Attribute Weighting Methods and Decision Quality in the Presence of Response Error: A Simulation Study

Jianmin Jia *, Gregory W. Fischer †, and James S. Dyer ‡

*Department of Marketing
Faculty of Business Administration
Chinese University of Hong Kong
Shatin, N. T., Hong Kong

† Center for Decision Studies
The Fuqua School of Business
Duke University, Durham, NC 27706

‡ Department of Management Science & Information Systems
Graduate School of Business
The University of Texas at Austin, Austin, TX 78712

May 1993

Revised April 1997

Journal of Behavioral Decision Making, in press

Attribute Weighting Methods and Decision Quality in the Presence of Response Error: A Simulation Study

Abstract

This paper uses a simulation approach to investigate how different attribute weighting techniques affect the quality of decisions based on multiattribute value models. The weighting methods considered include equal weighting of all attributes, two methods for using judgments about the rank ordering of weights, and a method for using judgments about the ratios of weights. The question addressed is: How well does each method perform when based on judgments of attribute weights that are unbiased but subject to random error? To address this question, we employ simulation methods. The simulation results indicate that ratio weights were either better than rank order weights (when error in the ratio weights was small or moderate) or tied with them (when error was large). Both ratio weights and rank order weights were substantially superior to the equal weights method in all cases studied. Our findings suggest that it will usually be worth the extra time and effort required to assess ratio weights. In cases where the extra time or effort required is too great, rank order weights will usually give a good approximation to the true weights. Comparisons of the two rank-order weighting methods favored the rank-order-centroid method over the rank-sum method.

KEY WORDS: attribute weights, multiattribute utility, decision quality, rank-order weights, preference uncertainty, response error, simulation, uncertain weights.

The standard decision analysis approach to assessing a multiattribute value or utility model assumes that preferences satisfy the usual axioms of rational choice (Keeney & Raiffa, 1976). One implication of these axioms is that the decision maker's preferences are known with certainty. Although external events are often viewed as uncertain, and thus can be assigned probabilities, the value or utility of each possible outcome is represented by a single point estimate, not by a probability distribution of possible values. In practice, however, decision makers are frequently uncertain about the relative desirability of different decision outcomes. They may even be uncertain as to whether they prefer one outcome to another. Two manifestations of preference uncertainty are evident in behavioral data – inconsistent preference judgments over time (response error) and delay in making judgments and choices (Fischer, 1976; Fischer, Luce, & Jia, 1996; Laskey & Fischer, 1987; Tversky & Shafir, 1992). In this article, we investigate how preference uncertainty may affect the assessment of weights in multiattribute value models.

In multiattribute decision analysis, preference assessment involves specifying a set of attributes describing the value-relevant properties of outcomes, assessing single-attribute value functions over the levels of each attribute, and assessing attribute weights (scaling constants) that govern the rate of substitution of value across attributes (Keeney & Raiffa, 1976). Preference uncertainty may adversely affect all three aspects of preference assessment. Our emphasis is on how preference uncertainty, manifested as response error in assessing weights, affects the accuracy of choice. Response error in weight assessments implies that weights elicited at one moment will differ from those elicited at another. If so, which weights should one use as the true indicator of the decision maker's preferences? One pragmatic approach to this problem is to rely on the well-documented robustness of linear models as a defense against it. According to the robustness argument, small changes in weights usually do not matter, so they can be ignored (Dawes & Corrigan, 1974; von Winterfeldt & Edwards, 1986).

A second approach is to rely on weighting methods that require less precise information regarding the decision maker's tradeoffs. For example, Dawes and Corrigan (1974) suggested that *equal weighting methods* will frequently produce decisions that are at least as good in quality while sparing one the effort (expense) of estimating a precise set of numerical weights. Others have argued that *rank-order weighting methods,* which rely only on ordinal information about attribute importance, provide a superior approach. There are several arguments favoring rank order methods (Barron & Barrett, 1996; Stilwell, Seaver, & Edwards, 1981). First, ranking methods are clearly easier and possibly more reliable than methods that require judgments sufficient to specify ratios of weights (Eckenrode, 1965). Second, respondents may be unwilling to provide more than ordinal information (Kirkwood & Sarin, 1985). Third, if the decision is being made by a group, they may be able to agree on the ranking of attributes but not on precise weights (Kirkwood & Sarin, 1985). Finally, evaluations generated by rank order methods correlate more highly with those generated by more precise numerical methods than do evaluations generated by the equal weights method (Barron & Barrett, 1996; Stillwell, Seaver, & Edwards, 1981).

The goal of this research is to investigate how preference uncertainty affects the relative performance of precise and approximate weight assessment methods. We use Monte Carlo simulation methods to investigate the following question: *What is the quality of the decisions produced by precise and approximate weight elicitation methods when the decision maker's judgments of attribute weights are subject to random response error?* We consider four approaches to weight elicitation and specification: ratio-scale weights, rank-sum weights, rank-order-centroid weights, and equal weights.

It is important to note that all approaches to weight elicitation result in weights defined on a ratio scale. This is a necessary property of weights in both riskless multiattribute value functions and risky multiattribute utility functions (Keeney & Raiffa, 1976).[1] However, weight assessment methods differ in terms of *the kind of information they preserve from the decision maker's judgments*. Some assessment methods (e.g., swing weights, corner point value assessments, and tradeoff weights) preserve ratio scale properties of the decision maker's judgments. Thus, we refer to these *as ratio weights methods*. In contrast, the rank-sum and rank-order-centroid methods preserve only *ordinal* properties of the decision maker's judgments. Thus, we refer to methods of this type as *rank-order methods*. With these methods, ratios among weights are established by applying a standard transformation of ranks into ratios. (The rank-sum and rank order centroid methods differ only in the ranks-to-ratios transformation that they use.) Finally, the *equal weights method* preserves only *categorical* information from a decision maker's judgments – either an attribute matters or it does not. Ratio properties are established by arbitrarily giving all attributes equal weight.

We begin this article by discussing a simple example that illustrates the four approaches to weight specification method and shows how they may produce different decisions. Next, we briefly review previous hypotheses and research findings regarding the relative efficacy of different attribute weighting methods. Then we propose a statistical model of preference uncertainty in which attribute weights are represented by a Dirichlet multivariate density function. We use this model to conduct two Monte Carlo simulation experiments in which we explore the quality of the decisions produced by the different methods. We conclude this article by summarizing the implications of our findings for both behavioral and prescriptive decision research.

The issues addressed in this article complement those addressed in a large body of recent behavioral research investigating *systematic response mode biases* in multiattribute preference assessment; for example, the prominence effect (Fischer & Hawkins, 1993; Tversky, Sattath, & Slovic, 1988) and the scale compatibility bias (Delquie, 1993; Fischer & Hawkins, 1993; Slovic, Griffin, & Tversky, 1990; Tversky, Sattath, & Slovic, 1988). Such systematic biases clearly complicate the task of identifying a useful measure of a decision maker's preferences. In this article, we show that random error arising from preference uncertainty also has important implications for how one should assess multiattribute preference models.

Attribute Weighting Methods: An Example and Brief Review

Consider a young college graduate who is choosing among a set of cities in which to live. The first column of Exhibit 1 displays the set of attributes relevant to the student's choice. For simplicity, we assume that this decision maker has constructed a set of attributes reflecting the attainment of each of these objectives and that each city under consideration has been assigned a measurable value[2] on each attribute (see the rightmost 3 columns of Exhibit 1). In the discussion that follows we first describe each weighting method and the rationale underlying it. Then we illustrate how to apply the weighting method to our example. We also discuss how the weighting methods differ in terms of the weights they assign, and show that these differences can lead to different rank orderings of realistic decision alternatives. Finally, we briefly review prior research on the efficacy of the four methods.

/Insert Exhibit 1 about here./

Ratio Scale Weighting Methods

Additive value and utility models have been used extensively in both prescriptive (Dyer & Sarin, 1979; Keeney & Raiffa, 1976) and descriptive studies of multiattribute decision making (Fischer, 1976; Tversky et al., 1988). The simplicity of additive models and the fact that they provide good approximations to many types of preference structures make them a very useful decision tool. The standard decision analysis approach to assessing an additive multiattribute value or utility model involves two types of assessments. First, assess a value or utility function over each attribute. Second, assess a set of weights that will be used in forming the weighted sum of the single-attribute values or utilities (Keeney & Raiffa, 1976).

