Research in Accounting Ethics, Vol. 2, 1996, pp. 109-130.

Value Judgments Using Belief Functions*

Rajendra P. Srivastava**
Ernst & Young Professor
Director, Ernst & Young Center for Auditing Research
and Advanced Technology
Division of Accounting and Information Systems
School of Business, The University of Kansas
Lawrence, KS 66045, USA

May 1995

*This article has been accepted for publication in *Research in Accounting Ethics* .

**The author would like to thank the accounting and finance workshop participants at The University of Kansas, in particular, Peter Gillett, Keith Harrison, Mark Hirschey, Maurice Joy, and George Pinches for their invaluable comments and suggestions.

Value Judgments Using Belief Functions

ABSTRACT

The main purpose of this article is to show how the belief-function framework can be used to model human behavior when making decisions that involve value judgments. A rule of decision making has been presented in the form of a proposition. This proposition is simple and intuitive. It predicts correctly the observed behavior, even the behavior that is considered to be irrational under the expected utility theory (EUT). We use the empirical data of Tversky and Kahneman (1981, 1986) to show how well the proposed rule predicts the observed behavior. Also, we discuss the relevance of this study to moral reasoning and to making ethical judgments using the empirical results of Arnold and Ponemon (1991).

Value Judgments Using Belief Functions

1. INTRODUCTION

The main purpose of this article is to show how the belief-function framework can be used to model human behavior when making decisions that involve value judgments. A rule of decision making has been presented in the form of a proposition. This proposition is simple and intuitive. It predicts correctly the observed behavior, even the behavior that is considered to be irrational under the expected utility theory (EUT).

The empirical work over several decades has shown that neither probability theory nor expected utility theory fully explains human decision making behavior. There are numerous situations where we run into problems if we use probability theory to model uncertainties encountered in the real world (see, e.g., Buchanan and Shortliffe 1984; Davis, Buchanan, and Shortliffe 1977; Einhorn and Hogarth 1986; Kahneman and Tversky 1979; and Shafer and Srivastava 1990; Srivastava 1993). In fact, the straight forward use of probability theory to model uncertainties in certain situations has often led to paradoxes. For example, the Ellsberg (1961, see also, Einhorn and Hogarth 1986) paradox stems from the fact that we are not able to model correctly the uncertainties involved in his situations using probabilities. Consider the following situation. Suppose you have two urns. Each urn contains 100 balls of red and black color. However, for one urn you are told that it has unknown proportions of red and black balls (a situation of complete ignorance), and for the other you are told that it contains 50 red and 50 black balls (a situation of complete knowledge). If we use probabilities to represent uncertainties involved in the above example then we are not able to distinguish between the two situations since we are forced to assign the same value 0.5 to the probability of picking a red ball, and to the probability of picking a black ball from either urn. This inability of probability theory to distinguish between the two situations led to the Ellsberg paradox as elaborated by Einhorn and Hogarth (1986). A belief-function treatment of the problem resolves the paradox (see, e.g., Srivastava 1994).

Schoemaker (1982) presents a good summary of the problems we have had in justifying some of the basic axioms of the expected utility theory (see also, Kahneman and Tversky 1979; Tversky

1969; Tversky and Kahneman 1986). Simon (1986) points out several situations where decisions are not made based on the utility maximization rule but on the context, taking into consideration all the factors of the environment in the decision process. Simon emphatically states under several situations that "Utility maximization is neither a necessary nor a sufficient condition for the conclusion that was reached (page S215)." He makes further comments on how decisions are procedurally made and how a neoclassical economist thinks that such a decision should be made:

> The rational person of neoclassical economics always reaches the decision that is objectively, or substantively, best in terms of the given utility function. The rational person of cognitive psychology goes about making his or her decisions in a way that is procedurally reasonable in the light of the available knowledge and means of computation (p. S211).

Simon's last sentence above is the essence of the decision making process using the belief-function framework. It aggregates all the information available in the context and then makes the decision. The present article deal with value judgments[1]. This domain of decision making is beyond the domain of utility theory. However, for the last several decades we have been using EUT to rationalize behavior and have failed in our efforts. As Simon (1986) puts it:

> Embracing a substantive theory of rationality has had significant consequences for neoclassical economics and especially for its methodology. Until very recently, neoclassical economics has developed no strong empirical methodology for investigating the processes whereby values are formed, *for the content of the utility function lies outside its self-defined scope* (emphasis added, p. S211).

We do not plan to repeat all the problems with EUT that are already discussed in the literature. Rather, we show how one can use the belief-function framework to make value judgments. We use extensively the empirical data of Tversky and Kahneman (1986, 1981) to demonstrate the usefulness of the belief-function framework for decision making. One situation is of special interest. When the same information is presented in two versions, positive and negative, the decision maker has two different responses which are inconsistent according to EUT. They violate the invariance axiom and thus have puzzled both the theorists and empiricists. However, as stated by Tversky and Kahneman (1986) this behavior has been observed repeatedly and even with the same group:

> On several occasions we presented both versions to the same respondents and discussed with them the inconsistent preferences evoked by the two frames. Many respondents expressed a wish to remain risk averse in the "lives saved" version and risk seeking in the "lives lost" version, although they also expressed a wish for their answers to be consistent (p. S260).

Prospect theory has been developed by Kahneman and Tversky (1979) to account for preferences that are anomalous in the normative theory. It is interesting to see that the rule proposed in this article for making decisions that involve value judgments predicts all the observed anomalous behaviors.

