Decision Analysis Using Belief Functions

Thomas M. Strat
Artificial Intelligence Center, SRI International, Menlo Park, California

ABSTRACT

A primary motivation for reasoning under uncertainty is to derive decisions in the face of inconclusive evidence. Shafer's theory of belief functions, which explicitly represents the underconstrained nature of many reasoning problems, lacks a formal procedure for making decisions. Clearly, when sufficient information is not available, no theory can prescribe actions without making additional assumptions. Faced with this situation, some assumption must be made if a clearly superior choice is to emerge. This paper offers a probabilistic interpretation of a simple assumption that disambiguates decision problems represented with belief functions. It is proved that it yields expected values identical to those obtained by a probabilistic analysis that makes the same assumption. A strict separation is maintained between evidence that carries information about a situation and assumptions that may be made for disambiguation of choices. In addition, it is shown how the decision analysis methodology frequently employed in probabilistic reasoning can be extended for use with belief functions. This generalization of decision analysis allows the use of belief functions within the familiar framework of decision trees.

KEYWORDS: ***belief functions, decision analysis, decision making, decision tree, Dempster–Shafer theory, evidential reasoning, reasoning under uncertainty***

INTRODUCTION

Decision analysis provides a methodological approach for making decisions. Uncertain states of nature are represented by probability distributions, and each possible state is assigned a value or *utility*. The best decision is the one that yields the greatest *expected utility*. By enumerating all available choices in a

Address correspondence to Thomas M. Strat, AI Center, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025.

International Journal of Approximate Reasoning 1990; 4:391–417

655 Avenue of the Americas, New York, NY 10010 0888-613X/90/$3.50

decision tree and assessing the probabilities and utilities of the states of nature that may result, one can mechanically determine the optimal sequence of actions that should be taken (Howard [1], Lapin [2], LaValle [3], Raiffa [4]).

In practice, these simple requirements are hard to satisfy (Horvitz et al. [5]). Sometimes reliable estimates of the probabilities involved are hard to come by. For example, few statistics are available for determining the probability of a nuclear reactor core meltdown. Assessing the utility of many-faceted states of nature is equally challenging. How should one give a unique value to the anticipated quality of married life? These limitations have hindered the more widespread application of decision analysis.

Shafer's theory of belief functions (Lowrance et al. [6], Ruspini [7], Shafer [8], Smets [9]) allows one to express partial beliefs when it is impossible or impractical to assess complete probability distributions confidently. Using belief functions, one can bound the probabilities of events for which the assignment of a precise probability would be misleading. The theory provides a facility to express one's beliefs only to the degree to which there is supporting evidence, thereby resulting in an appropriate description of an uncertain event. For example, there might be reason to assign a probability to a reactor malfunction without saying what the chance is that it may lead to a core meltdown.

Despite its representational advantages, the theory of belief functions lacks a formal basis upon which decisions can be made in the face of ambiguity (Barnett [10]). Computing the expected utility of a random event that has been represented with belief functions results in an *expected utility interval* (EUI). To choose between two actions one must compare their respective EUIs. If they do not overlap, the choice is clear. But when the EUIs overlap, the decision maker is confronted with a dilemma: the available evidence does not support either choice. Ideally, one should collect more information until the intervals no longer overlap and the choice becomes clear. However, sometimes one is forced to choose without benefit of additional information. What should be done?

In this situation there is no recourse except to make an assumption to eliminate the ambiguity. Various authors have expressed preference for different assumptions (such as renormalization, generalized insufficient reason (Dubois and Prade [11], Smets and Kennes [12]), minimax (Wald [13]), and optimism/pessimism (Hurwicz [14]). More elaborate schemes have been suggested, but they also amount to the introduction of unfounded assumptions (Loui et al. [15], Pittarelli [16], Yager [17]). Here I advocate the interpolation of a point-valued utility within the EUI. I make no claim that it leads to superior decisions, but I do claim that it is no less viable than the alternative assumptions. I show that it gives the same expected utility (and hence leads to the same decisions) as would be obtained by assuming that there is some probability that ambiguity will be resolved in one's favor.

I further show how decision analysis can be generalized to accommodate a belief function representation of uncertainty. This involves two modifications:

allowing an interval as the utility of a state or set of states, and allowing a belief function in place of a probability distribution. The result is a complete decision analysis procedure compatible with either probabilistic or belief function representations of uncertainty.

I should point out that decision theory (and its associated utility theory) is not the only approach for making decisions under uncertainty. For example, Lesh [18] has proposed a model based on an ignorance-preference coefficient that is empirically derived. Shafer [19] has advocated a "constructive" decision theory that seeks support for actions that achieve goals. Loui et al. [15] suggest representing beliefs not by distribution but by a sequence of progressively more decisive distributions. In this paper I am concerned with providing for the use of belief functions within the general framework of decision analysis.

It is worth noting that none of the material described in this paper depends on the use of Dempster's rule, which is commonly used in Shafer's theory to combine independent bodies of evidence (Shafer [8]). The computation of expected utility interval, and the procedure for using EUIs in decision analysis, only require that a belief function representation of the problem be available. Dempster's rule *could* be used to construct that belief function, but it is not required for decision analysis.

In the sections that follow the theory is developed and its use is illustrated with simple examples. In Section 2 we derive the expected utility interval that results from the use of belief functions. We then see how making an assumption about the probability of nature's cooperation leads to the same expected utility as interpolation within the EUI. In Section 3, this result is used to generalize decision analysis and is illustrated within a decision problem concerning whether or not to drill for oil. I conclude with a discussion of the benefits and limitations of this approach and compare its use with other approaches to decision making under uncertainty.

2. EXPECTED VALUE

Decision analysis provides a methodological approach for making decisions. The crux of the method is that one should choose the action that will maximize the expected utility. In this section I review the computation of expected utility using a probabilistic representation of a simple example and show how a belief function gives rise to a range of expected utilities. I then show how a simple assumption about the inclination of nature leads to a means for choosing a single-point expected utility for belief functions.

2.1. Expected Value Using Probabilities

EXAMPLE—CARNIVAL WHEEL NO. 1 A familiar game of chance is that of the carnival wheel pictured in Figure 1. This wheel is divided into 10 equal sectors,

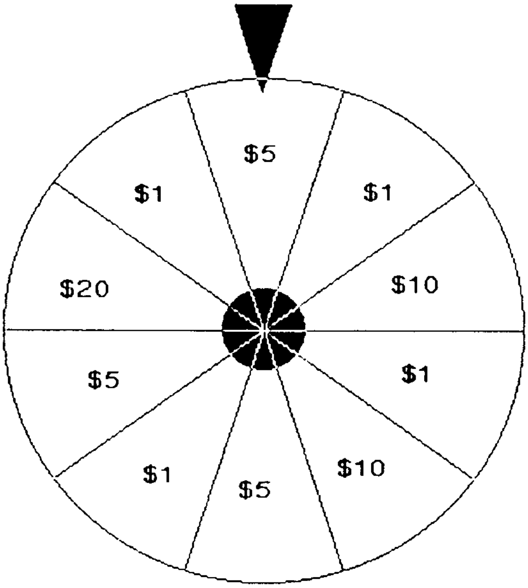

Figure 1. Carnival wheel No. 1.

each of which is labeled with a dollar amount as shown. For a \$6.00 fee, the player gets to spin the wheel and receives the amount shown in the sector that stops at the top. Should we be willing to play?