The general form of an additive multiattribute value model can be represented as:[3]

$$V(x_1, x_2, \ldots, x_m) = \Sigma_i w_i \, v_i(x_i) \qquad (1)$$

where x_i is an attribute value for attribute i; $v_i(x_i)$ is a single-attribute value function defined over levels of x_i and $0 \leq v_i(x_i) \leq 1$; and w_i is a scaling constant that weights the value function for attribute i, $0 \leq w_i \leq 1$, $i = 1, 2, \ldots, m$, and $\Sigma_i w_i = 1$.

The weights in Eq (1) are assumed to reflect the relative importance of moving an attribute from its worst to best level and thus are defined on a ratio scale.[4] Many approaches to obtaining numerical weights are possible, including direct tradeoff methods, direct judgments of swing weights, and lottery-based utility assessments (see Keeney & Raiffa, 1976; von Winterfeldt & Edwards, 1986). Whichever method is used, the decision maker is asked to make judgments that provide ratio-scale information regarding the relative importance of attributes.

For example, with the swing weight method (von Winterfeldt & Edwards, 1986), the decision maker begins by rank ordering attributes in terms of their associated value ranges. Assuming that each attribute is at its worst possible level, the decision maker is asked which attribute she would most prefer to change from its worst to its best level. The attribute chosen has the most important value range. Call this attribute X_1 and arbitrarily assign it a weight of 100. Next the decision maker is asked which attribute she would next most prefer to change from its worst to best level. Call this X_2. To quantify the relative value ranges of X_1 and X_2, the decision maker next assigns a relative importance weight between 0 and 100 to X_2 (e.g., 70). Proceeding in this fashion, the decision maker rank orders the attributes and assigns relative importance weights to their value ranges (see the third column of Exhibit 1). The final step in the swing weight procedure is to normalize the relative importance weights, $r_1, \ldots, r_m$, to obtain the normalized swing weights $w_1, \ldots, w_m$. The standard normalization is

$$w_i = \frac{r_i}{\sum_{k=1,m} r_i}\ , i = 1, 2, \ldots, m, \qquad (2)$$

where $0 \le w_i \le 1$ and $\sum_i w_i = 1$. These normalized weights are presented in the fourth column of Exhibit 1.

To summarize, a variety of different methods may be used to obtain ratio-scale assessments of attribute weights. They can be assessed using the swing weight method, inferred from direct tradeoffs, or inferred from direct judgments of the value of appropriately selected corner outcomes (see Keeney & Raiffa, 1976). We will refer to the weights that result from any of these assessment methods as *ratio weights* because they preserve ratio scale properties of the decision maker's judgments.

Approximate Weighting Methods

Ideally, the weights in the additive value model (1) should reflect precise quantitative information regarding the decision maker's preferences (e.g., see Keeney and Raiffa, 1976). However, accurate determinations of attribute weights are difficult to obtain in practice because assessed weights are always subject to response error. This has led to a variety of suggestions regarding how to assess approximate attribute weights for the additive model, including equal weights (Dawes and Corrigan, 1974; Einhorn and Hogarth, 1975) and rank-order weights (Stillwell, Seaver and Edwards, 1981; Barron and Barrett, 1996). Using approximate weights can greatly simplify a decision procedure and may also result in satisfactory decision quality.

If a decision maker is uncertain as to her tradeoffs, then one may think of attribute weights as being sampled from a distribution of possible weights. The distribution of the weight w_i may be thought of as resulting from the addition of a random error component to a *true weight* – i.e., the weight that describes the decision maker's preferences when error is averaged out.[5] In general, an expected value can be used for approximating an uncertain quantity. This logic also applies when the uncertain quantity is the weight associated with an attribute. Similar approaches have been taken in analytical studies by Kleinmuntz (1990) and Ravinder, Kleinmuntz, and Dyer (1988) on the effects of error on decompositions of subjective probability and by Ravinder and Kleinmuntz (1991) on the effects of error on additive decompositions of utility.

Ideally, a quantitative method for assessing weights will yield unbiased estimates of the expected values of the underlying distributions of weights. Similarly, a good method for approximating weights should yield values close to the expected values of the weight distributions. This provides a theoretical basis for evaluating alternative approaches to weight construction.

The equal weights method. As we just argued, choice of a weighting method depends on one's knowledge of the underlying distributions of true weights. If one has no information about the true weights, then the true weights could be represented as a uniform distribution on the unit m-simplex, defined by the domain of $0 \le w_i \le 1$, $i = 1, 2,..., m$, and $w_i = 1$. (In the following, we will simply refer to this as the simplex of weights).

A simplex is a closed geometric object; for instance, a line segment (which has two vertices) or a triangle (which has three). With two attributes, the no-information simplex of weights is the set of points lying on the line segment whose vertices are (1, 0) and (0, 1). All points on this line have coordinates that sum to one – e.g., (1/3, 2/3). This is the "unit 2-simplex." If we have no knowledge about the weights, except that they sum to 1, then our knowledge can be represented by a uniform probability density function over this line. The expected value of this distribution is the centroid (center of mass) of the line, whose coordinates are (1/2, 1/2). More generally, under the assumption of no information about weights, the expected value of the weights distribution is the equal weights vector defined by

$$w_i = 1/m, \quad i = 1, 2, \ldots, m \quad . \tag{3}$$

The method requires minimal knowledge of the decision maker's priorities and minimal input from the decision maker. All that is required is the judgment that an attribute matters. If it does, it receives equal weight to all other attributes considered. The equal weights vector for our city choice example is shown in column 7 of Exhibit 1. Not surprisingly, the equal weights differ substantially from the ratio weights. The equal weights method was popularized by an influential article by Dawes and Corrigan (1974), who argued that this method often produced decisions nearly as good as those based on optimal (e.g., least squares) attribute weights.

Rank-order centroid (ROC) weights. Barron and Barrett (1996) investigated an alternative assumption regarding knowledge of the decision maker's preferences. If we know the rank order of the true weights, but have no other quantitative information about them, then we may assume that the weights are uniformly distributed on the simplex of rank-order weights, $w_{(1)} \geq w_{(2)} \geq \ldots w_{(m)} \geq 0$, where $\Sigma_i w_{(i)} = 1$, i is the rank position of $w_{(i)}$, and m is the number of attributes.

For example, if $m = 2$, then $w_{(1)} \geq w_{(2)}$ implies that $0.5 \leq w_{(1)} \leq 1$. If we know nothing else about $w_{(1)}$, then a plausible assumption is that the probability distribution of $w_{(1)}$ is uniform between 0.5 and 1. Taking expected values, we obtain $E(w_{(1)}) = 0.75$, which implies that $E(w_{(2)}) = .25$.[6] Barron and Barrett (1996) generalized this argument to $m > 2$ attributes, showing that the expected values of the true weights can be calculated using

$$w_{(i)} = \frac{1}{m}\sum_{k=i}^{m}\frac{1}{k}, \quad i = 1, 2, \ldots, m \ . \qquad (4)$$

They call these rank-order centroid (ROC) weights because these weights reflect the centroid (center of mass) of the simplex defined by the ranking of the attributes. For example, with $m = 3$, $w_{(1)} = (1+1/2+1/3)/3 = 11/18$; $w_{(2)} = (1/2+1/3)/3 = 5/18$; and $w_{(3)} = (1/3)/3 = 2/18$.

The application of the ROC formula (4) to $m > 3$ attributes is illustrated in Exhibit 1. The ROC weights for the city choice problem are quite similar to the ratio weights in this example. In this case, however, the ROC weights are "steeper," assigning relatively greater weight to the more important attributes. It should be noted, however, that the ROC weights depend only on the ranking of attributes in the ratio weights, not on the quantitative values. Thus, it is possible to construct sets of ratio weights that are flatter, steeper, or exactly equal to the ROC weights.

Rank-sum (RS) weights. Einhorn and McCoach (1977) and Stillwell, Seaver and Edwards (1981) investigated the rank-sum weighting method. With this method, the decision maker ranks attributes in terms of their relative importance (value range), then each attribute is weighted in proportion to its position in the rank order. That is

$$w_{(i)} = \frac{m+1-i}{\sum_{k=1,m} k} = \frac{2(m+1-i)}{m(m+1)}, \quad i = 1, 2, ..., m \tag{4}$$

where (i) is the rank position of attribute i, and where $0 \le w_{(i)} \le 1$ and $\sum_i w_{(i)} = 1$.