The remaining part of the paper is divided into five sections and an appendix. In Section 2, we discuss the belief function approach to decision making. In Section 3, we discuss a proposition for value judgments. In Section 4, we discuss more examples of value judgments. In Section 5, we discuss the relevance of this study to moral reasoning and ethical decisions. In Section 6, we provide a summary. Finally, in the appendix, we present an introduction to belief functions.

2. THE BELIEF-FUNCTION APPROACH TO DECISION MAKING

Recently, several approaches to making decisions under the belief-function framework have been proposed (see, i.e., Jaffray 1989, 1994; Nguyen and Walker 1994; Smets 1990a, 1990b; Strat 1990, 1994; and Yager 1990 for details). However, since the value judgments are outside the domain of utility theory, we do not want to use the expected utility theory approach of decision making here. Rather, we want to explore how the belief-function framework can help us understand human behavior in value judgments.

In general, the theory of belief functions provides a framework for aggregating evidence; the evidence that is available at the time of decision making. Thus, the belief function framework not only provides a way to model uncertainties encountered by decision makers but also provides a framework to integrate all that we know about the contextual situation in a specific decision problem. The contextual situation has been emphasized by Simon (1986) and many others, especially by researchers in AI (Artificial Intelligence).

There are three basic functions in the belief-function framework that are important in our context (see Appendix A for an introduction). The *basic probability assignment function*[2], m(A), represents the uncertainty mass assigned to an assertion A. This assignment may be a single assessment of the decision maker for the level of support on A or a combination of such assessments based on several items of evidence that are relevant to A. The belief function, Bel(A), represents the total belief or support for A considering all the evidence that directly relate to the assertion. The plausibility function, Pl(A), represents the maximum possible support that could be assigned to the assertion given the current knowledge of the situation. For example, Bel(A) = 0.3, and Pl(A) = 0.9 imply that we have direct evidence that A is true with a level of support of 0.3 but the maximum possible support that could be assigned to A is 0.9. It simply means that we have 0.6 (= 0.9 - 0.3) level of support unassigned that could be assigned to A if further evidence in its favor were collected.

3. A PROPOSITION FOR VALUE JUDGMENTS USING BELIEF FUNCTIONS

The proposition listed below determines the rule for decision making under uncertainty involving value judgments.

Proposition 1: When making a choice of an alternative from a group of alternatives based on the value judgment, the decision maker will select the alternative that has the highest belief in favor of the alternative being good/fair. If this rule results in the decision maker being indifferent among the alternatives, then he or she will choose the alternative that has the lowest belief of it being bad/unfair (i.e., the highest plausibility of it being good/fair).

This proposition is intuitive. We collect all that we know from the contextual situation about all the alternatives. Next, we determine the beliefs in favor of them being good/fair and select the one that has the highest belief in favor of it being good/fair. This process may involve combining both positive and negative items of evidence. The formal way to combine this information is to use Dempster's rule of combination (see, e.g., Shafer 1976; Shafer and Srivastava 1990; Srivastava and Shafer 1992). However, the decision maker uses his or her intuitive judgment to aggregate all the evidence in favor of and against the alternative to determine the total belief. In the case where we have the same belief for each alternative of being good/fair then we should use the beliefs for the alternatives being bad/unfair and select the one that has the least belief for it being bad/unfair, i.e., select the alternative that has the highest plausibility of it being good/fair. We demonstrate below how well this decision making rule fits with the empirical data.

We consider problems 5 and 6 in Tversky and Kahneman (1986; see also Tversky and Kahneman 1981, p. 453) to illustrate the process. The description of the situation is as follows:

> Problem 5 (N =152):
>
> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:
>
> If Program A is adopted, 200 people will be saved. [72%]
>
> If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28%]

The symbol N above represents the total number of subjects used in the study. The figures in parentheses next to program descriptions represent the percent of subjects that prefer that particular program.

The information about the two programs above has been presented in a positive way, because it describes how many people will be saved. Death incidences are implied in the statement, but are not made explicit. As a decision maker, when you receive this information and need to respond quickly you start processing this information as a positive piece of evidence in favor of the program being described. Before we can make the decision about our choice of the program, we need the belief for each program being good. The historical information that we have on each program does not directly provide these beliefs. We need to estimate them from the evidence. This situation is quite different from the situation of decision making with monetary payoffs under a game of chance. There, the decision maker knows all the probabilities and thus all the beliefs and plausibilities or can compute the beliefs and plausibilities given the information about the game (Srivastava 1994). In the present problem, we need to estimate them from the data as described below.

Based on the statistics, 200 people will be saved for sure in Program A. This is a positive item of evidence in support of the program being good. How strong is this support? We cannot exactly quantify it but let us say it supports the hypothesis that Program A is good with a level of belief 'a', i.e., $Bel(G_A) = a$ $(1 > a > 0)$. Similarly for Program B, we have information that if implemented it will save

600 people with 1/3 probability and none with 2/3 probability. This information is again interpreted as a positive item of evidence in favor of B but the level of support is not known precisely. Let us say, the support is some amount 'b', i.e., $Bel(G_B) = b$ $(1 > b > 0)$. Now, the question is which is higher of the two, is a > b or b > a? Let us reason it out and see which inequality is true. Based on what we know about the two programs, we see that in A we are sure to save 200 people and in B we have only 1/3 probability to save on average 200 people. This analysis suggests that Program A is better than B, i.e., $Bel(G_A) = a > Bel(G_B) = b$. Which means we will select Program A based on Proposition 1. This is what the empirical data show. Note that 72 percent chose Program A and 22 percent chose Program B (see Problem 5 above).