The analysis of this problem lends itself readily to a probabilistic representation. From inspection of the wheel (assuming each sector really is equally likely), we can construct the following probability distribution:

$$p(\$1) = 0.4$$

$$p(\$5) = 0.3$$

$$p(\$10) = 0.2$$

$$p(\$20) = 0.1$$

The expected value $E(x)$ is computed from the formula

$$E(x) = \sum_{x \in \Theta} x p(x) \tag{1}$$

where Θ is the set of possible outcomes. The expected value of the carnival wheel is $5.90 as shown here:

x	$p(x)$	$xp(x)$
1	0.4	0.4
5	0.3	1.5
10	0.2	2.0
20	0.1	2.0
	$E(x) =$	5.90

Therefore, we should refuse to play, because the expected value of playing the game is less than the $6.00 cost of playing.[1] Let us now modify the problem slightly in order to motivate a belief function approach to the problem.

2.2. Expected Value Intervals

EXAMPLE—CARNIVAL WHEEL NO. 2 Another carnival wheel is divided into 10 equal sectors, each having $1, $5, $10, or $20 printed on it. However, one of the sectors is hidden from view (Fig. 2). How much are we willing to pay to play this game?

This problem is ideally suited to an analysis using belief functions. In a belief function representation, a unit of belief is distributed over the space of possible outcomes (commonly called the *frame of discernment*). Unlike a probability distribution, which distributes belief over elements of the outcome space, this distribution (called a *mass function*) attributes belief to subsets of the outcome space. Belief attributed to a subset signifies that there is reason to believe that the outcome will be among the elements of that subset, without committing to any preference among those elements. Formally, a mass distribution m_Θ is a mapping from subsets of a frame of discernment Θ into the unit interval:

$$m_\Theta \colon 2^\Theta \mapsto [0, 1]$$

such that

$$m_\Theta(\emptyset) = 0 \qquad \text{and} \qquad \sum_{A_i \subseteq \Theta} m_\Theta(A_i) = 1$$

Any subset to which nonzero mass has been attributed is called a *focal element*. One of the ramifications of this representation is that the belief in a hypothesis

[1] We assume that the monetary value is directly proportional to utility because of the small dollar amounts involved. We could instead have chosen to work with utilities to account for nonlinearities in one's preferences for money.

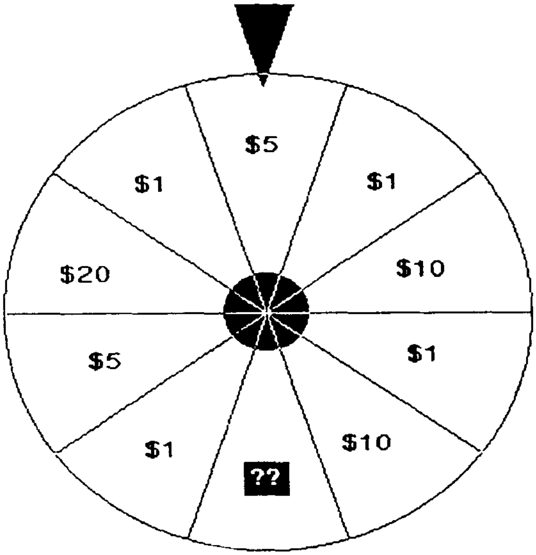

Figure 2. Carnival wheel No. 2.

A ($A \subseteq \Theta$) is constrained to lie within an interval [Spt(A), Pl(A)], where

$$\mathrm{Spt}(A) = \sum_{A_i \subseteq A} m_\Theta(A_i), \qquad \mathrm{Pl}(A) = 1 - \mathrm{Spt}(\neg A) \tag{2}$$

These bounds are commonly referred to as *support* and *plausibility*.

The frame of discernment Θ for wheel No. 2 is {\$1, \$5, \$10, \$20}. The mass function for wheel No. 2 is

$$m(\{\$1\}) = 0.4$$
$$m(\{\$5\}) = 0.2$$
$$m(\{\$10\}) = 0.2$$
$$m(\{\$20\}) = 0.1$$
$$m(\{\$1, \$5, \$10, \$20\}) = 0.1$$

and its associated belief intervals are

$$[\mathrm{Spt}(\{\$1\}), \mathrm{Pl}(\{\$1\})] = [0.4, 0.5]$$
$$[\mathrm{Spt}(\{\$5\}), \mathrm{Pl}(\{\$5\})] = [0.2, 0.3]$$
$$[\mathrm{Spt}(\{\$10\}), \mathrm{Pl}(\{\$10\})] = [0.2, 0.3]$$
$$[\mathrm{Spt}(\{\$20\}), \mathrm{Pl}(\{\$20\})] = [0.1, 0.2]$$

Before we can compute the expected value of the wheel represented by this belief function, we must somehow assess the value of the hidden sector. We know that there is a 0.1 chance that the hidden sector will be selected, but what value should we attribute to that sector? If the carnival hawker had been allowed to assign a dollar value to that sector, he would surely have assigned \$1. On the other hand, if we (or a cooperative friend) had been allowed to do so, it would have been \$20. Any other assignment method would result in a value between \$1 and \$20, inclusive. Therefore, if we truly do not know what assignment method was used, the strongest statement that we can make is that the value of the hidden sector is between \$1 and \$20. Using interval arithmetic we can apply the expected value formula of Eq. (1) to obtain an *expected value interval* (EVI):

$$E(x) = [E_*(x), E^*(x)] \tag{3}$$

where[2]

$$E_*(x) = \sum_{A_i \subseteq \Theta} \inf(A_i) m_\Theta(A_i)$$

$$E^*(x) = \sum_{A_i \subseteq \Theta} \sup(A_i) m_\Theta(A_i)$$

The expected value interval of wheel No. 2 is

$$E(x) = [0.4(1) + 0.2(5) + 0.2(10) + 0.1(20) + 0.1(1),$$
$$0.4(1) + 0.2(5) + 0.2(10) + 0.1(20) + 0.1(20)]$$
$$E(x) = [5.50, 7.40]$$

2.3. Expected Value Using Belief Functions

As many researchers have pointed out, an interval of expected values is not very satisfactory when we have to make a decision. Sometimes it provides all the information necessary to make a decision; for example, if the game costs \$5 to play, then clearly we should be willing to play regardless of who gets to assign a value to the hidden sector. Sometimes we can defer making the decision until we have collected more evidence; for example, if we could peek at the hidden sector and then decide whether or not to play. But the need to make a decision based on the currently available information is often inescapable; for example,

[2] We use $\inf(A_i)$ or $\sup(A_i)$ to denote the smallest or largest element in the set $A_i \subseteq \Theta$. Θ is assumed to be a set of scalar values (Strat [20]).

should we spin wheel No. 2 for a \$6 fee? We will present our methodology for decision making using belief functions after pausing to consider a Bayesian analysis of the same situation.