The normalized rank-sum weights for the city choice problem are displayed in Exhibit 1. The rank-sum weights are much flatter than the ratio weights in Exhibit 1. Of course, the relationship of rank-sum weights to directly assessed ratio weights depends on the ratio weights themselves. It is possible to construct examples in which ratio weights are flatter, steeper, or precisely equal to the rank-sum weights.

The rank-sum weights are also much flatter than the ROC weights. In general, rank-sum weights are flatter than ROC weights. For instance, with 3 attributes, ROC = (0.611, 0.278, 0.111) and rank-sum = (0.500 , 0.333, 0.167). A choice between rank-sum and ROC weights depends in part on one's beliefs about the steepness of the true weights guiding a decision maker's preferences. The greater the concentration of value in the first few attributes, the more attractive the ROC method.

The ROC approach to rank-order weights has a clear statistical basis and interpretation (Barron & Barrett, 1996) whereas previous uses of rank-sum weights have taken a more heuristic approach. However, we have developed a deeper theoretical basis for rank-sum weights, based on the method of *order statistics* (David, 1991). If one generates each weight uniformly on the same interval [a, b] (e.g., $a = 0$ and $b = 100$), and then orders them by size, one will obtain the order statistics of the weights (not normalized). One may then renormalize the weights to sum to 1. Surprisingly, the expected values of these renormalized weights correspond to the rank-sum weights of Eq (4). Thus, there is a clear statistical basis for the rank-sum approach, although it has not been recognized by previous researchers who have used this approach.

To state this argument more formally, let α_i, $i = 1, 2, \ldots, m$, be independent and uniformly distributed on the same interval $[a, b]$, where $b > a \geq 0$ are constants, and $\alpha_{(1)} > \alpha_{(2)} > \ldots > \alpha_{(m)}$ denote the m ordered observations of these random variables, i.e., the order statistics (David, 1991). Then the mean of $\alpha_{(i)}$ is $E(\alpha_{(i)}) = \frac{m+1-i}{m-1}(b-a)$ and the expected value of the sum of the weights is $\sum_{i=1}^{m} E(\alpha_i) = \frac{m}{2}(b-a)$. Thus the normalized weights are $w_{(i)} = E(\alpha_{(i)}) / \sum_{i=1}^{m} E(\alpha_{(i)})$ $= \frac{2(m+1-i)}{m(m+1)}$, $i = 1, 2, \ldots, m$, which corresponds to the rank-sum weights of Eq (4).

Do Differences Among Weights Matter?

As we have already noted, the four sets of weights in Exhibit 1 differ in their steepness. The ROC weights are steepest, followed by the ratio weights, rank-sum weights, and then equal weights. Do such differences matter? As the present example illustrates, they can. In this example, the ratio weights indicate a ranking of the cities in which v(Large) > v(Medium) > v(Small). Although the preference for Large over Medium is weak, the preference for both over Small is strong. The ROC weights closely resemble the ratio weights and lead to the same ranking of the cities. However, the value differences are larger under the ROC model.

In contrast to the ratio weight and ROC methods, the equal weights method dictates exactly the opposite rank ordering, and the value differences are relatively large. Finally, the rank-sum weights may be viewed as a compromise in which Medium is slightly preferred to Large, and Large slightly preferred to Small. In short, despite the widespread belief that linear models are insensitive to differences among weights, it is easy to construct examples in which different approaches to weight assignment have radically different implications for preference and action.

It is important to note that in this example, the ROC and rank-sum weights both depend on the same ordering of attributes as the ratio weights, so the differences among the three methods in this example result only from how they translate attribute ranks into ratios. Because the different methods can lead to different decisions, it is important to determine which approach gives the best results under different circumstances.

Past Comparisons of Quantitative and Approximate Weights

Dawes and Corrigan (1974) investigated characteristics of linear models and found that, in the examples they investigated, linear models were robust with respect to deviations from "optimal" weights; that is, using non-optimal weights led to little loss of value. They argued that even a linear model with random weights might perform quite well and advocated the use of equal weights linear models because such models yielded predictions very highly correlated with those obtained from models based on the "optimal" weights in their examples.[7] Einhorn and Hogarth (1975) derived the minimum correlation between differential and equal weighting models, which provides a lower bound on how well an equal weighting model will perform if attributes are positively correlated. Wainer (1976) offered additional arguments supporting the use of equal weights. In all of this work, the emphasis was on decision and prediction domains in which attributes were either uncorrelated or positively correlated with one another.

However, equal weighting schemes have been criticized because they ignore the potential for negative correlations among attributes. Negative correlations arise, for example, within non-dominated sets of choice alternatives. With negative correlations among attributes, differences in weights can have a large impact on the alternatives chosen (for a review, see von Winterfeldt and Edwards, 1986, Chapter 11). Following this argument, Stillwell, Seaver and Edwards (1981) proposed several rank-weighting techniques as approximations of weight assessments. They found that the rank order weights performed substantially better than equal weights in their decision problems. Barron and Barrett (1996) used simulation methods to compare several rank-order approximation methods (including their ROC method) with the equal weights approach. In their simulation studies, they showed that the ROC method was much more accurate than the equal weights method in selecting the best alternative, and more accurate than the rank-sum method described earlier.

A number of investigators have compared equal weights and rank-order weights with directly assessed ratio-scale weights. Some argued that equal weights performed better than the stated ratio weights (Einhorn and Hogarth, 1975; Horsky and Rao, 1984). Others found that the ratio weights yielded more accurate predictions than equal weights (e.g., Cook and Stewart, 1975; Schoemaker and Waid, 1982). In another experiment, Einhorn and McCoach (1977) found that ratio, rank order weights, and equal weights yielded similar results.

One shortcoming of some previous studies is that researchers made strong assumptions about the relationship between the true weights and directly assessed ratio weights. For instance, Stillwell, Seaver and Edwards (1981) used directly assessed ratio weights as a benchmark for comparing rank-weighting techniques. This implicitly suggests that the directly assessed ratio weights are the true weights. As we argued above, however, directly assessed quantitative weights are subject to random (and possibly systematic) response error. Thus, treating assessed ratio weights as the true weights is questionable.

A different problem arises in Barron and Barrett's (1996) simulation methods. Their simulations assume that the true rank order of weights is known with certainty but that nothing more than that is known. Thus, they sample their true weights from the simplex satisfying the weight constraints. This approach assures that the ROC method always agrees with the ranking of the true weights in their simulations (though not with the magnitudes of the true weights). It also assures that the ROC weights correspond to the expected value of the true weights in the simulation. This gives the ROC method an important edge over other weighting schemes (e.g., equal weights and rank-sum) which are based on different assumptions about the distribution of weight values.

Barron and Barrett's assumption that the ranking of weights is known with certainty is a strong one. We believe that it is more reasonable to assume that both the magnitudes and ordering of weights may be subject to uncertainty. On the other hand, the assumption that only the ranking is known also seems problematic. Often we may be quite confident that some differences in importance are larger than others. An important advantage of the simulation approach we propose in the next section is that it captures uncertainty about both the ranks and magnitudes of attribute weights.

Simulation Framework for Evaluating Decision Quality:

A Dirichlet Model of Uncertain Attribute Weights

Preference Uncertainty and Decision Quality

Ideally, evaluations of decision quality resulting from the use of different weighting methods would be based on comparisons with a true measure of preference. In practice, however the true measure is extremely difficult to define, so previous behavioral research has usually relied on the convergent validity criterion to test different weighting methods. To the extent that different methods yield similar estimates, all gain in credibility. However, a number of studies have shown that different weighting approaches lead to inconsistent estimates (e.g., Eckenrode, 1965; Hobbs, 1980; Schoemaker and Waid, 1982; Jaccard, Brinberg and Ackerman, 1986; Borcherding, Eppel and von Winterfeldt, 1991; Fischer, 1995).

In practice, it seems reasonable to assume that some weighting methods should be more accurate than others. Thus, some authors used a logical approach to determine the criterion weights; for example, regression weights in prediction settings (Dawes and Corrigan, 1974; Einhorn and Hogarth, 1975), or "expert" weights in a purely evaluative context (Borcherding, Eppel and von Winterfeldt, 1991). Clearly, however, both regression weights and expert weights are subject to random (and possibly systematic) error and thus may fail to represent a decision maker's true beliefs about attribute weights.