Let us consider Problem 6 of Tversky and Kahneman (1986):

Problem 6 (N =155):

If Program C is adopted 400 people will die. [22%]

If Program D is adopted there is 1/3 probability that no body will die, and 2/3 probability that 600 will die. [78%]

This information is presented in a negative way because it only talks about how many people will die if a particular program is adopted. Also, as in the previous case, it does not directly give us the beliefs for the programs being good or bad. However, we can estimate these beliefs based on what we know about the programs. For Program C, 400 people will die for sure. This immediately implies that it is related to the program being bad (not good). Let us say the belief assigned to the hypothesis that Program C is not good is 'c', i.e., $Bel(\sim G_C) = c$. Also, since there is no direct evidence in support of Program C being good, the corresponding belief is zero: $Bel(G_C) = 0$.

With a similar argument as above we can assign a zero value to the belief that Program D is good and a value of 'd' to the belief that it is not good: $Bel(G_D) = 0$, and $Bel(\sim G_D) = d$. The only question is whether c > d or d > c? Once we establish it, we can select the program. Let us again reason it out. Based on what we know about the two programs, we see that in C we are sure that 400 people will die but in B we have only 2/3 probability that on average 400 people will die. This information makes us feel that Program C is worse than program D, i.e., $Bel(\sim G_C) = c > Bel(\sim G_D) =$

d. Here, although we do not know how good are the two programs ($Bel(G_C) = 0$, and $Bel(G_D) = 0$) but we do know, in relative terms, how bad are the two programs. Thus, if we want to select a program to be adopted we will of course choose the one that is not as bad as the other. Which means we will choose Program D. Proposition 1 suggests the same choice and the empirical data support it too: 22% selected C and 78% selected D (see Problem 6 above).

However, if we had a well-informed individual, he or she will immediately recognize, even if he or she received only one version of the story, that if 200 people will be saved in Program A for sure then 400 will die too. Similarly for Program B, if 600 people will be saved with the probability 1/3 then 600 will die too with probability 2/3. If we analyze the above information altogether in the belief-function framework then we will have the following beliefs: $Bel(G_A) = a$, $Bel(\sim G_A) = c$, $Bel(G_B) = b$, and $Bel(\sim G_B) = d$. Let us assume that the same inequalities hold as we discussed earlier, i.e., $a > b$, and $c > d$. In probability theory, the above inequalities will not make sense. We can not express such feelings about the evidence in probability terms. This is one of the reasons why Einhorn and Hogarth (1986) proposed their ambiguity model.

How can we determine which alternative is better in the above situation? Let us use Proposition 1. It tells us to pick an alternative with the highest belief that it is good/fair. In the present case, we have $Bel(G_A) = a > Bel(G_B) = b$. Thus we should select Program A based on the proposition. Let us further investigate and see if A is the right choice. Consider now the worst scenario case and look at the plausibility (see Appendix A for the definition) for each program to be bad. The plausibility value for each program to be bad represents the maximum possible belief that could be assigned to it being bad. The plausibilities that the two programs are bad are: $Pl(\sim G_A) = (1 - a)$, and $Pl(\sim G_B) = (1 - b)$. We find that in the worst scenario case, Program A is not as bad as Program B because the plausibility of A being bad is less than the plausibility of B being bad: $Pl(\sim G_A) = (1 - a) < Pl(\sim G_B) = (1 - b)$. This inequality is true because $a > b$. Thus, based on the above argument we find Program A is to be overall better than Program B. This is what Proposition 1 suggests.

4. MORE EXAMPLES OF VALUE JUDGMENTS

In this section, we consider the rest of problems discussed by Tversky and Kahneman (TK) (1986) related to value judgment and show how Proposition 1 predicts all of the observed behaviors.

4.1 Problem 1 (Problem 1 of TK comes from McNeil et al. 1982)

> Survival frame: (N = 247)
>
> Surgery: Of 100 people having surgery 90 live through the post-operative period, 68 are alive at the end of the first year and 34 are alive at the end of five years. [82%]
>
> Radiation Therapy: Of 100 people having radiation therapy all live through the treatment, 77 are alive at the end of one year and 22 are alive at the end of five years. [18%]
>
> Mortality frame: (N = 336)
>
> Surgery: Of 100 people having surgery 10 die during surgery or the post-operative period, 32 die by the end of the first year and 66 die by the end of five years [56%].
>
> Radiation Therapy: Of 100 people having radiation therapy, none die during treatment, 23 die by the end of one year and 78 die by the end of five years [44%].

In order to explain the results, let us first consider the evidence based on the survival frame. Looking at the Surgery statistics we feel that all the evidence points towards the treatment being good. This can be expressed as a positive belief that the Surgery is good: $Bel_s(G_S) = a$ ($1 > a > 0$). There is no negative evidence, therefore the belief that Surgery is bad (not good) is zero: $Bel_s(\sim G_S) = 0$ (G stands for good, the subscript S of G stands for surgery, and the subscript s of Bel stands for survival frame). Similarly for Radiation Therapy, all the evidence again points towards the treatment being good and none to it being bad. These beliefs can be written as: $Bel_s(G_R) = b$ ($1 > b > 0$), and $Bel_s(\sim G_R) = 0$.