If we are to use the probabilistic definition of expected value from Eq. (1), we are forced to assess probabilities of all possible outcomes. To do this, we must make additional assumptions before proceeding further. One possible assumption is that all four values of the hidden sector (\$1, \$5, \$10, \$20) are equally likely, and we could evenly distribute among those four values the 0.1 chance that the hidden sector is chosen. This is an example of the generalized insufficient reason principle advanced by Dubois and Prade [11] and by Smets [12]. The resulting computation of expected value with this assumption is shown below; the expected value is \$6.30.

x	$p(x)$	$xp(x)$
1	0.425	0.425
5	0.225	1.125
10	0.225	2.250
20	0.125	2.500
	$E(x) =$	6.30

An alternative assumption is that the best estimate of the probability distribution for the value of the hidden sector is the same as the known distribution of the visible sectors. Using this assumption, the result is \$6.00:

x	$p(x)$	$xp(x)$
1	4/9	4/9
5	2/9	10/9
10	2/9	20/9
20	1/9	20/9
	$E(x) =$	6.00

Rather than making one of these assumptions, we may wish to parameterize by an unknown probability ρ our belief that either we get to choose the value of the hidden sector or the carnival hawker does. Let ρ be the probability that the value assigned to the hidden sector is the one that we would have assigned if given the opportunity, so $1-\rho$ is the probability that the carnival hawker chose the value of the hidden sector. That is,

$$p(\text{hidden sector is labeled \$20}) = \rho$$

$$p(\text{hidden sector is labeled \$1}) = 1 - \rho$$

The expected value of wheel No. 2 can then be recomputed using probabilities

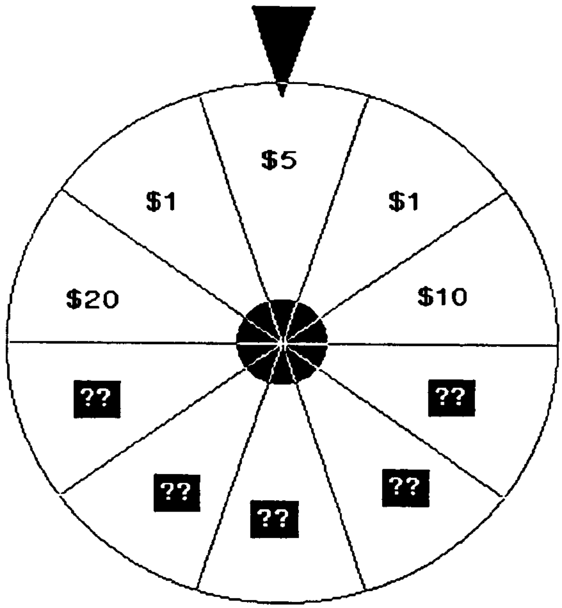

Figure 3. Carnival wheel No. 3.

and Eq. (1) as illustrated here:

x	$p(x)$	$xp(x)$
1	$0.4 + 0.1(1 - \rho)$	$0.5 - 0.1\rho$
5	0.2	1.0
10	0.2	2.0
20	$0.1 + 0.1\rho$	$2.0 + 2\rho$
	$E(x) =$	$5.50 + 1.90\rho$

To decide whether to play the game, we need only assess the probability ρ. For the carnival wheel it would be wise to allow that the hawker has hidden the value from our view; thus we might assume that $\rho = 0$. So $E(x) = 5.50$, and we should not be willing to pay more than \$5.50 to spin the wheel.

EXAMPLE—CARNIVAL WHEEL NO. 3 A third carnival wheel is divided into 10 equal sectors, each having \$1, \$5, \$10, or \$20 printed on it (Fig. 3). This wheel has five sectors hidden from view. However, we do know that none of these sectors is a \$20, that the first hidden sector is either a \$5 or a \$10, and that the second hidden sector is either a \$1 or a \$10. How much are we will to pay to spin wheel No. 3?

A probabilistic analysis of wheel No. 3 requires that we make additional assumptions. Estimating the conditional probability distribution for each hidden sector would provide enough information to compute the expected value of the wheel. Alternatively, estimating just the expected value of each hidden sector would suffice as well. However, doing so can be both tedious and frustrating: tedious because there may be many hidden sectors, and frustrating because we are being asked to provide information that, in actuality, we do not have. (If we knew the conditional probabilities or the expected values, we would have used them in our original analysis.) What is the minimum information necessary to establish a single expected value for wheel No. 3?

The probability ρ that we used to analyze wheel No. 2 can be used here as well.

DEFINITION *Let ρ = the probability that ambiguity will be resolved as favorably as possible; then $1 - \rho$ = the probability that ambiguity will be resolved as unfavorably as possible.*

Estimating ρ is sufficient to restrict the expected value of a belief function to a single point. It is easy to see that the expected value derived from this analysis as ρ varies from 0 to 1 is exactly the value obtained by linear interpolation of the EVI that results from using belief functions. The following derivation shows that this is true in general.

THEOREM 1 *Given a mass function m_Θ defined over a scalar frame Θ of utilities, and an estimate of ρ (the probability that all residual ambiguity will turn out favorably), the expected utility given m_Θ is*

$$E(x) = E_*(x) + \rho[E^*(x) - E_*(x)] \tag{4}$$

Proof Consider a mass function m_Θ defined over a frame of discernment Θ. Now consider any focal element $A \subseteq \Theta$ such that $m_\Theta(A) > 0$. Since ρ is the probability that a cooperative agent will control which $x \in A$ will be selected, and $1 - \rho$ is the probability that an adversary will be in control, then the probability that x will be chosen given that focal element A occurs is

$$p_\Theta(x|A) = \begin{cases} \rho & \text{if } x = \sup(A) \\ 1 - \rho & \text{if } x = \inf(A) \\ 0 & \text{otherwise} \end{cases}$$

Considering all focal elements in m_Θ, we can construct a probability distribution $p_\Theta(x)$ as follows:

$$p_\Theta(x) = \sum_{A_i \subseteq \Theta} p_\Theta(x|A_i) p_\Theta(A_i)$$

$$p_\Theta(x) = \sum_{A:\,\sup(A)=x} \rho m_\Theta(A) + \sum_{A:\,\inf(A)=x} (1-\rho)m_\Theta(A)$$

Using Eq. (1) we have

$$\begin{aligned}
E(x) &= \sum_{x\in\Theta} x p_\Theta(x) \\
&= \sum_{x\in\Theta} x \left(\sum_{A:\,\sup(A)=x} \rho m_\Theta(A) + \sum_{A:\,\inf(A)=x} (1-\rho)m_\Theta(A) \right) \\
&= \sum_{x\in\Theta} \left(\sum_{A:\,\sup(A)=x} \sup(A)\rho m_\Theta(A) + \sum_{A:\,\inf(A)=x} \inf(A)(1-\rho)m_\Theta(A) \right)
\end{aligned}$$

The double summations can be collapsed to a single summation because every $A \subseteq \Theta$ has a unique $\sup(A) \in \Theta$ and a unique $\inf(A) \in \Theta$.

$$\begin{aligned}
E(x) &= \sum_{A\subseteq\Theta} \sup(A)\rho m_\Theta(A) + \inf(A)(1-\rho)m_\Theta(A) \\
&= \sum_{A\subseteq\Theta} \inf(A)m_\Theta(A) + \rho \sum_{A\subseteq\Theta} [\sup(A) - \inf(A)]m_\Theta(A) \\
&= E_*(x) + \rho[E^*(x) - E_*(x)]
\end{aligned}$$

The important point here is that the probabilistic analysis provides a meaningful way to choose a distinguished point within an EVI that results from the use of belief functions. That distinguished point can then be used as the basis for comparison of several choices when their respective EVIs overlap.

2.4. Discussion

Because of its interval representation of belief, Shafer's theory poses difficulties for a decision maker who uses it. Lesh [18] has proposed a different method for choosing a distinguished point to use in the ordering of overlapping choices. Lesh makes use of an empirically derived "ignorance preference coefficient" τ that is used to compute the distinguished point called "expected evidential belief (EEB)":

$$\mathrm{EEB}(A) = \frac{\mathrm{Spt}(A) + \mathrm{Pl}(A)}{2} + \tau \frac{[\mathrm{Pl}(A) - \mathrm{Spt}(A)]^2}{2}$$

A choice is made by choosing the action that maximizes the "expected evidential value (EEV)":

$$\mathrm{EEV} = \sum_{A_i \subseteq \Theta} A_i\, \mathrm{EEB}(A_i)$$

There are some important differences between Lesh's approach and the present approach for evidential decision making. The ignorance preference parameter τ can be seen as a means for interpolating a distinguished value within a *belief* interval $[\mathrm{Spt}(A), \mathrm{Pl}(A)]$, while the cooperation probability ρ is used to interpolate within an interval of *expected utilities* $[E_*(x), E^*(x)]$. Second, Lesh's parameter τ has been derived empirically rather than theoretically. In contrast, the cooperation parameter ρ has been explained as a probability of a comprehensible event—that the residual ambiguity will be favorably resolved. It leads to a simple procedure involving linear interpolation between bounds of expected utility and is derived from probability theory.