Computer simulation can provide a powerful tool for investigating psychological constructs that cannot be readily observed or measured (for a good example, see Payne et al., 1990). In the present context, we can use simulation to explore the consequences of preference uncertainty which manifests itself in the form of random response error regarding attribute weights. As we argued earlier, people typically experience some uncertainty regarding their own preferences, particularly when choices require tradeoffs between objectives. In such cases, one may think of attribute weights as being sampled from a distribution of possible weights. The distribution of the weight w_i may be thought of as resulting from the addition of a random error component to a true weight. Given this assumption, it seems reasonable to equate the decision maker's "true weights" with the expected values of these distributions of weights. As noted earlier, similar approaches have been taken in analytical studies by Kleinmuntz (1990), Ravinder, Kleinmuntz, and Dyer (1988), and Ravinder and Kleinmuntz (1991).

Barron and Barrett (1996) used this approach to conduct an extensive simulation experiment to evaluate the quality of decisions resulting from equal weights, rank-sum weights, rank-reciprocal weights (not considered in this article), and ROC weights. As we noted earlier, however, their simulation was based on the strong assumption that the true rank order of the weights was known with certainty. A more realistic assumption is that knowledge of both the ranks and magnitudes of weights is subject to uncertainty. Thus, the rank order information obtained from the subject may not be consistent with the true rank order. This could make a substantial difference when comparing different weighting methods in terms of decision quality. In this article, we report the results of a computer simulation experiment based on the more realistic assumption that both the ranks and magnitudes of weights are uncertain and thus can be known only probabilistically.

Simulation Process

The simulation experiment reported here concerns riskless choices among discrete sets of alternatives whose consequences are known with certainty. Thus, risk attitudes do not play a role. To preview, the simulation process involves five major steps. First, we specify a set of *n* decision alternatives, each described by a vector of *m* value attributes. Second, we select a set of true weights for the attributes, sampled from a probability distribution of possible true weights. Third, we establish a probability distribution for each attribute weight, with expected value equal to the true weight of the attribute. These distributions represent a decision maker's uncertainty about his own preferences. Then, we generate a vector of quantitative ratio weights, sampled from the probability distributions for each weight. Fourth, we generate rank-sum and ROC weights based on the rank ordering of the ratio weights. We also generate equal weights. Fifth, we use each set of weights to select a choice alternative, then evaluate these choices by comparing them with the choice dictated by the true weights vector. We will now describe each step in greater detail.

Step 1. Specify the set of choice alternatives. A set of alternatives in our simulation is defined by the number of attributes, *m*, and the number of alternatives, *n*. In the simulations reported in this article, we chose $m = 3, 6, 9$ and $n = 5, 10, 15, 20$. A uniform (0, 1) distribution was used to generate attribute values x_i. Then the single attribute value function $v_i(x_i)$ was determined by the normalization:

$$v_i(x_i) = (x_i - x_{i, min}) / (x_{i, max} - x_{i, min}) \quad (4)$$

where $x_{i, min}$ and $x_{i, max}$ are the minimum and maximum values in attribute *i* respectively. All attribute values and alternatives were generated independently in each iteration.

Step 2. Specify the true weights. We will assume that the true weights are *uniformly distributed* on the simplex of weights. There are at least two reasons to do this: First, in real scenarios, the true weights are rarely known. Thus, the only reasonable course of action is to generate them uniformly on the whole simplex of weights. Second, to make our simulation results as general as possible, we wish to consider all possible true weights. Therefore, we used the following procedure to generate the true weights for each iteration of the simulation. First, we randomly sampled $m - 1$ numbers from the uniform distribution U[0, 1]. Then we ranked these numbers to split the unit interval (0, 1) into *m* subintervals. The *lengths* of these subintervals are uniformly distributed on the simplex and have a joint distribution that is a an *m*-variate Dirichlet distribution (David, 1991, pp. 98-99; DeGroot, 1970, p. 64). We used these subinterval lengths to define the vector of true weights, $\omega_1, \omega_2, \ldots, \omega_m$. We may arbitrarily index these weights in order of their size.

For example, suppose that $m = 3$. Then we would sample two numbers from U[0, 1]; for example, 0.9 and 0.3. These numbers divide (0, 1) into three subintervals – (0, 0.3), (0.3, 0.9), and (0.9, 1.0) – whose lengths are 0.3, 0.6, and 0.1, respectively. Ordering these lengths by size, we obtain the true weights vector $\omega_1 = 0.6$, $\omega_2 = 0.3$, and $\omega_3 = 0.1$.

Step 3. Generate the ratio weights. We assume that assessed weights are sampled from a Dirichlet distribution whose single-attribute means correspond to the true weights. Thus, the assessed weights are unbiased but subject to random error. To generate such weights for each iteration of the simulation, we used the following procedure. First, we generated a set of Gamma random variables, $r_i \sim \text{Gamma}(\omega_i, \lambda)$, $i = 1, \ldots, m$, where ω_i is the true weight for attribute *i*, and λ is a scaling parameter that controls the precision of the assessed weights. The greater λ, the less error variance in the assessed weights. The expected value of each of these Gamma variates is $\mu(r_i) = \lambda\omega_i$. Then we normalized the r_i to obtain weights that sum to 1; e.g., $w_i = r_i / \sum_j r_j$. The resulting vector of assessed ratio weights has an *m*-variate Dirichlet distribution, with mean vector $\mu(w_i) = \omega_i$, $i = 1, \ldots, m$ (see DeGroot, 1970, pp. 49-50 and 63-64). Note that these sampled ratio weights may differ from the true weights not only in magnitude but also in rank order.

Step 4. Compute the ROC, rank-sum, and equal weights. We computed the rank-sum and ROC weights using the rank order of the assessed ratio weights. Thus, all three sets of weights were based on exactly the same initial information. The ratio weights extracted both the ranks and relative magnitudes of weights; the ROC and rank-sum weights used only the ranking information. Finally, we also computed equal weights, which depended only on the number of attributes.

Step 5. Compute the preferred alternative for each weighting method, then compare with the alternative favored by the true weights model. By applying the additive model (1) to a set of choice alternatives, we can determine the ranking of decision alternatives under each weighting method. These rankings can then be compared with the ranking of alternatives under the true weights model to obtain a measure of decision quality for each weighting method.

A variety of measures of decision quality are possible. One is the *rank order correlation* (Kendall's tau) between the ranking of a set of alternatives by the true weights model and another model. A second is *selection accuracy,* or the proportion of times that a weighting procedure chooses exactly the same alternative as the true weights model. In our pilot simulations, we found that these two measures were highly correlated ($\tau = 0.973$) for 44 sets of simulation results. Thus, only one of them is needed to measure decision quality. We chose selection accuracy because it is easier to understand and simpler to compute. We will use the terms "decision quality" and "selection accuracy" interchangeably. Computing this measure of quality for each weighting method completes one iteration of the simulation process.

Properties of the Dirichlet Model

The Dirichlet model described above has a variety of properties that make it attractive for simulating the distribution of assessed ratio weights. First, the marginal density function over each attribute weight, w_i, has a Beta distribution with mean equal to ω_i, the true weight (DeGroot, 1970). Thus, the model allows us to isolate the effects of random error on weight assessments.

Second, a Beta distribution is defined over the interval (0,1). Thus, $0 \leq w_i \leq 1$, as the first normalization condition of the weights in the additive model (1) requires. Further, the sum of the weights based on a Dirichlet distribution is constrained to be 1. Thus, the Dirichlet weights satisfy both normalization conditions of the additive model (1).

Third, a Beta distribution is attractive as a model of uncertainty about individual weights because it may take on a wide range of shapes. For instance, if w_i is close to 0, there is more room for uncertainty on the high side than on the low. For most parameter values, a Beta with mean close to 0 is skewed to the right. Similarly, a Beta with mean close to 1 is skewed to the left, reflecting the intuition that in this case there is more room for error on the low side. A Beta with a mean of 0.5 is symmetric about 0.5.

Finally, the sum of the Dirichlet parameters may be interpreted as an "equivalent sample size" of experience upon which one's beliefs are based. The larger the sum of the parameters, the less uncertain the weights. Thus, there is a simple and consistent framework for systematically varying the degree of response error present in a simulated weight assessment.[8] Because of the bounded nature of weights (between 0 to 1), the amount of response error present in the simulated weights will be adjusted automatically according to the sampled values of the true weights. For instance, a ratio weight would be subject to a larger response error when the true weight is located in the middle of the range [0, 1] rather than when it is close to 0 or 1. In short, the general properties of a Dirichlet distribution make it very attractive for modeling uncertainty about attribute weights.

