed. These settings will be investigated in future work.

## REFERENCES

1. Alo, R., Kleyle, R.M. and de Korvin, A. (1985), "Decision Making Mechanisms and Stopping Times in a Generalized Information System", Mathematical Modelling, Vol. 6, pp. 259-271.
2. Alo, R., de Korvin, A. and Kleyle, R. (1987), "Emergence of a Dominant Course of Action in a Generalized Feedback Loop when Goal Uncertainty is Present", Journal of the American Society for Information Science, Vol. 38, pp. 111-117.
3. Davis, L.S., Rosenfeld, A. "Computer Vision Systems". Hanson, A.R., and Riseman, E.M. eds., 101, Academic Press, New York (1978).
4. Duda, R.O., Reboh, R., et al (1979) "A Computer Based Consultant For Mineral Exploration" SRI Final Report For Project II 6415, Artificial Intelligence Ctr., SRI International, Menlo Park, California.
5. Duda, R.O., et al (1976) "Subjective Bayesian Methods For Rule-Based Inference Systems" Proc. Nat. Comp. Conf. pp. 1075-1082. AFIPS Press, Reston, Va.
6. Garvey, T.D., (1976) "Perceptual Strategies For Purposive Vision" SRI International Tech. Note No. 117, Artificial Intelligence, Ctr, Menlo Park, California.
7. Kleyle, R.M. and de Korvin, A. (1984), "Switching Mechanisms in a Generalized Information System", Cybernetics and Systems, Vol. 15, pp. 145-167.
8. Kleyle, R.M. and de Korvin, A. (1986), "A Two Phase Approach to Making Decisions Involving Goal Uncertainty", Journal of Information Science, Vol. 12, pp. 161-171.
9. Kleyle, R.M., de Korvin, A. "A Belief Function Approach To Information Utilization In Decision Making" Submitted.
10. Pearl, J. "How To Do With Probabilities What People Say You Can't" The Second Conf. on Artificial Intelligence Appl. pp. 6-11, IEEE Comp. Soc. Press Miami, Florida (1985).
11. Selfridge, P.G (1982) "Reasoning About Success and Failure In Aerial Image Understanding" Ph.D. dissertation, Computer Science Dept. Univ. of Rochester.

12. Schafer, G. (1976), A Mathematical Theory of Evidence. Princeton, NJ, Princeton University Press.
13. Tanimoto, S.L., Computer Vision Systems. Hanson, A.R., and Riseman, E.M., eds Academic Press, New York (1978).
14. Waltz, D.L. Computer Vision Systems. Hanson, A.R., and Riseman, E.M., eds Academic Press, New York (1978).
15. Yakimovsky, Y., Feldman, J.A., (1973) in Proc. Third Int. Joint Conf. on Artificial Intelligence, 580. The International Joint Conf. on Artificial Intelligence.
16. Zadeh, L.A. (1986), "A Simple View of the Dempster-Shafer Theory of Evidence and its Implication for the Rule of Combination", The AI Magazine, pp. 85-90.
17. Zucker, S.W. in Pattern Recognition and Artificial Intelligence, Chen, C.H. ed Academic Press, New York.