Now, how do we find which treatment is better unless we know the relative values of the two beliefs that the programs are good? Let us look at the factors involved in the above judgment: 90 versus all live during the treatment, 68 versus 77 are alive at the end of first year, and 34 versus 22 are alive at the end of five years. We see that 34 people in Surgery will be alive at the end of five years compared to only 22 in Radiation Therapy. However, the other two items of evidence are not as favorable for Surgery. Assuming that a treatment is better than another if more people that are treated by it live longer than those treated by the other, then the belief that Surgery is good is higher than the belief that Radiation Therapy is good: $Bel_s(G_S) = a > Bel_s(G_R) = b$. Therefore, the choice should be Surgery according to Proposition 1. This is what the data show: 82 percent chose Surgery and only 18 percent chose Radiation Therapy (see Problem 1 above).

On the other hand, when we consider the evidence produced in mortality frame, we find that all the evidence in the case of Surgery point to it being bad (not good). Therefore, we have a non-zero belief that Surgery is not good and a zero belief that Surgery is good: $Bel_m(G_S) = 0$, and $Bel_m(\sim G_S) = c$ $(1 > c > 0)$, where the subscript 'm' stands for mortality. For Radiation Therapy, we have both kinds of evidence; that none die during the treatment is a good sign, however, there are two other items of evidence that tell us that Radiation Therapy is not good. Thus, we have a mixed situation. We can express this feeling by non-zero beliefs for both the treatment being good, and the treatment being bad, i.e., $Bel_m(G_R) = b'$, $(1 > b' > 0)$ and $Bel_m(\sim G_R) = d$ $(1 > d > 0)$. The value of b' may be very small but it is not zero.

Proposition 1 suggests that Radiation Therapy is better because the belief that Radiation Therapy is good is higher than the belief that Surgery is good, i.e., $Bel_m(G_R) = b' > Bel_m(G_S) = 0$. The preferred choice based on this argument is Radiation Therapy. However, there would be many individuals who may evaluate the situation by considering both aspects of the evidence. In that case Surgery has a higher support for it being good because 34 instead of 22 people live more than five years. Thus, they will prefer Surgery over Radiation Therapy. We will have a mixed response in this case. This is evident from the empirical data. While there is a significant increase in response to Radiation Therapy from 18% in the survival frame to 44% in the mortality frame, still 56% percent preferred Surgery in the mortality frame. Thus, the data support our reasoning.

4.2 Problem 12 (TK)

> Problem 12 (N = 72). In the treatment of tumors there is sometimes a choice between two types of therapies: (i) a radical treatment such as extensive surgery, which involves some risk of imminent death, (ii) a moderate treatment, such as limited surgery or radiation therapy. Each of the following problems describes the possible outcome of two alternative treatments, for three different cases. In considering each case, suppose the patient is a 40-year-old male. Assume that without treatment death is imminent (within a month) and that

only one of the treatments can be applied. Please indicate the treatment you would prefer in each case.

Case 1

Treatment A: 20% chance of imminent death and 80% chance of normal life, with an expected longevity of 30 years. [35%]

Treatment B: certainty of a normal life, with an expected longevity of 18 years. [65%]

Case 2

Treatment C: 80% chance of imminent death and 20% chance of normal life, with an expected longevity of 30 years. [68%]

Treatment D: 75% chance of imminent death and 25% chance of normal life, with an expected longevity of 18 years. [32%]

Case 3

Consider a new case where there is a 25% chance that the tumor is treatable and a 75% chance that it is not. If the tumor is not treatable, death is imminent. If the tumor is treatable, the outcomes of the treatment are as follows:

Treatment E: 20% chance of imminent death and 80% chance of normal life, with an expected longevity of 30 years. [32%]

Treatment F: certainty of normal life, with an expected longevity of 18 years. [68%]

In all the above cases, again, we do not have the beliefs whether a treatment is good or bad. We need to assess them. For Treatment A in Case 1, the evidence that 20% will die supports that the treatment is not good (~G), but the evidence that 80% will live a normal life for 30 years supports that the treatment is good (G). Thus, we have the following beliefs: $Bel(G_A) = a$, $Bel(\sim G_A) = a'$. Both 'a' and 'a" are non-zero and, of course, less than 1. For treatment B in Case 1, the evidence that everyone lives a normal life of 18 years is a strong support for the treatment to be good. There is no direct evidence that the treatment is bad. For this case, we have the following beliefs: $Bel(G_B) = b$, $Bel(\sim G_B) = 0$. Since the decision is to be based on the belief of the treatment being good, we need to determine which belief is higher. Is a > b, or b > a? This is a difficult question. In one case we have 80% people live a normal life for 30 years, and in the other case we have everyone living a normal life for 18 years. However, if we give equal importance to every one's life then everyone living 18 years of normal life is better than only 80% living a normal life for 30 years. This argument we believe is pretty strong because we are talking about people's lives and every one should have equal opportunity to live. However, the situation may be reversed if we had everyone living only for few years, say five years, of normal life. Thus, in light of our discussion, $Bel(G_B) = b > Bel(G_A) = a$ and the choice will be Treatment B: The empirical data agree with our result: 65 percent chose Treatment B and 35 percent chose Treatment A.