The use of a single parameter to choose a value between two extremes is similar in spirit to the approach taken by Hurwicz with a probabilistic formulation [14]. Hurwicz suggested that rather than computing the expected utility of a variable for which a probability distribution is known, one could interpolate a decision index between two extremes by estimating a single parameter related to the disposition of nature. When this parameter is zero, one obtains the Wald minimax criterion—the assumption that nature will act as strongly as possible against the decision maker [13]. In contrast to the Hurwicz approach in which one ignores the probability distribution and computes a decision index on the basis of the parameter only, in my approach the expected utility interval is computed, and interpolation between extremes occurs only within the range of residual ambiguity allowed by the focal elements of a belief function. Thus my approach is identical to the use of expected utilities when a probability distribution is available, it is identical to Hurwicz's approach when there are known constraints on the distribution, and it combines elements of both when the distribution is a belief function.

There may be circumstances in which a single parameter is insufficient to capture the underlying structure of a decision problem. In these cases it would be more appropriate to use a different probability to represent the attitude of nature for each source of ambiguity. Let ρ_i be the probability that ambiguity within each focal element A_i will be decided favorably, $(\forall A_i)\ A_i \subseteq \Theta$. Then we obtain

$$E(x) = \sum_{A_i \subseteq \Theta} \inf(A_i) m_\Theta(A_i) + \sum_{A_i \subseteq \Theta} \rho_i [\sup(A_i) - \inf(A_i)] m_\Theta(A_i) \quad (5)$$

in place of Eq. (4).

3. DECISION ANALYSIS

In the preceding section I have defined the concept of an expected utility interval (EUI) for belief functions and have shown that it bounds the expected utility that would be obtained with any probability distribution consistent with that belief function. Furthermore, I have proposed a parameter (the probability that residual ambiguity will be decided in our behalf) that can be used as the basis for computing a unique expected utility, when the available evidence warrants only bounds on that expected utility. In this section I will show how the EUI can be used to generalize probabilistic decision analysis.

Decision analysis was first developed as a means by which one could organize and systematize one's thinking when confronted with an important and difficult choice (Howard [1], Raiffa [4]). Its formal basis has made it adaptable as a computational procedure by which computer programs can choose actions when provided with all relevant information. Simply stated, the analysis of a decision problem under uncertainty entails the following steps:

1. List the viable options available for gathering information, for experimentation, and for action.
2. List the events that may possibly occur.
3. Arrange the information you may acquire and the choices you may make in chronological order.
4. Decide the value to you of the consequences that result from the various courses of action open to you.
5. Judge the chances that any particular uncertain event will occur.

3.1. Decision Analysis Using Probabilities

First we will illustrate the use of decision analysis on a problem that can be represented with probabilities to acquaint the reader with the method and terminology.

OIL DRILLING EXAMPLE 1 A wildcatter must decide whether or not to drill for oil. He is uncertain whether the hole will be dry, have a trickle of oil, or be a gusher. Drilling a hole costs \$70,000. The payoffs for hitting a gusher, a trickle, or a dry hole are \$270,000, \$120,000, and \$0, respectively. At a cost of \$10,000, the wildcatter could take seismic soundings that would help determine the underlying geologic structure. The soundings will determine whether the terrain has no structure, open structure, or closed structure. The experts have provided us with the joint probabilities shown below. We are to determine the optimal strategy for experimentation and action (Lapin [2]).

State	No struct.	Open	Closed	Marginal
Dry	0.30	0.15	0.05	0.50
Trickle	0.09	0.12	0.09	0.30
Gusher	0.02	0.08	0.10	0.20
Marginal	0.41	0.35	0.24	1.00

In decision analysis, a decision tree is constructed that captures the chronological order of actions and events (Lapin [2], LaValle [3]). A square is used to represent a decision to be made, and its branches are labeled with the alternative choices. A circle is used to represent a chance node, and its branches are labeled with the conditional probability of each event, given that the choices and events along the path leading to the node have occurred.

To compute the best strategy, the tree is evaluated from its leaves toward its root.

- The value of a leaf node is the utility of the state of nature it represents.
- The value of a chance node is the expected utility of the probability distribution represented by its branches as computed using Eq. (1).
- The value of a choice node is the maximum of the utilities of each of its sons. The best choice for the node is denoted by the branch leading to the son with the greatest utility. Ties are broken arbitrarily.

This procedure is repeated until the root node has been evaluated. The value of the root node is the expected utility of the decision problem; the branches corresponding to the maximal value at each choice node give the best *strategy* to follow (i.e., choices to make in each situation).

The evaluated decision tree for the oil drilling example is portrayed in Figure 4. It can be seen that the expected value is $22,500 and that the best strategy is to take seismic soundings, to drill for oil if the soundings indicate open or closed structure, and not to drill if the soundings indicate no structure.

3.2. Decision Analysis Using Belief Functions

To use the decision procedure just described, it must be possible to assess the probabilities of all uncertain events. That is, the set of branches emanating from each chance node in the decision tree must depict a probability distribution. In many scenarios, however, estimating these probability distributions is difficult or impossible, and the decision maker is forced to assign probabilities even though he knows they are unreliable. Using belief functions, one need not estimate any probabilities that are not readily available. The representation better reflects the evidence at hand, but the decision analysis procedure cannot be used

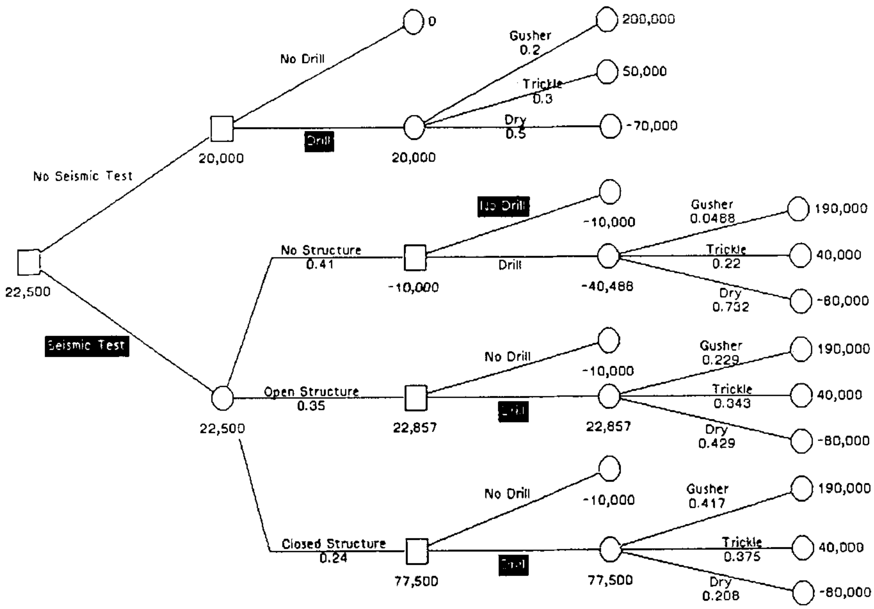

Figure 4. Decision tree for oil drilling example 1.

with the resulting interval representation of belief. In this section I describe a generalization of decision analysis that accommodates belief functions.

