Environment and Planning B: Planning and Design, 1989, volume 16, pages 155–170

The meanings of trade-offs in multiattribute evaluation methods: a comparison

S-K Lai, L D Hopkins
Department of Urban and Regional Planning, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
Received 19 September 1988

Abstract. Multiattribute evaluation methods are being used without full understanding of the meanings of the weights which characterize the relative importance of the attributes. The weights in the methods do not carry the same meaning, even though the questions asked in the procedures are similar. The weights imply the trade-offs among attributes with respect to certain transformed measures, and the decisionmaker is presumed to be able to specify the appropriate trade-off judgments based on preferences. Five multiattribute evaluation methods are compared deductively based on a common framework in which the meanings of weights are clarified: multiattribute utility theory (MUT), weighting and rating, the analytic hierarchy process, concordance analysis, and computation of equivalent alternatives. For each method, the evaluation procedures are described in mathematical terms and weights are derived and interpreted by means of the preference structure represented by MUT. Transformations between the methods make clear the meanings of weights and make possible iterative procedures for conducting empirical comparisons of the methods.

1 Introduction

Many techniques are currently being advocated for making decisions based on more than one attribute. Each technique requires, implicitly or explicitly, that decisionmakers assess some kind of trade-off among these attributes. The exact meaning of these trade-offs is not always made explicit. It is not clear, therefore, that the decisionmaker is providing the trade-off information with the same meaning that the method requires. Comparisons in the current literature lack a common framework to investigate the internal logic of the methods in assessing results (for example, Belton, 1986; Fischer, 1977; Schoemaker and Waid, 1982). Differences in results between methods may come both from the internal logic and from the ways in which the decisionmaker's preferences are elicited in the methods. Straightforward deductive comparisons demonstrate that the logic of different methods implies different meanings for the weights. These differences in meaning are important for useful applications of the methods. The key question, put simply, is: Does the method apply the elicited preferences in the way that the decisionmaker intended when expressing them?

This question requires investigation both of the logical meaning of weights and of the empirical ability of decisionmakers to respond to that meaning. In this paper we first describe multiattribute utility theory (MUT), a method in which the meaning of trade-offs is quite explicit. We then introduce weighting and rating (WR), the analytic hierarchy process (AHP), concordance analysis (CA), and computation of equivalent alternatives (CEA). For each method we define the concrete meaning of the weights that the decisionmaker is asked to provide and we relate that meaning to the decisionmaker's preferences.

The meaning of weights thus derived is important in the empirical comparison of the methods because the questions asked in the evaluation procedures should correctly reflect that meaning. The transformation between the methods developed in the deductive comparison suggests an experimental design using iterative procedures.

The responses from the decisionmaker using one method can be tested against the equivalent responses that should be given in another method. Experiments can then be conducted to determine which method a decisionmaker can implement most successfully.

2 Deductive comparison of methods

In this section, each method is first briefly introduced and detailed implementation steps are then described according to consistent terms. Decisionmaking under uncertainty is not considered because the focus is on the meaning of trade-offs and, except for MUT, few applications of these methods deal with uncertainty. The term 'value function' is, therefore, used instead of 'utility function' throughout the paper. The weights are given different labels in each method, such as 'scaling constants', 'relative importance', and 'trade-offs'. It is our purpose to clarify the meanings of these terms. Table 1 briefly depicts the steps that are required in each method.

To ensure consistency, the terms that are common to each method must be correctly defined. The original measures for attributes, either in natural or constructed units, are referred to as attribute levels. A space is defined as a collection of vectors in which each point represents an alternative. The alternative can be represented by different coordinates in different spaces and, therefore, transformed from one space to another by transformation functions. Unidimensional transformation functions transform attributes into one kind of scale whereas multidimensional transformation functions transform more than one attribute into a unidimensional scale. Standardization refers specifically to the transformation that results in the sum of transformed measures being equal to unity. Since any

Table 1. The evaluation steps in the methods.

Multiattribute utility theory (MUT)
(1) Estimate unidimensional value functions.
(2) Given the unidimensional value functions, derive scaling constants for the attributes.
(3) Apply the decision rule of additive value function to rank alternatives.

Weighting and rating (WR)
(1) Assign weights of relative importance to the attributes.
(2) Select unidimensional transformation functions to transform the raw scores into scores.
(3) Apply the additive decision rule to rank alternatives.