We will not discuss Cases 2 and 3 in detail. However, if we follow a similar logic as above, we can predict the results. The only important point to note is that whether 20% people living a normal life for 30 years is better than 30% living only 18 years. This is again a difficult question. Many may believe that it is better for fewer people to live a longer life than more people to live a shorter life, especially when there is no guarantee who lives and who dies. With this argument we can explain the findings of Case 2. For Case 3, the empirical data can be explained using the same logic as used in Case 1. The decision maker has no control over whether the tumor is treatable or not treatable. However, when it is treatable he should prefer to choose alternative F based on the arguments presented earlier.

4. 3. Discounts and Surcharges

There is ample empirical evidence that when it comes to preferences people prefer things that are positive, good news, discounts, et cetera, over things that are negative, bad news, surcharges, et cetera, even if the information may have the same economic consequence on the well-being of the decision maker (Bazerman 1983; Kahneman, Knetsch, Thaler 1986; Schelling 1981; Thaler 1980). However, this behavior has been considered as irrational behavior because economic theories do not predict such a behavior. Based on the decision rule proposed in this article, such a behavior is normal; the decision maker makes the decision based what he or she knows about the situation, bringing in all cognitive capabilities to process the information. If the information appears to be negative then that is how he or she is going to process it. The decision maker may not be a sophisticated economist who looks at all the consequences even though they are not given in the information and then makes the decision.

Let us consider a simple example of a judgment about being fair or unfair. This example and the empirical data have been taken from Kahneman, Knetsch, and Thaler (1986). The subjects are answering the question whether the company is being unfair to employees.

> Positive information:
>
> A company is making a small profit. It is located in a community experiencing a recession with substantial unemployment and inflation of 12%. The company decides to increase salaries only 5% this year. [22%]
>
> Negative information:
>
> A company is making a small profit. It is located in a community experiencing a recession with substantial unemployment but no inflation. The company decides to decrease wages and salaries 7% this year. [62%]

In the case of positive information, we get the impression that the company is doing a favor to the employees by giving a 5% raise to the employees rather than laying off any one. However, some may interpret the information that they are giving only a 5% raise when the inflation is 12% as an item of evidence supporting unfairness of the company. In this situation, we may have a non-zero value for both the beliefs: the belief that the company is being fair (F), and the belief that the company is not being fair (~F): $Bel_P(F) = a$, and $Bel_P(\sim F) = b$ where $1 > a > 0$, and $1 > b > 0$ and the subscript of Bel stands for positive evidence. Some may argue for $b = 0$ which is acceptable.

Let us analyze the negative evidence. It tells us that the company is being unfair to the employees by decreasing the wages and salaries by 7%. This evidence has a direct support for the company being unfair and no support for the company being fair, which yields the following beliefs: $Bel_N(F) = 0$, and $Bel_N(\sim F) = c$. Based on Proposition 1 if we had to choose whether the company was being fair, we see that in the positive case we have support in favor of the company being fair, i.e., $Bel_P(F) = a > 0$, whereas we have no evidence to support that the company is being fair in the negative case ($Bel_N(F) = 0$). Thus, in the positive case the company is being fair, but in the negative case the company is being unfair. The empirical data support our argument (see the problem description above).

5. MORAL REASONING AND ETHICAL DECISIONS

Moral reasoning and ethical decisions are similar to value judgments; the decision maker takes into consideration all the information relevant to the context based on the individual values and beliefs (e.g., see the discussion in Arnold and Ponemon 1991). Kohlberg (1969) has developed a theory of cognitive moral reasoning and development. Arnold and Ponemon (1991) use Kohlberg's model to examine the internal auditor's perceptions of whistle-blowing behavior within the context of his or her level of moral reasoning. It is interesting to note that their experimental situation fits very well within the domain of this study. Their experiment involves judgments related to whether an individual discovering a management fraud will report the findings to higher authorities, i.e. blow the whistle. The subjects were 106 internal auditors from public and private sector. Since the subjects acted as third persons in making their judgments, the experimental design is beyond the domain of the traditional expected utility theory. Rather, the situation is similar to value judgments as discussed in Sections 3 and 4. Before providing a belief-function treatment of their problem we briefly describe one version of their case presented to the subjects with two levels of treatment (Arnold and Ponemon 1991, p. 7):

> Tim is an internal auditor with an organization that is a primary contractor for the U.S. Government. Tim recently completed an audit of a subsidiary business unit (ABC plant) which is completing significant (large dollar) contracts for various government agencies. The billings of the subsidiary have been audited previously and no major problems were detected. During the present audit, Tim discovered, within the subsidiary's billing system, a series of bogus (inflated or falsified) invoices to customers that have already been paid. Tim reported this finding to the director of internal audit. The director said that he would report it to authorities within the company. After a few days, the director told Tim, "Forget about it!"
>
> Tim demands further action. The director says,
>
> (Treatment level - affiliation) "If your findings are reported, the company will be forced to close the ABC plant." Two of Tim's very close friends are employees in the plant.
>
> (Treatment level–penalty) "If you pursue this, you will be fired!"

The subjects were asked to make a prediction as to whether Tim *did* or *did not* blow the whistle. This question is similar to value judgment; the subjects were asked to make a value judgment whether the individual in the case acted ethically (E) or not (~E). We do not plan to discuss all their cases in the present article. However, in order to illustrate the belief-function treatment of such a problem, we consider the scenario where the individual discovering the fraud is an internal auditor and faces two situations if blows the whistle: (1) two friends lose their jobs, or (2) Tim gets fired.