Prob.	Test Result	Capacity
0.5	Red	Dry
0.2	Yellow	Dry or Trickle
0.3	Green	Trickle or Gusher

OIL DRILLING EXAMPLE 2 As in the first oil drilling example, a wildcatter must decide whether or not to drill for oil. His costs and payoffs are the same as before: drilling costs \$70,000, and the payoffs for hitting a gusher, a trickle, or a dry well are \$270,000, \$120,000, and \$0, respectively. However, at this site, no seismic soundings are available. Instead, at a cost of \$10,000, the wildcatter can make an electronic test that is related to the well capacity as shown below. We are to determine the optimal strategy for experimentation and action.

Several issues arise that prevent us from constructing a well-formed decision tree for this example. First, consider the branch of the tree in which the test is conducted and the result is green (Fig. 5). If we drill for oil, then we know we will find either a trickle or a gusher, but we cannot determine the probability of either from the given information. We are tempted to label the branch with the disjunction (Trickle $\vee$ Gusher) with probability 1.0. But what should be the

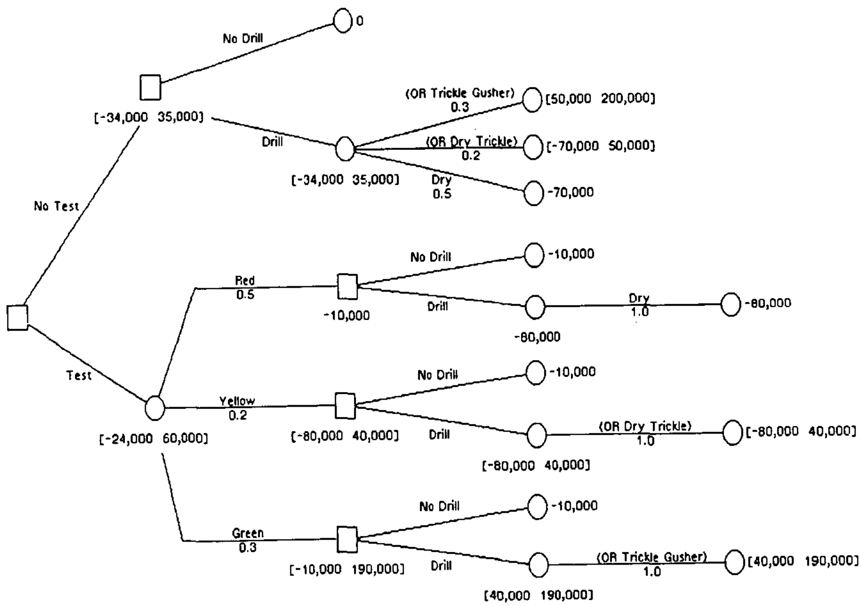

Figure 5. Modified decision tree for oil drilling example 2.

payoff of that branch? All we can say is that the payoff will be either \$40,000 (if a trickle) or \$190,000 (if a gusher). Ordinary decision analysis requires a unique value to be assigned, but we have no basis for computing one. So the first modification we make to the construction of decision trees is to allow disjunctions of events on branches emanating from chance nodes, and to allow intervals as the payoffs for leaf nodes. We will discuss later how to evaluate such a tree.

To see the second issue, consider the branch of the tree in which the test is not conducted. If we drill for oil, there is a chance that we will hit a gusher, a trickle, or a dry well, but what is the probability distribution? We know only that

$$p(\text{Dry}|\text{Red}) = 1.0 \qquad p(\text{Red}) = 0.5$$

$$p(\text{Dry} \vee \text{Trickle}|\text{Yellow}) = 1.0 \qquad p(\text{Yellow}) = 0.2$$

$$p(\text{Trickle} \vee \text{Gusher}|\text{Green}) = 1.0 \qquad p(\text{Green}) = 0.3$$

There is not enough information to use Bayes' rule to compute the probability distribution for the well capacity. Without adding a new assumption at this point, the strongest statement that can be made is

$$0.5 \leq p(\text{Dry}) \leq 0.7$$

$$0.0 \leq p(\text{Trickle}) \leq 0.5$$

$$0.0 \leq p(\text{Gusher}) \leq 0.3$$

Using belief functions, this can be represented as

$$m(\{\text{Dry}\}) = 0.5$$
$$m(\{\text{Dry, Trickle}\}) = 0.2$$
$$m(\{\text{Trickle, Gusher}\}) = 0.3$$

which yields the required belief intervals

$$[\text{Spt}(\{\text{Dry}\}), \text{Pl}(\{\text{Dry}\})] = [0.5, 0.7]$$
$$[\text{Spt}(\{\text{Trickle}\}), \text{Pl}(\{\text{Trickle}\})] = [0.0, 0.5]$$
$$[\text{Spt}(\{\text{Gusher}\}), \text{Pl}(\{\text{Gusher}\})] = [0.0, 0.3]$$

the second modification we make to decision trees is to allow the branches emanating from a chance node to represent a mass function. The masses must still sum to 1, but the events need not be disjoint.[3] The completed decision tree for oil drilling example 2 is shown in Figure 5.

The tools of Section 2 can be used to evaluate a decision tree modified in this manner.

- The value of a leaf node is the utility of the state of nature it represents. This may be a unique value or, in the case of a disjunction of states, an interval of values.
- A chance node represents a belief function. Its value is the EUI computed with Eq. (3):

$$E(x) = [E_*(x), E^*(x)]$$

- A decision node represents a choice of the several branches emanating from it. The utility of each branch may be a point value or an interval. The value of a decision node is the expected utility computed using Eq. (4) and an estimate of ρ:

$$E(x) = E_*(x) + \rho[E^*(x) - E_*(x)]$$

  The action on the branch that yields the greatest $E(x)$ is chosen. Ties are broken arbitrarily.

In summary, a decision tree and decision analysis procedure for belief functions have been described. Two modifications were made to adapt ordinary decision trees: intervals are allowed where utilities occur, and belief functions

[3] Recall that a probability distribution is an assignment of belief over mutually exclusive elements of a set, whereas a mass function is a distribution over possibly overlapping subsets.

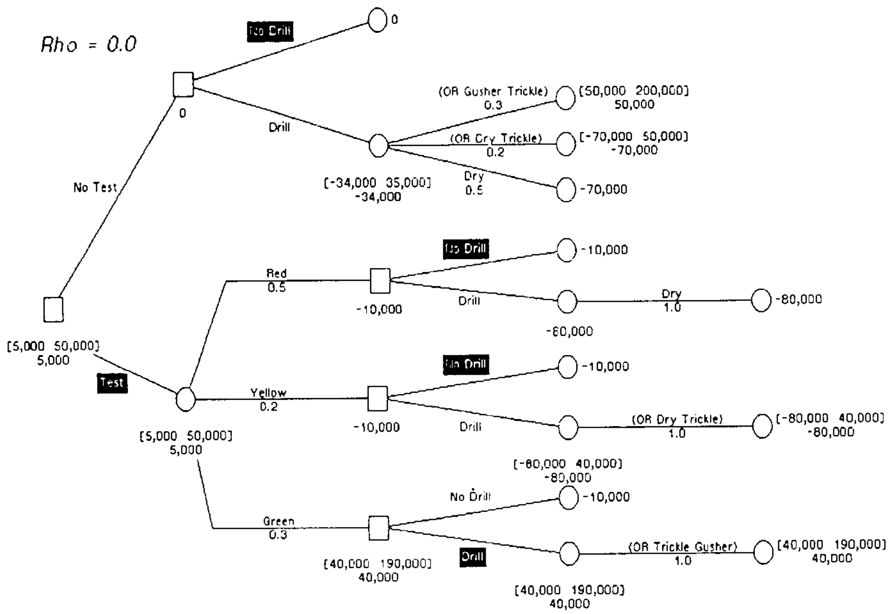

Figure 6. Decision tree for oil drilling example 2 assuming $\rho = 0.0$.

are allowed where probability distributions occur. A unique strategy can be obtained by estimating the probability ρ.[4]

3.3. Generalized Decision Tree Examples

Figures 6–8 show the evaluated decision tree for several values of ρ; each node is labeled with its expected value or expected value interval. In the cases where the expected value is an interval, the evidential expected value $E(x)$ is also shown (using the assumed ρ). Preferred decisions are highlighted with a black background.