Analytic hierarchy process (AHP)
(1) Construct pairwise comparison matrices of alternatives with respect to the attributes and of attributes with respect to the overall focus.
(2) Derive transformed weights from the matrices by means of the eigenvector approach.
(3) Apply the additive decision rule to rank alternatives.

Concordance analysis (CA)
(1) Assign weights of relative importance to the attributes.
(2) Construct the concordance matrix.
(3) Transform the concordance matrix into the dissimilarity matrix.
(4) Apply multidimensional scaling technique to search for a best-fitted configuration.
(5) Rank alternatives according to the configuration.

Computation of equivalent alternatives (CEA)
(1) Eliminate dominated alternatives.
(2) Select a promising alternative with the highest likelihood of being dominant.
(3) Select another alternative to be compared with the promising alternative.
(4) Make trade-off judgments among attributes for the promising alternative so that one alternative dominates the other.
(4′) Delete dominated alternatives.
(5) Stop when one alternative remains. Otherwise, go to (1).

equivalent ratios such as 2:1 and 4:2 could be used, standardization arbitrarily specifies one set. A special term is given to the transformation functions in MUT as value functions to distinguish them from the transformation functions in other methods. Scores refer specifically to dimensionless scales. Trade-offs among the attributes are defined as the ratios of the relative contributions of attributes to a unit of the overall measure of an alternative. Mathematically, the trade-offs are the marginal rates of substitution of attribute measures relative to a constant amount of the overall measure of the alternative. The determination of weights implies, therefore, trade-offs among the attributes.

In the deductive comparison, we assume that the decisionmaker's preferences can be represented by unidimensional and multidimensional value functions. The decisionmaker is assumed to be perfectly consistent in making value judgments based on these value functions, which are invariant throughout the evaluation procedure. Procedures for coping with inconsistency can be incorporated in any of these methods, but are not pertinent to the present discussion. Inconsistency will be pertinent in the empirical work.

In what follows we will first show how the unidimensional value functions in MUT are elicited and how the alternatives are ranked accordingly. MUT is the only method that elicits the complete set of the value functions. We define, therefore, the preference structure as the unidimensional and multidimensional value functions defined in MUT value space, which serves as the criterion in the deductive comparison. For each of the other methods, we will implement the steps required by the method and compare the weights in the method with those derived from a different but logically equivalent procedure incorporating the preference structure, so that the meaning of weights can be related to the decisionmaker's preferences.

The attribute levels of alternatives are represented in an effectiveness matrix in which rows are the alternatives and columns are the attributes. There are n attributes and m alternatives. The effectiveness matrix for the set of alternatives is given as below, where x_{ij} is the attribute level of attribute j for alternative i. Note that the units of these attributes may be different. All the methods assume such a matrix as a starting point, though creating the effectiveness matrix is a far from trivial task (for example, see Keeney and Raiffa, 1976):

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} . \qquad (1)$$

2.1 Multiattribute utility theory (MUT)

Multiattribute utility theory is the most thoroughly justified method. It is derived directly from utility theory (Keeney and Raiffa, 1976). MUT elicits the decisionmaker's preferences by first asking questions in terms of attribute levels. The unidimensional value functions can then be approximated through a curve-fitting technique, such as the mid-value splitting technique. These functions transform the attribute levels into an interval-value scale as shown in figure 1. By multiplication of these values by the weights or scaling constants, the trade-offs among the attribute values are taken into account in the multidimensional value function. The scaling constants are derived from a linear system based on the decisionmaker's judgments between indifferent alternatives. The alternatives are finally ranked by substituting the derived scaling constants and the attribute levels into the approximated unidimensional value functions and the multidimensional value function.

Mathematically, MUT constructs the following additive value function:

$$Z^m(X_i) = \sum_{j=1}^{n} w_j^m V_j^m(X_i) , \qquad \text{for } i = 1, 2, ..., m . \tag{2}$$

where

Z^m is the multidimensional value function,

X_i is the attribute vector of alternative i, $X_i = (x_{i1}, x_{i2}, ..., x_{in})$,

V_j^m is the unidimensional value function for attribute j,

w_j^m is the weight or scaling constant for attribute j.