The decision maker (DM, the subject) has to make the judgment whether Tim will act ethically (E), i.e., blow the whistle under situation 1 or situation 2. A belief-function treatment of the problem follows. Let us assume that DM does not know anything else on Tim in terms of his moral and ethical values except the information given above about the two situations. In situation 1, the knowledge that Tim's two friends will lose their jobs if he reports the fraud provides direct evidence that Tim will not report the fraud, that is, he will act unethically (~E). Thus, this evidence provides a positive non-zero belief, say 'a', that Tim will act unethically, i.e., $Bel_1(\sim E) = a$ $(1 > a > 0)$ where the subscript on Bel represents the first situation. Since there is no evidence in support of E, the belief that Tim will act ethically (E) is zero, i.e., $Bel_1(E) = 0$.

In situation 2, the evidence that Tim will be fired provides direct evidence that he will not report the fraud, i.e., he will not act ethically. Thus, the belief that Tim will act unethically is again non-zero, say an amount b, i.e., $Bel_2(\sim E) = b$, and again since there is no evidence in favor of E, $Bel_2(E) = 0$. The question is which alternative will the DM pick? The answer depends on the relative values of a and b. Let us analyze the situation. It seems reasonable to assume that without any other information about Tim's background on his moral and ethical values, he will be more willing to act unethically when it comes to losing his own job in comparison to having his friends lose their jobs. This reasoning suggests that $a < b$. Thus, using Proposition 1, we find that DM will pick situation 1. This is because the belief that Tim will act ethically is zero in both situations ($Bel_1(E) = Bel_2(E) = 0$) and the belief that he will act unethical in situation 1 is lower than the belief of acting unethical in situation 2 ($Bel_1(\sim E) = a < Bel_2(\sim E) = b$). This implies that it is more plausible that Tim will act ethical in situation one than in situation 2. In other words, DM will suggest that it is more plausible that Tim will report the fraud in situation 1 than in situation 2. This is what Arnold and Ponemon found: the subjects perceived whistle-blowing to be less likely under conditions of job termination than under plant closure and employee layoffs (1991, p. 10). One can analyze the other cases of Arnold and Ponemon in a similar fashion.

6. SUMMARY

In summary, we have described a rule for making decisions that involve value judgments using belief functions. The rule is presented in the form of a proposition which is simple and intuitive. We have shown how well the rule predicts the observed behavior by using the empirical data of Tversky and Kahneman (1986, 1981). It is interesting to see that the proposed rule predicts correctly even those behaviors that are considered irrational under EUT. Also, we have discussed, using the empirical results of Arnold and Ponemon (1991), how the present work can be used to model behaviors when making ethical judgments.

FOOTNOTES

1. Moral reasoning and making ethical judgments are similar to value judgments; they all depend on the contextual information available to the decision maker.

2. The *basic probability assignment function* (m-values) in the belief-function framework is similar to probability function with one difference that m-values are defined, in general, on all possible subsets of the frame of interest whereas probabilities are defined only on singletons (see the appendix for details). In the case where non-zero m-values exist only for the singletons, the belief-function framework reduces to the Bayesian framework (Shafer and Srivastava 1992).

3. This appendix is taken from Srivastava (1993).

4. In the case of n elements in the frame, we will have $P(a_i) = 0$, and $\sum_{i=1}^{n} P(a_i) = 1$, where a_i represents the ith element of the frame.

5. For a frame of n elements, we will have, in general, m-values for each individual elements, each set of two elements, each set of three elements, and so on, to the m-value for the entire frame. All such m-values add to one, i.e., $\sum_{A \subseteq \Theta} m(A) = 1$, where A represents all the proper subsets of the frame Θ. The m-value for the empty set is zero.

REFERENCES

Arnold, Sr., D. F., and L. A. Ponemon. 1991. Internal Auditors' Perceptions of Whistle-Blowing and the Influence of Moral Reasoning: An Experiment. *Auditing: A Journal of Practice and Theory,* Vol. 10, No. 2 (Fall):1-15.

Bazerman, M.H. 1983. Negotiator Judgment. *American Behavioral Scientist,* vol. 27:211-28.

Buchanan, B.G., and E.H. Shortliffe. 1984. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*t. Reading, Massachusetts: Addison-Wesley.

Davis, R., B. Buchanan, and E.H. Shortliffe. 1977. Production Rules as a Representation for a Knowledge-Based Consultation System. *Artificial Intelligence,* vol. 8:15-45.

Einhorn, H. J. and R. M. Hogarth. 1986. Decision Making under Ambiguity. *The Journal of Business,* Vol. 59, No. 4, Pt. 2 (October):S225-S250.

Ellsberg, D. 1961. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics,* Vol. 75:643-69.

Jaffray, J-Y. 1994. Dynamic Decision Making with Belief Functions. *Advances in the Dempster-Shafer Theory of Evidence,* edited by R.R. Yager, M. Fedrizzi, and J. Kacprzyk. New York, NY: John Wiley and Sons.

Jaffray, J-Y. 1989. Utility Theory for Belief Functions. *Operations Research Letters,* Vol. 8:107-12.

Kahneman, D., J.L. Knetsch, and R.H. Thaler. 1986. Fairness and the assumptions of economics. *The Journal of Business,* Vol. 59, No. 4, Pt. 2 (October):S85-S300.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica,* vol. 47, no. 2:263-91.