If we opt not to test, then our choice is either to not drill (expected value 0) or to drill (EVI $[-34{,}000,\ 35{,}000]$). The better choice depends on what value of ρ is assumed. As can be seen in the figures, if $\rho = 0.0$, then it is better to not drill, but if $\rho = 0.5$ or $\rho = 1.0$, then drilling is the better choice.

If we choose to test and the result is yellow, then our choice is to not drill (expected value $-10{,}000$) or to drill (EVI $[-80{,}000, 40{,}000]$). In this case it is better to not drill if either $\rho = 0.0$ or $\rho = 0.5$ and to drill if $\rho = 1.0$.

[4] When all utilities are point-valued and all belief functions are true probability distributions, no assumption is required, and the strategy will be identical to that prescribed by ordinary decision analysis.

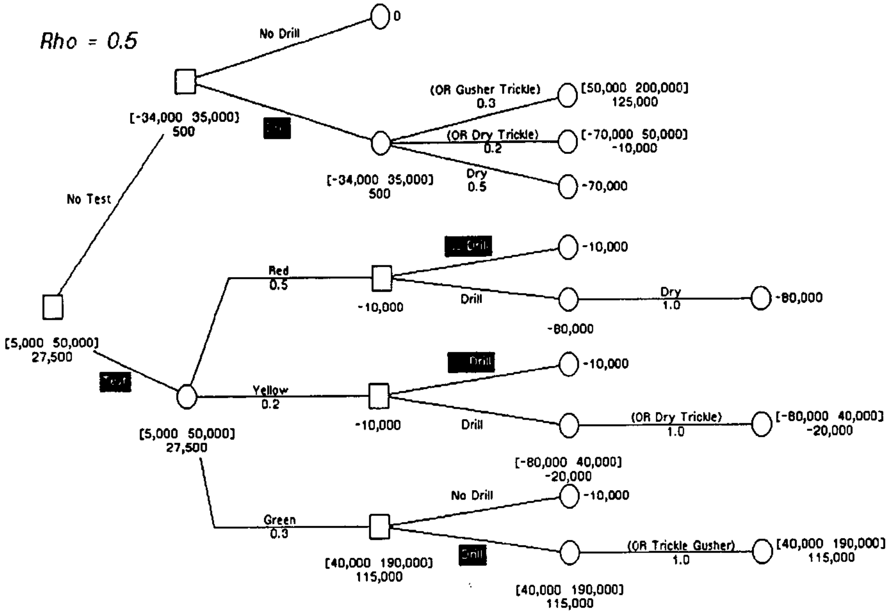

Figure 7. Decision tree for oil drilling example 2 assuming $\rho = 0.5$.

If the test result is red, then one should not drill regardless of ρ ($-10,000$ is always better than $-80,000$), If the test result is green, then one should always drill ($-10,000$ is never as good as the interval $[40,000, \quad 190,000]$).

3.4. Comparing Two Choices

Instead of assuming a value for ρ first, and calculating the choices that result, one may ask the reverse question. At what value of ρ would I change my decision? This can be answered in general by examining a choice between two states having EUIs.

THEOREM 2 *Let the expected utility intervals of two choices be as follows*:

$$\text{Choice1:} \qquad [E_{1*}(x), E_1^*(x)]$$

$$\text{Choice2:} \qquad [E_{2*}(x), E_2^*(x)]$$

Assume without loss of generality that choice 1 has the smaller interval, that is, $[E_2^*(x) - E_{2*}(x)] > [E_1^*(x) - E_{1*}(x)]$. *Then choice 2 is preferred over choice 1 if and only if*

$$\rho > \frac{E_{1*}(x) - E_{2*}(x)}{E_2^*(x) - E_1^*(x) + E_{1*}(x) - E_{2*}(x)} \tag{6}$$

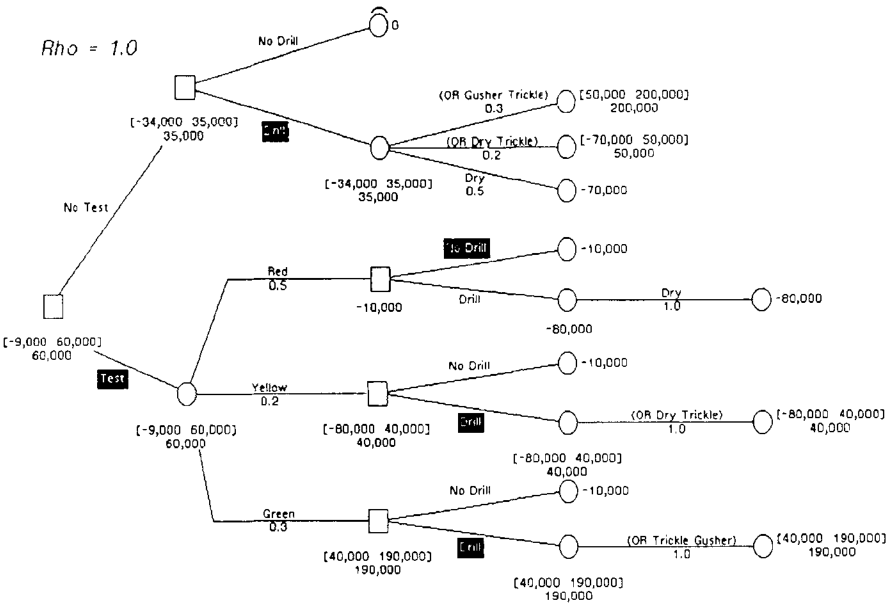

Figure 8. Decision tree for oil drilling example 2 assuming $\rho = 1.0$.

Proof Using Theorem 1 and solving for ρ gives the point ρ_c at which one is indifferent between choice 1 and choice 2:

$$E_1(x) = E_{1*}(x) + \rho[E_1^*(x) - E_{1*}(x)]$$
$$E_2(x) = E_{2*}(x) + \rho[E_2^*(x) - E_{2*}(x)]$$

$$\rho_c = \frac{E_{1*}(x) - E_{2*}(x)}{[E_2^*(x) - E_1^*(x)] + [E_{1*}(x) - E_{2*}(x)]} \tag{7}$$

The expected value of both choices at ρ_c is

$$E_c(x) = \frac{E_{1*}(x)E_2^*(x) - E_1^*(x)E_{2*}(x)}{E_{1*}(x) - E_{2*}(x) + E_2^*(x) - E_1^*(x)}$$

Now consider the choice at $\rho = \rho_c + \delta$, where $\delta > 0$:

$$E_1(x) = E_c(x) + \delta[E_1^*(x) - E_{1*}(x)] \tag{8}$$
$$E_2(x) = E_c(x) + \delta[E_2^*(x) - E_{2*}(x)] \tag{9}$$

Since $[E_2^*(x) - E_{2*}(x)] > [E_1^*(x) - E_{1*}(x)]$ and $\delta > 0$, it must be the case that $E_2(x) > E_1(x)$. Therefore, choice 2 is preferred. A similar argument shows that choice 1 is preferred when $\rho < \rho_c$.