In evaluating the unidimensional value functions by means of the mid-value splitting technique, the decisionmaker has first to set the range for each attribute. He or she then estimates the midpoint of the value interval between the minimum and the maximum of that attribute so that the difference between the minimum and the midpoint, in terms of value, is the same as that between the midpoint and the maximum. The decisionmaker then determines the midpoints for the smaller intervals formed by the previously decided midpoints. The procedure continues until the decisionmaker thinks that the curve of the value function can be approximated. All questions are considered in terms of the original attribute levels.

Scaling constants are needed to rescale these values so that the units of the values for different attributes are equal. The scaling constants are derived indirectly by asking the decisionmaker trade-off questions. For example, consider figure 2. One of the two alternatives, say X_1, contains the maximum and the minimum of the attribute level for the attributes 1 and 2, respectively. The decisionmaker is asked to specify in terms of attribute level of attribute-2 an alternative, X_2, with the minimum level of attribute 1, so that X_2 is indifferent to X_1. The values of the minimum and maximum attribute levels are, by definition, equal to 0 and 1, respectively, because the values are standardized to identify a unique unit of value. Substituting these values into the multidimensional value function Z^m results in an equation of two unknowns, the scaling constants for the two different attributes. If the sum of the weights is set equal to 1, the scaling constants can then be obtained by solving the two equations for the two unknowns. Similar questions can be applied in the case of more than two attributes.

In MUT, then, the meaning of the weights or scaling constants, w_j^m, is that one unit of value in the multidimensional value function Z^m equals w_j^m times a unit of value for unidimensional value function V_j^m. Therefore, the weights imply the trade-offs among attributes with respect to the value of the unidimensional value functions. They are sensitive only to the ranges of the attribute levels, because these ranges determine the range and therefore the unit of the value function.

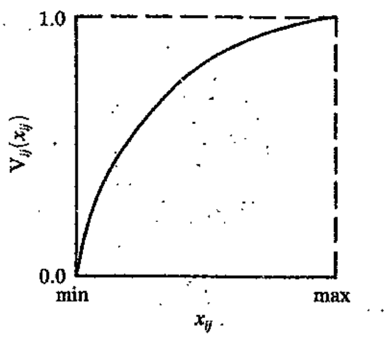

Figure 1. The unidimensional value function for attribute j in multiattribute utility theory.

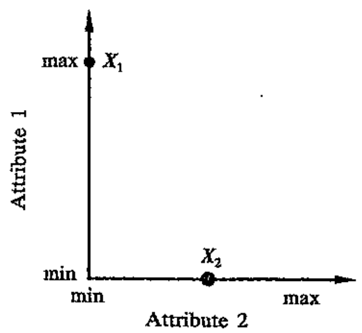

Figure 2. An example of two alternatives of two attributes for deriving scaling constants.

They are not sensitive to the attribute levels themselves. The trade-offs among the attributes relative to the attribute levels can be recovered from the multidimensional value function by taking the partial derivatives of the function with respect to a certain attribute level. If the unidimensional value functions are nonlinear, the trade-offs thus derived are not constants but depend on the levels of the attributes. It would be difficult, therefore, for the decisionmaker to estimate the scaling constants w_j^m directly because that would require thinking about the trade-offs among the attributes in terms of units of value rather than units of the attribute level. We will return in section 3 to the question of whether the decisionmaker intends to express this meaning.

To explain the meaning of weights in a more concrete way, consider the choice of a set of apartments according to two attributes: distance to work (walking time in minutes) and total floor area (square feet), denoted as d and f, respectively. The ranges of the two attributes for the alternatives under consideration are $[-15\text{ min}, -5\text{ min}]$ and $[100\text{ ft}^2, 300\text{ ft}^2]$. The minus sign indicates that larger numbers of minutes are less preferred. Suppose that all the alternatives are noninferior, which means that no alternative dominates others.