Simulation Experiment 1: Error Free versus Random Weights

Simulation Design

Our first experiment explored the two most extreme cases imaginable: one in which the assessed ratio weights were error free, and thus precisely equal to the true weights; and one in which the assessed weights were completely random, and thus uncorrelated with the true weights. These two weight assessment conditions (true or random) were crossed with two other experimental factors: number of attributes (3, 6, or 9) and number of alternatives (5, 10, 15, 20). We ran two replications of 500 trials for each experimental condition, using Latin hypercube sampling. These simulations were run using @*Risk*, a software package for conducting simulation experiments.

Simulation Results

Weights known with certainty. The results of this simulation are displayed in Exhibit 2. Columns 3-6 summarize the results when the true ratio weights are known with certainty. In the absence of error, the model based on the ratio weights always selected the correct alternative. Of greater interest are the comparisons among the three approximate methods. In every case, the ROC method was the best of the approximate methods and the equal weights method the worst. These results reproduce those of Barron and Barrett (1996). Even though our sampling assumptions differed from theirs, when the rank ordering of attributes was known with certainty, the ROC method was clearly the best of the approximate methods. The advantage of the ROC method was most pronounced when the number of attributes was large. In other cases, the differences among methods were small.

/Insert Exhibit 2 here/

Random weights. The other extreme case considered here was one in which the assessed ratio weights were purely random, and thus uncorrelated with the true weights. In this case, the equal weights model performed best, followed by the rank-sum, ROC, and ratio weights methods. If one knows nothing about the true weights, one cannot do better than weight the attributes equally. Any departure from equal weighting makes things worse, even if that departure is random. The superiority of rank-sum weights over ratio weights and ROC weights arose, in this case, because rank-sum weights were flatter than ratio weights and ROC weights, and thus more closely approximated the equal weights vector.

Conclusions. In short, if weights are known with certainty, ratio weights are best and the ROC method is the best of the three approximate methods. At the other extreme, if nothing is known about weights, the equal weights method is best. We believe, however, that most real situations lie between these extremes. In most cases, one can rank objectives, but not with perfect certainty. Moreover, one can usually go beyond ranking to establish quantitative relations among weights, but again uncertainty will arise as to the exact ratios among weights. The next simulation experiment explores this more realistic case. In this simulation, we use the Dirichlet model of uncertainty about weights that we introduced in the previous section.

Experiment 2: Uncertain Weights Generated from Dirichlet Distributions

Simulation Design

In this experiment, we considered all possible combinations of three factors: weighting model (true ratio weights, sampled ratio weights, ROC weights, rank-sum weights, and equal weights), number of attributes (3, 6, 9), and number of alternatives (5, 10, 15, 20).

A fourth factor, amount of response error, was partially tied to the number of attributes. Amount of response error was operationalized as the sum of the Dirichlet parameters. For 3 attributes, we considered two levels of response error: sum = 3 (high error) and sum = 6 (low error). For 6 attributes, we considered sum = 12 (high error) and sum = 24 (low error). For 9 attributes, we considered sum = 18 (high error) and sum = 24 (low error). The sum of the Dirichlet parameters may be thought as the equivalent sample size upon which the distribution is based. The larger the sample, the less the uncertainty. Consider the case of 3 attributes. If the sum of parameters is equal to 3 and all the true weights are equal to 1/3, then the distribution of the stated ratio weights would be uniform on the simplex of weights and the standard deviation of each weight would be 0.235. If the sum of the parameters is equal to 6 and the true weights are equal to 1/3, then the standard deviations of each weight would be 0.178.

This simulation experiment was also run using the @*Risk* software. Again we ran two replications of 500 trials for each experimental condition, using Latin hypercube sampling. [9]

Simulation Results

The results differ somewhat depending on the degree of response error present. Thus, we discuss the two cases separately.

Low response error. When response error is relatively low (top panel of Exhibit 3), the ratio weights correspond relatively closely to the true weights. In this case, several important results emerge:

/Insert Exhibit 3 here./

1. For all combinations of number of alternatives and number of attributes, the ratio weight method had the highest selection accuracy. It surpassed the other methods in every comparison.

2. The ratio weights method was substantially superior to the equal weights method. With larger numbers of attributes and alternatives, the ratio weights chose the best option as much as 25% more often.

3. The advantage of ratio weights over the two rank-order methods was more modest. With 3 attributes, the advantage was only on the order of 2%-5%. With larger numbers of attributes and alternatives, the advantage became as large as 5%-7.5%.

4. *Among the approximate methods, the ROC method was consistently superior to the rank-sum method*. However, the differences between the ROC and rank-sum methods were small, ranging from less than 1% to as much as 4%.

4. *Both rank-order methods were substantially better than the equal weights method*. With 3 attributes, they afforded 5% to 15% better selection accuracy. With larger numbers of attributes and alternatives, the rank-order methods offered 20%-25% improvements over equal weights. Of course, the ratio weight method was better still. These results supports previous findings that ratio weights and rank order weights perform better than equal weights (Barron and Barrett, 1996; Cook and Stewart, 1975; Schoemaker and Waid, 1982; Stillwell, Seaver and Edwards, 1981).

High response error. When response error is relatively large (bottom panel of Exhibit 3), the ratio weights still have expected value equal to the true weights but have greater variance about these true weights. In this case, several important results emerged:

1. *Even with relatively high uncertainty about weights, the equal weights method was consistently inferior to all of the other methods*. The equal weights had roughly a 10% lower selection accuracy when the number of attributes was small and slightly more than a 20% lower selection accuracy with a larger number of attributes.

2. *With 3 attributes, there was no consistent winner among the ratio weight, ROC weight, and rank-sum weight methods*. The differences among the methods were small and the winner differed from case to case. This result is similar to those from an experimental study by Einhorn and McCoach (1977).

3. *With more than 3 attributes, the ratio weight method was best in every case studied, but its superiority over the ROC and rank-sum methods was small.* The advantage of ratio weights

over ROC weights ranged from less than 1% to a bit over 2%. The advantage of ratio weights over rank-sum weights was only a bit larger; the largest difference was 4%.

4. With more than 3 attributes, the ROC method was consistently superior to the rank-sum method, but the differences were very small. They ranged from less than 1% to a bit over 2%.

Other results. Several other results are evident when we look across the high and low uncertainty panels of Exhibit 3.

1. The ratio weight method benefited more than the rank-order methods when the response error decreases. Exhibit 4 provides a clear picture of this for the 3 attributes x10 alternatives case. Recall that larger sums of the Dirichlet parameters are associated with lower levels of response error – i.e., with greater precision in the assessed weights. Exhibit 4 reveals that as the precision of the assessed weights increases, the selection accuracy of the ratio weights increases more rapidly than that of the ROC or rank-sum weights. Further, the ROC method improves somewhat faster than the rank-sum method. However, both the ROC and rank-sum methods reach an upper bound of selection accuracy as error in the assessed weights disappears. This limit is about 80% for both in this case (see row 2 of Exhibit 2). Ratio weighting methods, such as swing weights or tradeoff weights, have no such a limit on selection accuracy (see Exhibit 2). Therefore these methods have more potential accuracy than ROC, rank-sum, or equal weights.

/Insert Exhibit 4 here./

2. The selection accuracy of all weighting methods decreased as the number of alternatives increased. This was true for every combination of method, number of attributes, and level of uncertainty. The effect was more pronounced when the number of attributes was large and uncertainty was high. This result is consistent with findings reported in some empirical studies

(Cook and Steward, 1975; Adelman, Sticha and Donnell, 1984).

3. Number of attribute had a less consistent impact on selection accuracy. The selection accuracy of the equal weights method declined by about 10% as the number of attributes increases from 3 to 9. But the other methods were much less affected. When uncertainty about weights was high, selection accuracy of ratio weights, ROC weights and rank-sum weights declined slightly as the number of attributes increased from 3 to 9, but the effect was small (usually about 1%-2%). When uncertainty about weights was low, the selection accuracy of ratio weights, ROC weights and rank-sum weights was even less affected by the number of attributes.