Letting

$$a = E_{1*}(x) - E_{2*}(x) \qquad \text{and} \qquad b = E_2^*(x) - E_1^*(x)$$

gives

$$\rho_c = a/(a+b) \tag{10}$$

Thus, choice 1 is preferable if

$$\rho < a/(a+b)$$

and choice 2 is preferable if

$$\rho > a/(a+b)$$

If $a/(a+b) > 1.0$, then choice 1 is always preferred (no assumption of ρ is necessary). If $a/(a+b) < 0.0$, then choice 2 is always preferred. It follows that whenever one EUI is slightly "higher" than another, that is, when

$$E_{1*}(x) > E_{2*}(x) \qquad \text{and} \qquad E_1^*(x) > E_2^*(x)$$

then the action that gives rise to it is always preferred.

Returning to oil drilling example 2, the decision of whether or not to drill when the test result is yellow involves a choice between

$$\begin{array}{ll} \text{No drill:} & E(x) = [-10{,}000, \quad -10{,}000] \\ \text{Drill:} & E(x) = [-80{,}000, \quad 40{,}000] \end{array}$$

By Theorem 2, $\rho_c = 0.583$, and one should drill only if $\rho > 0.583$.

When $\rho > 0.583$, the decision as to whether or not to conduct the test involves a choice between

$$\text{No test:} \quad E(x) = [-34{,}000, \quad 35{,}000]$$

and

$$\text{Test:} \quad E(x) = [-9{,}000, \quad 60{,}000]$$

Here, Test is the preferred choice because its EUI is higher.

4. DISCUSSION

The value of the result of an action is frequently measured in money (e.g., in dollars), but people often exhibit preferences that are not consistent with maximization of expected monetary value. The theory of *utility* accounts for this behavior by associating for an individual decision maker a value (measured in *utiles*) with each state s, $u = f(s)$, such that maximization of expected utility yields choices consistent with that individual's behavior (Howard [1]). Utility theory can satisfactorily account for a person's willingness to expose himself to risk and should be used whenever one's preferences are not linearly related to value. This attitude toward risk should not be confused with one's attitude toward ambiguity, which is the quality that is modeled by ρ.

4.1. On Making Assumptions

It is interesting to compare the types of assumptions made in a probabilistic analysis with the ρ assumption proposed here for belief functions. When using probability, a maximum entropy assumption is often made. Sometimes this assumption is justified, and it should properly be considered part of the evidence, not an assumption. When this is the case, a maximum entropy belief function can be used as well (Dubois and Prade [11]). At other times, the maximum entropy assumption is not justified but is used simply because *some* assumption must be made and maximum entropy has some desirable properties (Smets and Kennes [12]). In these cases, the choice of elements in the sample space (the set of possibilities) introduces distortion into the expected value that will result. That is, adding a few more possibilities into the sample space will change the expected value of the maximum entropy distribution over that sample space. For example, if we chose to allow for the possibility of \$2 being among the possibilities for the hidden sector of carnival wheel No. 2, the sample space would be $\{1, 2, 5, 10, 20\}$ instead of $\{1, 5, 10, 20\}$, and the expected value of the maximum entropy distribution of that wheel would be \$6.16 instead of \$6.30. On the other hand, for any choice of ρ, the evidential expected value using either of the two preceding sample spaces would be identically $(5.50 + 1.90\rho)$ dollars. Of course, adding possibilities outside the interval $[1, 20]$ would change the evidential expected value. For example, allowing for the possibility of \$50 in the hidden sector would change the maximum entropy expected value to \$7.12 and would change the evidential expected value to $(5.50 + 4.90\rho)$ dollars. The point is that both assumptions introduce bias into the decision criteria. This should not be surprising because both are unjustified assumptions. There is no basis on which to prefer one over the other; both assumptions are entirely plausible.

Having made this point, there are some consequently weak arguments for recommending the use of the assumption of the probability of nature's cooperation ρ. Because the EUI spans the range of all expected utilities that could be obtained by adding *any* assumption to a probabilistic analysis, there always exists some value of ρ, $0 \leq \rho \leq 1$, that yields the same expected utility $E(x)$ as a probabilistic analysis. Therefore, the decisions that are prescribed depend only on one's ability to estimate ρ, not on his election to use Eq. (3). Furthermore, the use of a single parameter means that the decision maker is asked to provide only one additional piece of information.

The parameter ρ has been explained as a probability, giving it a formal grounding that earlier decision schemes for belief functions have lacked. Furthermore, I believe that it is the probability of a meaningful event. Selecting $\rho = 0$ is appropriate when an adversary controls the situation (as in game playing, for example) or when a decision maker wishes only to minimize expected loss, and is equivalent to the maximin criteria of Wald. An optimistic decision maker would prefer to choose $\rho = 1$ to maximize his chance of realizing the greatest possible expected payoff without worrying about what losses might be possible. Intermediate values of ρ can be used to compromise between these extremes.

4.2. On the Limitations of the Approach

Despite the appeal of a computationally efficient decision analysis procedure for belief functions, there remain some issues that are not addressed. As in classical decision analysis, it remains necessary to enumerate the potential states of nature and to assign utilities (actually utility intervals, which should be easier to assign in practice). This task can be overwhelming when complex scenarios are considered. Furthermore, it should not be forgotten that the assignment of a value to ρ (when it is necessary) remains an assumption unwarranted by the evidence at hand, just as maximum entropy or any other assumption is unwarranted when insufficient information is available.

It is inherent in the methodology described that the determination of what is best or worst is considered after the decision-maker's choice is postulated. That is, the reaction of nature is allowed to depend on the decision that is to be taken. This is sometimes reasonable and sometimes not. For example, conducting regional test marketing for a new product may affect the national demand by virtue of publicity or increased competition. As a result, there may be no single underlying probability distribution that can simultaneously give rise to the expected utilities obtained for each choice. This should not be particularly worrisome as long as this consideration suits the problem at hand. If not, the EUIs computed with the method described here may be wider than prescribed by the evidence. In that case, it is necessary to conduct a more complicated case-based analysis that is analogous to the linear programming problems that arise in game theory. See Jaffray [21] for further discussion of this approach.

4.3. On the Automation of Decision Analysis

A probabilistic analysis of a decision problem (e.g., the second oil drilling example) follows the paradigm: assess, assume, combine, decide. An *assessment* of a probability distribution is made for each piece of evidence; *assumptions* are made about the distributions of missing pieces of evidence; the assumptions and evidence are *combined* to obtain a distribution of payoffs, and a *decision* is made on the basis of the expected utility of the payoff. In contrast, a belief function analysis follows the paradigm: assess, combine, assume, decide. An *assessment* of a belief function is made for each piece of evidence; these pieces of evidence are then *combined* to obtain a belief function over the possible payoffs; then an *assumption* is made (about the benevolence of nature); and a *decision* is made using that assumption and the EUI of the payoffs.