In MUT, the unidimensional value functions, V_j^m, are elicited and the functional values for the worst and the best attributes of the ranges are set to be 0 and 1, respectively. The weights, w_d^m and w_f^m, represent the relative contribution of a unit, in terms of value, of distance to work and that in total floor area to a unit of the overall value of any alternative. The ratios of the two weights also imply how many units of value the decisionmaker is willing to give up for one attribute, say distance in walking time, in order to gain a certain amount of value in total floor area. By means of the scaling constants and the unidimensional functions, the trade-offs between attribute levels can be derived at any combination of attribute levels. For example, the decisionmaker may be willing to live 3 minutes farther away from school in order to gain 20 ft^2 of total area at the combination of $(-13\text{ min}, 150\text{ ft}^2)$, but would give up less distance, say 2 minutes, to gain the additional living space if living in a bigger apartment that is also closer to work, say $(-10\text{ min}, 200\text{ ft}^2)$. The diminishing marginal rate of the unidimensional value functions implies that the smaller the apartment is, the greater the value of a unit of total floor area; the farther the distance from work, the greater the value of a unit of walking distance. The diminishing marginal rate makes the value functions nonlinear.

2.2 *Weighting and rating (WR)*

In weighting and rating, the decisionmaker first directly assigns the weights of 'relative importance' to the attributes (for example, see Miller, 1980). Unidimensional transformation functions are selected to transform the attribute levels to an interval scale in which ranges of the attribute levels, such as the differences between the maximum and minimum levels of the attributes, are selected as scaling factors, resulting in a scale from zero to one. From the attribute levels, first subtract the corresponding minima and then rescale by dividing the results by the range. The alternatives are then ranked based on the overall scores calculated by substituting the directly estimated 'relative importance' weights and the transformed attribute levels or scores into a multidimensional transformation function.

Therefore, weighting and rating identifies as the best alternative the one that maximizes the sum of the weighted scores

$$S(X_i) = \sum_{j=1}^{n} w_j^w T_j(X_i), \qquad \text{for } i = 1, 2, \ldots, m, \tag{3}$$

where
S is the multidimensional transformation function of scores,
X_i is the attribute vector of alternative i,
T_j is the unidimensional transformation function for attribute j,
w_j^w is the weight or relative importance for attribute j in WR.

The unidimensional transformation functions, T_j, are not confined to the maximum–minimum scaling approach introduced above. Some analysts have used the mean of the attribute levels over the alternatives for each attribute as the scaling factor. Once a function has been chosen, the transformation of the attribute levels to scores is purely arithmetic, without any preference inputs from the decisionmaker.

Though similar in structure to MUT, the way in which the weights are derived in WR is different. These weights are directly assigned by the decisionmaker considering them as the 'relative importance' among the attributes. Therefore, the weights are actually insensitive to attribute ranges, but the transformation is sensitive to changes in attribute range. In MUT there is a relationship between the scaling constants and the ranges of the attributes because a change in ranges changes the linear system from which the scaling constants are solved. This immediately suggests that there is a difference in the meanings of weights in the two methods.

To examine this, we solve for the WR weights implied by a given preference structure defined in MUT. In other words, if the decisionmaker says the MUT weights are w_j^m, what should that decisionmaker say the WR weights are? Let the vector spaces composed of attribute levels and scores be called the 'original' and the 'transformed' attribute spaces, respectively. According to the steps in WR, the alternatives are transformed from the original attribute space into the transformed attribute space by the unidimensional transformation functions, T_j. That transformed attribute space can, in theory, be further transformed into the MUT value space by a set of unidimensional value functions, which should be different from those in MUT that transform the original attribute space directly to the value space. For the results derived from the two procedures to be consistent, the preference structures thus derived should be the same. In other words, the decisionmaker is presumed to be able to specify preferences in terms of trade-offs with respect to these attribute scores. The evaluation of the alternatives based on the multidimensional value function with respect to scores, for example, equation (3) should be consistent with that in MUT. Otherwise, the scores and the associated weights in the multidimensional transformation function are not indicative of the preference structure.

Suppose that the unidimensional value functions of the scores are V_j^w, for $j = 1, 2, ..., m$. The approach introduced above implies that the composite function $V_j^w o T_j$ is equivalent to V_j^m. That is,

$$V_j^m(X_i) = V_j^w o T_j(X_i), \qquad \text{for } i = 1, 2, ..., m, \quad j = 1, 2, ..., n. \tag{4}$$

Substituting equation (4) into equation (2), we have

$$Z^m(X_i) = \sum_{j=1}^{n} w_j^m V_j^w o T_j(X_i), \qquad \text{for } i = 1, 2, ..., m. \tag{5}$$