Conclusions

Main Findings

First, the selection accuracy of quantitatively stated ratio weights was as good as or better than that of the best approximate methods under all conditions studied, except when the assessed weights are purely random (uncorrelated with the true weights). This extreme case seems unlikely to arise in real world contexts. When ratio weights are assessed with a small or moderate amount of error, choices based on these ratio weights are better in all cases than the decisions based on ROC weights, rank-sum weights, or equal weights. When ratio weights are assessed with a large degree of random error, decisions based on these weights are at least equal in quality to those resulting from ROC weights and rank-sum weights, and still substantially better than those based on equal weights.

Second, because linear decision models are quite robust with respect to a change of weights, using approximate weights yields satisfactory decision quality under a wide variety of circumstances. In most cases studied, the ROC and rank-sum weights were not much worse than the ratio weights, especially in high response error conditions. In most cases, ROC weights had a better selection accuracy than the rank-sum weights, but the differences were small.

Third, despite the robustness of linear models, even noisy information about the ranking of attributes improves decisions substantially. Under almost all conditions studied, both the ratio weights method and the two rank-based methods were substantially superior to the equal weights method. The lone exception occurred when assessed weights were purely random. As noted earlier, this is an extreme case unlikely to arise in real-world contexts.

Finally, when response error is present, decision quality decreases as the number of attributes and/or the number of alternatives increases. This is true for all methods studied, especially the equal weights method. As tasks become more complex, selection accuracy becomes more sensitive to errors in the specification of attribute weights. However, selection accuracy depends much more on the number of alternatives than on the number of attributes. One caveat is necessary here. Although selection accuracy declines substantially as the number of alternatives becomes large, the cost of errors is likely to be smaller with large sets of alternatives. The larger the choice set, the smaller the expected difference between the best and second best alternatives.

We should qualify these findings by noting two restrictions to our methods. First, we considered only additive value structures, in part because approximate weighting methods (such as rank-sum and ROC) are limited to this domain. Only by using quantitative weight elicitation methods can we detect and parameterize degrees of non-additivity (Fischer, 1976; Keeney & Raiffa, 1976).

Second, our simulations only considered sets of attributes that were *uncorrelated in the population* from which attribute values were sampled. However, because the choice sets were relatively small, most correlations will be non-zero in particular choice sets. Some will be positive, others negative, as chance determines the sample of attribute values defining a specific choice set. Because there is no a priori reason to assume a particular correlational structure, this seems a reasonable approach. Further, we know from past analytical and simulation work how imposing a correlation structure would affect the results. In general, strong positive correlations among attributes reduce sensitivity to attribute weights, and strong negative correlations increase sensitivity (e.g., Barron & Barrett, 1996). Thus, we would expect differences among the methods studied here to be greatest when attributes are negatively correlated and weakest when attributes are positively correlated. The actual results presented should fall between these extremes.

Choosing a Method: Decision Quality versus Effort

These results raise a final question. Which weighting method should one use? As the Payne, Bettman, and Johnson (1993) program of research has shown, people make systematic tradeoffs between accuracy (decision quality) and effort when choosing judgment strategies. The greater the stakes and the greater the resources (e.g., time) available, the more people tend to use more effortful but more accurate strategies. Strategy selection by experimental participants need not involve a conscious choice among strategies, but could also reflect an adaptive process of learning from experience. In a decision analysis context, on the other hand, choice of methods is more explicit. One such choice concerns how to assess attribute weights.

The results reported in this article suggest the following ranking of methods in terms of accuracy: quantitative ratio weights > ROC weights > rank-sum weights > equal weights. The differences are smallest when assessment error is high and largest when it is low. The differences are also increasingly pronounced when the number of attributes and alternatives is large. The ranking of methods in terms of effort is similar: quantitative ratio weights > ROC weights = rank-sum weights > equal weights. Thus, there is a direct tradeoff between accuracy and effort. The equal weights method is easier, but it is also substantially inferior to other methods. If decision quality is a major goal, then at least a rank-based weighting method is needed. Based on our experience in running experiments and working in real decision analysis settings, we believe that the difference in effort needed to go from ranked weights to quantitative weights (e.g., swing weights or tradeoff weights) is usually quite modest. Thus, in most cases, we believe that a careful weighing of costs and benefits is likely to favor assessing the magnitudes as well as the ranking of weights. A little extra effort can produce a large gain, especially if response error is low and the number of attributes and alternatives is large. The argument in favor of ratio weights is especially compelling if uncertainty about the magnitudes of weights is low to moderate. However, if time is at a great premium, or if the decision maker resists quantification of priorities, the ROC and rank-sum methods offer an excellent compromise that yields good quality and requires only a modest assessment effort.

If one resorts to one of the rank-based methods, which is better? The assessment effort required is identical, and computational costs are negligible. Thus, accuracy alone governs the choice. The Barron and Barrett (1996) simulations indicate a strong advantage for the ROC method. However, their simulations employed a sampling process in which the rank-order of assessed weights always corresponded to the rank-order of the true weights. This creates favorable conditions for the ROC method. We found a similar but weaker advantage for ROC weights in simulations in which both the ranking and magnitudes of weights were subject to error.

Does this mean that ROC weights will always be better than rank-sum weights? No. Two considerations determine which method is better. First, as our simulation of the extreme case of purely random weights indicates, if knowledge of weights is highly limited, rank-sum weights are better (see Exhibit 2). Therefore, very high levels of assessment error probably favor rank-sum over ROC weights. Second, the choice of a method depends on one's beliefs about the steepness of the distribution of true weights. ROC weights are steeper, giving greater weight to the more important attributes than the rank-sum method (see the example in Exhibit 1). Thus, if one believes the distribution of true weights is steep, ROC weights are likely to be better. If not, rank-sum weights are likely to be better. But this assumption goes beyond ranking to include some information about the relative magnitudes of weights. Behavioral evidence suggests that the implicit weights governing people's choices are often quite steep (Fischer & Hawkins, 1993; Tversky, Sattath, & Slovic, 1988). However, the relative steepness of an ideal set of weights depends on many factors, including the relative ranges of outcomes on various attributes (Fischer, 1995; Keeney & Raiffa, 1976; von Winterfeldt and Edwards, 1986).

Simulation and Behavioral Experiments

The simulation method used in this study complements the information generated by traditional psychological experiments. Clearly, the simulation is not an exact analog to human judgment. However, key elements of the model are known to provide a good representation of human judgment. For example, linear models have been shown to provide good approximations to individual judgments in a wide array of contexts. Further, recent research has shown that the Dirichlet error model used here provides a good approximation to both test-retest reliability and response times in multiattribute judgments (Fischer, Luce, & Jia, 1996). Thus, the model provides a tool for exploring the impact of errors on weight assessment methods.

One advantage of the simulation approach is that it allows us to precisely specify and manipulate factors such as the true weights and amount of response error. These are very difficult to measure and even more difficult to manipulate in a behavioral experiment. By using simulation, we can investigate a wider range of situations (e.g., the large error case versus the small error case) than we could in psychological experiments. We can also isolate some factors (e.g., assuming there is no error in assessments of $v_i(x_i)$) while focusing on others (e.g., response error in assessments of w_i). Moreover, simulation can do some things that are impossible in behavioral experiments. One obstacle to behavioral validation of weight assessment methods is the difficulty of establishing a "gold standard" or benchmark for evaluating decision quality. In a simulation, by contrast, the true weights and preferences are known with certainty (to the experimenter, but not to the simulated actor). This makes it easy to specify measures of decision quality.

Finally, it is interesting to note that some previous behavioral experiments have produced results that closely resemble ours for the large response error conditions. This suggests that weight assessments are frequently subject to a considerable degree of response error. In addition, there is ample evidence of systematic between-response-task differences in weight assessments (e.g., between direct ratings of importance, tradeoff weights, and weights inferred from choices). This suggests that the error processes of different weighting methods may be quite different. The development of weight elicitation methods that are accurate and robust in the presence of response error is an important area for future research. Developing better models of the error process is an important element of this research. Good behavioral models can improve prescriptive models and methods for assessing preferences.

ACKNOWLEDGMENTS

The results described in this paper were presented at the ORSA/TIMS Joint National Meeting, Chicago, May, 1993. The authors thank Frank Yates, Robyn Dawes, John Butler and two anonymous reviewers for their helpful comments and suggestions. This work was supported in part by funds provided the Isle Maligne Society Research Fund at the Fuqua School of Business.

References

Adelman, L., Sticha, P. J. and Donnell, M. L. "The Role of Task Properties in Determining the Relative Effectiveness of Multiattribute Weighting Techniques," *Organizational Behavior and Human Performance*, **33** (1984), 234-262.