While the same decisions will be reached whether one makes assumptions first and then combines evidence or combines evidence and then adds those assumptions, the difference in paradigms has important implications for automating the procedure. First, in some decision problems the EUI of the top choice will not overlap the EUI of any other choice, that is, the decision follows from what is truly known and in no way depends upon the accuracy of any assumption that might be made. Using belief functions, the best decision in this case is immediately determinable without additional assumptions. Because Bayes' rule requires a prior distribution, this situation cannot be recognized without a more complex sensitivity analysis when a purely probabilistic representation is used. Second, when an assumption must be made because intervals do overlap, making it as late as possible allows one to maintain the assumption-free intermediate calculations for use in other computations. This is not an issue when the evidence will be used once and discarded but affords considerable computational savings when other decisions must be based on the original evidence plus new evidence as it comes along. Third, consider what must be computed if one chooses to use a different assumption (as needed for sensitivity analysis, for example). In a probabilistic analysis the assumptions and all evidence must be recombined before a decision can be made, because the assumptions are needed to combine the evidence. Using belief functions, one need only combine the new assumption with the already combined evidence before selecting the decision. This separation of evidence and assumption is similar in spirit to the distinction between credal and pignistic beliefs described by Smets [22].

5. SUMMARY

I have proposed a decision analysis methodology for Shafer's theory of belief functions. I started by defining the notion of expected utility interval (EUI) and showed it to properly bound the expected utility of any probability distribution that could be obtained by introducing additional assumptions. Because an EUI is often insufficient for decision making, a point value must be chosen to compare alternative choices. I then showed how a linear interpolation of a distinguished value within the EUI is equivalent to making an assumption of the benevolence or maleficence of nature. Letting ρ be the probability that ambiguity will be resolved favorably, I derived that distinguished point.

I have also shown how the theory can be used to generalize the decision trees used in probabilistic decision analysis. These tools allow a decision maker to defer unwarranted assumptions until the latest possible moment. In so doing he can sometimes avoid making any assumptions at all. Otherwise, he is forced to provide only enough additional information to allow a clear choice and has the benefit of all available information to selectively decide when he would like to make that assumption.

These techniques have been implemented and that software has been used to generate the decision trees shown in the figures in this paper. In addition, a new evidential operator for decision making has been added to the repertoire of the evidential reasoning technology developed at SRI International (Lowrance et al. [6]). Decision analysis has been incorporated into Gister,[5] SRI's evidential reasoning system, which uses the Dempster–Shafer theory of belief functions as its underlying representation.

What I have described is by no means a full theory of decision making for belief functions. Rather, I hope it may provide some insight that will someday lead to a better understanding of decision making with incomplete information.

ACKNOWLEDGMENTS

I thank John Lowrance, Thomas Garvey, Leonard Wesley, and Enrique Ruspini for their contributions to the ideas presented in this paper. I am also grateful for the close scrutiny provided by Philippe Smets, who called attention to the difference between this approach and the assumption of a single underlying probability distribution. This research was partially supported by the Defense Advanced Research Projects Agency under contract N00039-88-C-0248 in conjunction with the U.S. Navy Space and Naval Warfare Systems Command.

NOTATION

$p_\Theta(x)$	Probability distribution over sample space Θ, $x \in \Theta$
$m_\Theta(A)$	Mass function defined over frame of discernment Θ, $A \subseteq \Theta$
$\mathrm{Spt}(A)$	Support: $\mathrm{Spt}(A) = \sum_{A_i \subseteq A} m_\Theta(A_i)$
$\mathrm{Pl}(A)$	Plausibility: $\mathrm{Pl}(A) = 1 - \mathrm{Spt}(\neg A)$
$E(x)$	Expected value of a random variable whose outcome is governed by a probability distribution: $E(x) = \sum_{x \in \Theta} x p_\Theta(x)$

[5] Gister is a trademark of SRI International.

$E(x)$	Evidential expected value—the expected value of a variable governed by a belief function assuming that any residual ambiguity will be decided favorably with probability ρ: $E(x) = (1-\rho)E_*(x) + \rho E^*(x)$
$E^*(x)$	Upper bound of expected value: $E^*(x) = \sum_{A_i \subseteq \Theta} \sup(A_i) m_\Theta(A_i)$
$E_*(x)$	Lower bound of expected value: $E_*(x) = \sum_{A_i \subseteq \Theta} \inf(A_i) m_\Theta(A_i)$
EVI	Expected value interval: $[E_*(x), E^*(x)]$
EUI	Expected utility interval: same as EVI, when Θ is a frame of utilities
ρ	The probability that any residual ambiguity will be decided favorably
$1-\rho$	The probability that any residual ambiguity will be decided unfavorably
ρ_c	The value of ρ at which one would be indifferent between two choices

References

1. Howard, R. A., Decision analysis: applied decision theory, *Proceedings of the 4th International Conference on Operational Research*, 1966, pp. 55–77.

2. Lapin, L. L., *Quantitative Methods for Business Decisions*, Harcourt Brace Jovanovich, San Diego, 1981.

3. LaValle, I. H., *An Introduction to Probability, Decision, and Inference*, Holt, Rinehart and Winston, New York, 1970.

4. Raiffa, H., *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley, Reading, Mass., 1970.

5. Horvitz, E. J., Breese, J. S., and Henrion, M., Decision theory in expert systems and artificial intelligence, *Int. J. Approximate Reasoning*, 2(3), 247–302, 1988.

6. Lowrance, J. D., Garvey, T. D., and Strat, T. M., A framework for evidential-reasoning systems, *Proceedings AAAI-86*, Philadelphia, August 1986.

7. Ruspini, E. H., The logical foundations of evidential reasoning, SRI International Tech. Note 408, December 1986.

8. Shafer, G. A., *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, N.J., 1976.

9. Smets, P., Belief functions, in *Non-Standard Logics for Automated Reasoning* (P. Smets, E. Mamdani, D. Dubois, and H. Prade, Eds.), Academic, London, 1988, Chap. 9.

10. Barnett, J. A., Computational methods for a mathematical theory of evidence, *Proceedings*, VII IJCAI, Vancouver, Canada, August 1981, pp. 868–875.

11. Dubois, D., and Prade, H., On several representations of an uncertain body of evidence, in *Fuzzy Information and Decision Processes* (M. M. Gupta and E. Sanchez, Eds.), North-Holland, Amsterdam, 1982, pp. 167–181.

12. Smets, P., and Kennes, R., The transferable belief model: Comparison with bayesian models, to be published.

13. Wald, A., *Statistical Decision Functions*, Wiley, New York, 1950.

14. Hurwicz, L., A criterion for decision-making under uncertainty, Tech. Report 355, Cowles Commission, 1952.

15. Loui, R. P., Feldman, J. A., and Kyburg, H. G. E., Jr., Interval-based decisions for reasoning systems, *Proceedings of the Workshop on Uncertainty and Probability in Artificial Intelligence*, UCLA, Cal., August 1985, pp. 193–199.

16. Pittarelli, M., Decision making with linear constraints on probabilities, *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, Univ. Minnesota, August 1988, pp. 283–290.

17. Yager, R. R., Decision making under Dempster-Shafer uncertainties, Iona College Machine Intelligence Institute Tech. Report MII-915, undated.

18. Lesh, S. A., An evidential theory approach to judgment-based decision making, PhD thesis, Dept. of Forestry and Environmental Studies, Duke Univ., December 1986.

19. Shafer, G. A., Constructive decision theory, Univ. of Kansas Dept. of Mathematics Working Paper, December 1982.

20. Strat, T. M., Continuous belief functions for evidential reasoning," *Proceedings, AAAI-84*, Austin, Texas, August 1984.

21. Jaffray, J.-Y., Application of utility theory to belief functions, in *Uncertainty and Intelligence Systems* (B. Bouchon, L. Saitta, and R. Yager, Eds.), Springer-Verlag, Berlin, 1988, pp. 1–8.

22. Smets, P., Decision under uncertainty, *Proceedings of the Workshop on Uncertainty in Artificial Intelligence*, August 1989.