Kohlberg, L. 1969. Stages and Sequences: The Cognitive developmental approach to Socialization, *Handbook of Socialization Theory and Research* edited by D. Goslin. Chicago: Rand McNally.

McNeil, B.J., S.G. Pauker, H.C. Sox, Jr., A. Tversky. 1982. On the Elicitation of Preferences for Alternative Therapies. *New England Journal of Medicine,* vol. 306:1259-62.

Nguyen H.T., and Walker, E.A. 1994. On Decision Making Using Belief Functions. *Advances in the Dempster-Shafer Theory of Evidence,* edited by R.R. Yager, M. Fedrizzi, and J. Kacprzyk. New York, NY: John Wiley and Sons.

Schelling, T.C. 1981. Economic Reasoning and the Ethics of Policy. *Policy Interest,* vol. 63: 37-61.

Schoemaker, P. J. H. 1982. The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. *Journal of Economic Literature,* Vol. XX (June):529-563.

Shafer, G. 1976. *A Mathematical Theory of Evidence.* Princeton, N.J.: Princeton University Press.

Shafer, G. and R.P. Srivastava. 1990. The Bayesian And Belief-Function Formalisms: A General Perspective for Auditing. *Auditing: A Journal of Practice and Theory* (Supplement): 110-148.

Simon, H.A. 1986. Rationality in Psychology and Economics. *The Journal of Business,* Vol. 59, No. 4, Pt. 2 (October):S209-S224.

Smets, P. 1990a. The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 5 (May):447-458.

Smets, P. 1990b. Constructing the Pignistic Probability Function in a Context of Uncertainty. *Uncertainty in Artificial Intelligence 5,* ed. by M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer. North-Holland: Elsevier Science Publishers B.V.

Srivastava, R.P. 1994. Decision Making Under Ambiguity: A Belief-Function Perspective. Working Paper No. 262, School of Business, The University of Kansas, Lawrence (June).

Srivastava, R.P. 1993. Belief Functions and Audit Decisions. *Auditors Report,* Vol. 17, No. 1 (Fall):8-12.

Srivastava, R.P., and G. Shafer. 1992. Belief-Function Formulas for Audit Risk. *The Accounting Review* (April):249-283.

Strat, T.M. 1990. Decision Analysis Using Belief Functions. *International Journal of Approximate Reasoning,* vol. 4, no. 5:6.

Strat, T.M. 1994. Decision Analysis Using Belief Functions. *Advances in the Dempster-Shafer Theory of Evidence,* edited by R.R. Yager, M. Fedrizzi, and J. Kacprzyk. New York, NY: John Wiley and Sons.

Thaler, R.H. 1980. Towards a Positive Theory of Consumer Choice. *Journal Economic Behavior and Organization,* vol. 1:39-60.

Tversky, A. 1969. Intransitivity of preferences. *Psychological Review,* vol. 76:105-10.

Tversky, A, and D. Kahneman. 1986. Rational Choice and the Framing of Decisions. *The Journal of Business,* Vol. 59, No. 4, Pt. 2 (October):S251-S278.

Tversky, A, and D. Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science*, vol. 211:453-58.

Yager, R.R. 1990. Decision making Under Dempster-Shafer Uncertainties. *Technical Report MII-915,* Iona College, New Rochelle, NY.

APPENDIX A

THE BELIEF-FUNCTION FRAMEWORK[3]

Belief functions are not new; they have antecedents in the seventeenth century work of George Hooper and James Bernoulli (Shafer 1976). The works by Dempster in the 1960's and by Shafer in the 1970's make the current form of the belief-function formalism known as Dempster-Shafer theory of belief functions. We will give only the basics of belief functions. Since the present paper does not deal with combination of evidence, we will not give Dempster's rule of combination. Interested readers are suggested to see Shafer (1976) for details.

The basic difference between probability theory and the belief-function framework is in the assignment of uncertainties to a set of mutually exclusive and exhaustive states or assertions under consideration (we will call this set a *frame* and represent it by the symbol Θ). In probability theory, we assign uncertainty to each individual element of the frame and call it the probability of occurrence of the element. The sum of all these probabilities equals one.

Let us consider an auditing example. The accounts receivable balance is not materially misstated (ar) and it is materially misstated (~ar) are the two assertions representing a mutually exclusive and exhaustive set. Here the frame consists of the two elements[4]: Θ = {ar, ~ar}. In probability theory, we will assign probability to each element of the frame, i.e., P(ar) = 0, and P(~ar) = 0. Also, we know that P(ar) + P(~ar) = 1. In the belief-function formalism, uncertainty is not only assigned to the single elements of the frame but also to all other proper subsets of the frame and to the entire frame. We call these uncertainties m-values or the *basic probability assignment function.*

m-values (The Basic Probability Assignment Function)

Similar to probabilities, all these m-values add to one. For the example considered above[5], we will have m(ar) = 0, m(~ar) = 0, and m({ar, ~ar}) = 0, and, m(ar) + m(~ar) + m({ar, ~ar}) = 1. Let us assume that the auditor has performed analytical procedures relevant to the accounts receivable balance and finds no significant difference between the recorded value and the predicted value. Based on this finding, he feels that the recorded value appears reasonable and is not materially misstated. However, he does not want to put too much weight on this evidence. He feels he can assign a small level of assurance, say 0.3 on a scale of 0-1, that the account is not materially misstated. We can express this feeling in terms of m-values as: m(ar) = 0.3, m(~ar) = 0, and m({ar, ~ar}) = 0.7. The belief function interpretation of these m-values is that the auditor has 0.3 level of support to 'ar', no support to '~ar', and 0.7 level of support remains uncommitted which represents ignorance.