Barron, F. H. and Barrett, B. E. “Decision Quality Using Ranked Attribute Weights,” *Management Science*, **42** (1996), 1515-1525.

Borcherding, K., Eppel, T. and von Winterfeldt, D. "Comparison of Weighting Judgments in Multiattribute Utility Measurement," *Management Science*, **37** (1991), 1603-1619.

Cook, R. L. and Stewart, T. R. "A Comparison of Seven Methods for Obtaining Subjective Descriptions of Judgment Policy," *Organizational Behavior and Human Performance*, **13** (1975), 31-45.

David, H. A. *Order Statistics*, New York: John Wiley, (1991).

Dawes, R. M. "The Robust Beauty of Improper Linear Models in Decision Making," *American Psychologist*, **34** (1979), 571-582.

Dawes, R. M. and Corrigan, B. "Linear Models in Decision Making," *Psychological Bulletin*, **81** (1974), 95-106.

DeGroot, M. H. *Optimal Statistical Decisions.* New York: McGraw-Hill, 1970.

Delquie, P. “Inconsistent Tradeoffs between Attributes: New Evidence in Preference Assessment Biases.” *Management Science*, **39** (1993), 1382-1395.

Dyer, J. S. and Sarin, R. A. "Measurable Multiattribute Value Functions," *Operations Research*, **22** (1979), 810-822.

Eckenrode, R. T. “Weighting Multiple Criteria,” *Management Science*, **12** (1965), 180-192.

Einhorn, H. and Hogarth, R. M. "Unit Weighting Schemes for Decision Making," *Organizational Behavior and Human Performance*, **13** (1975), 171-192.

Einhorn, H. and McCoach, W. "A Simple Multiattribute Utility Procedure for Evaluation," in S. Zionts (eds.), *Multiple Criteria Problem Solving, Proceedings of a Conference*, Buffalo, NY, 1977.

Fischer, G. W. “Multi-Dimensional Utility Models for Risky and Riskless Choice,” *Organizational Behavior and Human Performance*, **17** (1976), 127-146.

Fischer, G.W. “Range Sensitivity of Attribute Weights in Multiattribute Value Models,” *Organizational Behavior and Human Decision Processes,* **62** (1995), 252-266.

Fischer, G. W., & Hawkins, S. “Strategy Compatibility, Scale Compatibility, and the Prominence Effect,” *Journal of Experimental Psychology: Human Perception and Performance*, **19 (**1993). 580-597.

Fischer, G. W., Luce, M. F., and Jia, J. "Attribute Conflict and Preference Uncertainty: Effects on Judgment Time and Error," Fuqua School of Business, Duke University, 1996.

Hobbs, B. F. "A Comparison of Weighting Methods in Power Plant Sitting," *Decision Science*, **11** (1980), 725-737.

Horsky, D. and Rao, M. R. "Estimation of Attribute Weights From Preference Comparisons," *Management Science*, **30** (1984), 801-822.

Jaccard, J., Brinberg, D. and Ackerman, L. J. "Assessing Attribute Importance: A Comparison of Six Methods," *Journal of Consumer Research*, **12** (1986), 463-468.

Kleinmuntz, D. N. “Decomposition and the Control of Error in Decision Analytic Models,” In R. M. Hogarth (Ed.), *Insights in Decision Making: A Tribute to Hillel J. Einhorn,* University of Chicago Press, 1990, pp 107-126.

Laskey, K. B., and Fischer, G. W. “Estimating Utility Functions in the Presence of Response Error,” *Management Science,* **33** (1987), 965-980.

Keeney, R. L. and Raiffa, H. *Decision with Multiple Objectives: Preference and Value Tradeoffs,* New York: Wiley, 1976.

Kirkwood, C., & Sarin, R. “Ranking with Partial Information: A Method and an Application.” *Operations Research,* **33** (1985), 38-48.

Payne, J. W, Bettman, J. R., and Johnson, E. J. *The Adaptive Decision Maker.* Cambridge University Press, 1993.

Payne, J. W, Johnson, E. J., Bettman, J. R., and Coupey, E. "Understanding Contingent Choice: A Computer Simulation Approach," *IEEE Transactions on Systems. Man. and Cybernetics.* **20** (1990), 296-309.

Ravinder, H. V., & Kleinmuntz, D. N. “Random Error In Additive Decompositions Of Multiattribute Utility,” *Journal of Behavioral Decision Making,* **4** (1991), 83-100.

Ravinder, H. V., Kleinmuntz, D. N., & Dyer, J. S. “The Reliability Of Subjective Probability Assessments Obtained Through Decomposition,” *Management Science,* **34** (1988), 186–199.

Schoemaker, P. J. H. and Waid, C. C. "An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Model," *Management Science,* **28** (1982), 182-196.

Slovic, P., Griffin, D., & Tversky, A. “Compatibility Effects In Judgment And Choice,” In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 5-27). Chicago: The University of Chicago Press, 1990.

Stillwell, W. G., Seaver, D. A. and Edwards, W. "A Comparison of Weight Approximation Techniques in Multiattribute Utility Decision Making," *Organizational Behavior and Human Performance*, **28** (1981), 62-77.

Tversky, A., Sattath, S., & Slovic, P. “Contingent Weighting in Judgment And Choice,” *Psychological Review,* **95** (1988), 371-384.

Tversky, A., and Shafir, E. “Choice Under Conflict: The Dynamics of Deferred Decisions,” *Psychological Science*, **6** (1992), 358-361.

von Winterfeldt, D. and Edwards, W. *Decision Analysis and Behavioral Research*, Cambridge University Press, 1986.

Wainer, H. "Estimating Coefficients in Linear Models: It Don't Make No Nevermind," *Psychological Bulletin*, **83** (1976), 312-317.

Footnotes

[1] Under standard assumptions, multiattribute value and utility functions are defined on an interval scale; i.e., they are unique to a positive linear transformation specifying the units and origin of the scale. The attribute weights (scaling constants) in these models reflect the impact (on the overall index of value or utility) of moving each attribute from its worst to best level, other things being equal. Thus, these weights are defined on a ratio scale; i.e., they are unique to a positive multiplicative transformation. The origin is uniquely determined because a weight of 0 means that an attribute has no impact. Under the standard normalization in which weights sum to 1.0, they are uniquely determined.

[2] A measurable value function reflects not only the preference ordering of outcomes, but also orders differences between pairs of outcomes (Dyer & Sarin, 1979). See von Winterfeldt and Edwards (1986) for an introduction to distinctions among various types of preference measures.

[3] The basic assumption for the additive model (1) is that the strength of preference for various levels of each attribute is independent of fixed levels of the other attributes, which is called "difference independence" (Dyer & Sarin, 1979). In particular, each v_i would be unaffected by changes in the levels of the other attributes. These results will also apply to the case of risky multiattribute decision making when the "additivity condition" is satisfied (e.g., see Keeney and Raiffa, 1976). In this latter case, the value functions v_i will be interpreted as von-Neumann and Morgenstern utility functions.

[4] A weight of 0 means that swinging the attribute from its worst to best value has no impact on the overall value index. See footnote 1.

[5] This definition of true weights addresses only the problem of random error. To the extent that systematic biases arise across different preference revealing processes (e.g., choice and matching), our approach suggests that the true weights for choice may differ from the true weights for matching.

[6] Equivalently, the simplex of weights is the line segment with vertices (1, 0) and (1/2, 1/2). The centroid (center of mass) of this line segment is its midpoint, (3/4, 1/4).

[7] Most of their examples involved prediction, so the optimal weights were least-squares regression weights.

[8] In our simulation process, the vector of Dirichlet parameters is $(\lambda\omega_1, \ldots, \lambda\omega_m)$, with sum = λ, where λ is the scaling parameter in the Gamma distributions used to generate the non-normalized swing weights. We varied λ from 3 to 36 across various conditions of our main simulation experiment.

[9] Because the sampling process of true weights in this experiment is the same as in Experiment 1, the expected results for the equal weights method are the same in this experiment as for the last. Therefore, we did not re-run this condition. Instead, we reused the Experiment 1 results for this method.

Authors' Biographies

Jianmin Jia is an Assistant Professor in the Department of Marketing at the Chinese University of Hong Kong. His research interests include decision analysis, behavioral decision making, and consumer choice modeling.