However, if we had to express the above feelings in terms of probabilities then we get into problems because we will assign P(ar) = 0.3 and P(~ar) = 0.7 which implies that there is a 70 percent chance that the account is materially misstated, but we know that this is not what the auditor is trying to say. The auditor has no reason to believe that the account is materially misstated. Thus, we can use m-values to express the basic judgment about the level of support or assurance the auditor obtains from an item of evidence for an assertion. An example of a negative item of evidence which will have a direct support for '~ar' would be the following set of inherent factors: (1) in the prior years the account has had major problems, and (2) there are economic reasons for management to misstate the account. In such a case we can express the auditor's feelings as m(ar) = 0, m(~ar) = 0.2, and m({ar, ~ar}) = 0.8, assuming that the auditor feels a low, say 0.2, level of support for '~ar'.

The auditor can express a mixed-type of evidence in terms of m-values without any problems. For example, consider that the auditor has accumulated several environmental factors, some in support of and some against the assertion that the accounts receivable balance is not materially misstated. He assesses that there is a moderate, say 0.4, level of support in favor of the assertion and a low level of support, say 0.1, for its negation, and feels that he cannot assign the remaining 0.5 level of support to any particular state. We can expressed this feeling as: m(ar) = 0.4, m(~ar) = 0.1, and m({ar, ~ar}) = 0.5. In probability theory, we cannot express such a feeling.

Belief Functions

So far we have talked about probabilities and m-values. Let us now define the belief function, Bel(A) for a subset A of elements of the frame. Bel(A) represents the total belief in A. This belief will be more than m(A). Actually, Bel(A) is equal to m(A) plus sum of all the m-values for the set of elements that are contained in A. In terms of symbols:

$$Bel(A) = \sum_{B \subseteq A} m(B) .$$

By definition, belief in the empty set is zero.

Let us consider an example to illustrate the definition of belief functions. Suppose, you have a friend who lives on the East Coast in the New Jersey area. The only contact you have with him is through greeting cards that he sends you periodically with no return address. You want to find the belief that your friend lives in New Jersey. After looking through all the cards you have received over the years, you can identify the following post-office seals marked on the cards: 10% of the cards are marked North Brunswick, 20% East Brunswick, 10% Philadelphia, and 15% Newark. Twenty-five percent of the cards have only the Brunswick part legible which means you cannot determine from what part of Brunswick the card was mailed. For the remaining 20 percent, nothing is legible on the seals. These numbers can be interpreted as non-zero m-values for different subsets of the frame that your friend lives somewhere on the East Coast near New Jersey. Based on just this evidence, you wish to form your total belief that the friend lives in New Jersey. This belief will be the sum of the m-values that he lives in North Brunswick, East Brunswick, Brunswick, and Newark. For this example, the belief is 0.70. Similarly, the belief that the friend lives in Brunswick which includes North Brunswick and East Brunswick will be 0.55 (10% North Brunswick, 20% East Brunswick, 25% Brunswick).

Going back to our auditing example of analytical procedures, the auditor's assessment of the level of support in terms of m-values was: m(ar) = 0.3, m(~ar) = 0, and m({ar, ~ar}) = 0.7. Based on analytical procedures alone, the belief that the account is not materially misstated is 0.3 (i.e., Bel(ar) = 0.3) and no support that the account is materially misstated (Bel(~ar) = 0). In general, a zero belief in the belief-function formalism means that there is no evidence to support the proposition. In other words, a zero belief in a proposition represents lack of evidence. In contrast, a zero probability in probability theory means that the proposition cannot be true which represents an impossibility. Also, one finds that beliefs for 'ar' and '~ar' do not necessarily add to one, i.e., Bel(ar) + Bel(~ar) = 1, whereas in probability, it is always true that P(ar) + P(~ar) = 1.

Belief functions differ from probabilities in representing ignorance. In probability theory, we represent ignorance by assigning equal probability to all the outcomes or elements of the frame. In the belief-function framework, we represent ignorance by assigning an m-value of one to the entire frame and an m-value of zero to all its proper subsets. The belief-function formalism becomes the Bayesian formalism when non-zero m-values exist only for single elements of the frame. In such a case, m-values become probabilities, i.e., $m(a_i) = P(a_i)$, and Dempster's rule in the belief-function formalism becomes Bayes' rule in the probability theory (Shafer 1976).

Plausibility Functions

By definition, the plausibility of A is equal to one minus the belief in ~A, i.e., Pl(A) = 1 - Bel(~A) where ~A represents the set of elements that are not in A. Intuitively, the plausibility of A is the degree to which A is plausible given the evidence. In other words, Pl(A) is the degree to which we do not assign belief to its negation ~A.

In our example of analytical procedures, we have Bel(ar) = 0.3, Bel(~ar) = 0. These values yield the following plausibility values: Pl(ar) = 1, and Pl(~ar) = 0.7. Pl(ar) = 1 indicates that 'ar' is maximally plausible since we have no evidence against it. However, Pl(~ar) = 0.7 indicates that if we had no other items of evidence to consider then the maximum possible risk that the account is materially misstated would be 0.7, even though we have no evidence that the account is materially misstated (Bel(~ar) = 0).