Gregory W. Fischer is a Professor in the Management area of the Fuqua School of Business of Duke University. His research interests are in judgment and decision making, decision analysis, and social cognition.

James S. Dyer is the Foster Parker Centennial Professor in the Department of Management Science and Information Systems at the University of Texas at Austin. His research activities involve multiple criteria decision making, risk management, and capital investment decisions.

Exhibit 1. A City Choice Example of Attribute Weighting Methods

	Attribute	Relative	Normalized Weighting Method				City Attribute Values		
Attribute Name	Rank	Weight	Ratio	ROC	Rank-Sum	Equal	Large	Medium	Small
Job opportunity	1	100	0.357	0.408	0.286	0.167	100	70	30
Housing quality	2	70	0.250	0.242	0.238	0.167	50	70	80
Cost of living	3	40	0.143	0.158	0.190	0.167	25	50	90
Culture & entertainment	4	35	0.125	0.103	0.143	0.167	100	60	30
Recreation	5	20	0.071	0.061	0.095	0.167	20	50	75
Crime	6	15	0.054	0.028	0.048	0.167	20	70	100
Weight Sum	--	280	1.000	1.000	1.000	1.000	--	--	--
Model							Overall Value		
Ratio Weights	--	--	--	--	--	--	66.8	64.5	58.0
ROC Weights	--	--	--	--	--	--	68.9	64.6	56.3
Rank-Sum Weights	--	--	--	--	--	--	62.4	62.9	61.0
Equal Weights	--	--	--	--	--	--	52.5	61.7	67.5

Note: Ratio Weights use quantitative weight information regarding relative attribute weights. The Rank-Sum and ROC Weights use only the attribute ordering information present in the Ratio Weights. The Equal Weights ignore all information about relative importance, weighting each attribute equally.

Exhibit 2. Selection Accuracy With True and Random Ratio Weights

Number of attributes	Number of alternatives	True ratio weights	Equal weights	ROC based on the true weights	Rank-Sum based on the true weights	Random ratio weights	ROC based on random weights	Rank-Sum based on random weights
3	5	1.000	0.670	0.850	0.839	0.538	0.541	0.586
3	10	1.000	0.608	0.821	0.798	0.502	0.512	0.554
3	15	1.000	0.568	0.790	0.738	0.453	0.487	0.513
3	20	1.000	0.544	0.766	0.735	0.443	0.442	0.472
6	5	1.000	0.605	0.854	0.824	0.508	0.516	0.541
6	10	1.000	0.548	0.836	0.784	0.403	0.401	0.451
6	15	1.000	0.471	0.798	0.738	0.346	0.366	0.428
6	20	1.000	0.442	0.780	0.716	0.337	0.345	0.423
9	5	1.000	0.562	0.866	0.788	0.450	0.466	0.520
9	10	1.000	0.504	0.814	0.778	0.340	0.368	0.450
9	15	1.000	0.470	0.802	0.728	0.327	0.323	0.399
9	20	1.000	0.454	0.778	0.692	0.313	0.315	0.373

Note: Entries in columns 3 to 7 are the proportion of times the method in question selected the same best alternative as a model based on the true weights.

Exhibit 3. Selection Accuracy in Experiment 2 with Attribute Weights Generated from a Dirichlet Distribution

Relatively low uncertainty cases

# of Attributes	# of Alternatives	Sum of Dirichlet parameters	Ratio Weights	ROC Weights	Rank-Sum Weights	Equal Weights
3	5	6	0.796	0.768	0.766	0.670
3	10	6	0.772	0.742	0.737	0.608
3	15	6	0.735	0.702	0.683	0.568
3	20	6	0.707	0.682	0.675	0.544
6	5	24	0.807	0.736	0.733	0.605
6	10	24	0.764	0.708	0.682	0.548
6	15	24	0.725	0.687	0.653	0.471
6	20	24	0.716	0.632	0.618	0.442
9	5	36	0.802	0.791	0.755	0.562
9	10	36	0.755	0.720	0.703	0.504
9	15	36	0.748	0.708	0.674	0.470
9	20	36	0.715	0.678	0.639	0.454

Relatively high uncertainty cases

# of Attributes	# of Alternatives	Sum of Dirichlet parameters	Ratio Weights	ROC Weights	Rank-Sum Weights	Equal Weights
3	5	3	0.761	0.748	0.754	0.670
3	10	3	0.716	0.710	0.726	0.608
3	15	3	0.674	0.684	0.671	0.568
3	20	3	0.631	0.625	0.642	0.544
6	5	12	0.748	0.741	0.728	0.605
6	10	12	0.688	0.668	0.660	0.548
6	15	12	0.652	0.643	0.625	0.471
6	20	12	0.619	0.608	0.601	0.442
9	5	18	0.756	0.744	0.716	0.562
9	10	18	0.713	0.695	0.667	0.504
9	15	18	0.674	0.665	0.651	0.470
9	20	18	0.645	0.616	0.613	0.454

Note: Entries in columns 4 to 7 are the proportions of choices in agreement with true-weights choices. Selection accuracy for the equal weights depends on the distribution of true weights, but is unaffected by the sampling distribution of the ratio weights. The equal weights statistics in this table are those generated in Experiment 1.

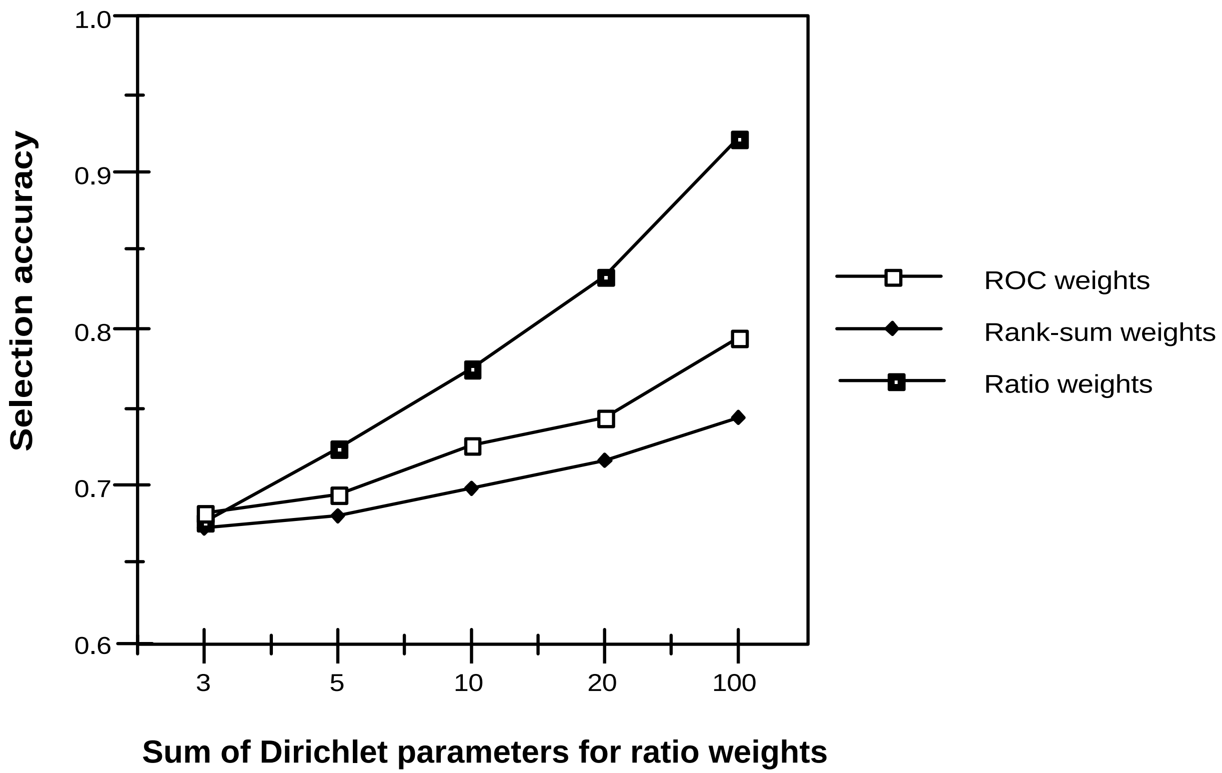

Exhibit 4. Selection accuracy for different weighting methods as a function of the amount of response error in the ratio weights. The larger the sum of the Dirichlet parameters, the less response error present in the ratio weight assessments. The simulated decisions involve 3 attributes and 10 alternatives.