To compare equations (5) and (3), both represent the multidimensional transformation functions of scores. In particular, the first is derived from the decisionmaker's preferences. For equation (3) to be valid compared with the preference structure, it is necessary for the trade-offs with respect to scores in equations (3) and (5) to be the same, implying that the weights are identical given that the sums of the weights are set to be a constant. The weights are the coefficients of the multidimensional transformation functions, which can be recovered by taking the partial derivatives of S and Z^m with respect to T, in equations (3) and (5). This gives

$$w_j^w = \left.\frac{\partial Z^m}{\partial T_j}\right|_{Z_m(X) = Z_m(X_i)}$$

$$= w_j^m \left.\frac{\partial V_j^w}{\partial T_j}\right|_{V_j^w \circ T_j(X) = V_j^w \circ T_j(X_i)} . \qquad (6)$$

The weights w_j^w are not constants if the V_j^w are nonlinear with respect to T_j, which implies that the trade-offs among the attributes vary with respect to the scores. The decisionmaker should then assign m sets of weights for the m alternatives, not just a constant set of w_j^w. Therefore, only when the unidimensional value functions V_j^w with respect to scores in WR are linear can the weights w_j^w be multipliers of the scaling constants in MUT and be said to be invariant across the alternatives.

In WR, then, the meaning of the weights w_j^w is that one unit of transformed attribute level in the multidimensional transformation function Z^w equals w_j^w times a unit of transformed attribute level for unidimensional transformation function T_j. These weights imply the trade-offs among the attributes with respect to the transformed attribute levels and vary with attribute level if unidimensional value functions V_j^w are nonlinear with respect to T_j. Even if the value function were linear, it would be difficult for the decisionmaker to estimate these weights directly without appropriate elicitation procedures as in MUT because the transformed attribute levels to which the weights apply are too abstract for a decisionmaker to consider. Does a decisionmaker asked to assess 'relative importance' actually provide weights with the meaning just described? We will return to this empirical question in section 3.

In the apartment example, the distances to work and the floor areas of the apartments are transformed by first subtracting from the attribute levels the minima of the two attributes, -15 min and 100 ft^2, and then dividing the results by the ranges of the two attributes, 10 min and 200 ft^2. For example, an apartment located 10 min from work with 200 ft^2 floor area is represented by the transformed attribute vector (0.5, 0.5). To estimate the weights w_d^w and w_f^w correctly, the decisionmaker has to make trade-off judgments in terms of these dimensionless numbers. If the value functions are nonlinear, the trade-offs with respect to the scores at (0.5, 0.5) should be different from these at (0.20, 0.25) or (13 min, 150 ft^2). These trade-offs can be verified by transforming the attributes and the trade-offs back to the original units by means of the transformation functions shown above. For this case, at level (0.5, 0.5), if we assume the value functions are given as in MUT, the decisionmaker is willing to give up 2 minutes of walking time for 20 ft^2 compared with 3 minutes for 20 ft^2 (0.20, 0.25).

2.3 *The analytic hierarchy process (AHP)*

The theoretical support for AHP has been well developed by Saaty (1980). Compared with MUT, AHP focuses more on consistency checks of the decisionmaker's judgments than on explicit representation of the preference structure. The decisionmaker's value judgements are transformed implicitly from a set of pairwise comparisons based on a verbally expressed, ratio-comparison scale that is restricted to nine integer numbers (Saaty, 1980). AHP requires the decisionmaker to construct a hierarchical value structure as shown in figure 3 and then asks questions about 'relative importance' among the attributes as in WR. The weights are derived through calculating the maximum eigenvalues for the comparison matrices, which implies that the weights have a meaning different from that in MUT. The alternatives are ranked by substituting the weights of the alternatives thus derived with respect to each attribute and those of the attributes with respect to the overall 'focus' into a multidimensional transformation function. The overall focus is defined as the decisionmaker's ultimate goal. The achievement of the alternatives toward this goal can be represented by the attribute levels. Though the measures derived from the unidimensional transformation functions are also called weights in AHP, they are related to the values in MUT as will be shown below and are called transformed or, more strictly, standardized values.

The multidimensional transformation function for AHP is as follows:

$$Z^a(X_i) = \sum_{j=1}^{n} w_j^a V_j^a(X_i) , \qquad \text{for } i = 1, 2, ..., m . \tag{7}$$

where

Z^a is the multidimensional transformation function in AHP,
X_i is the attribute vector of alternative i,
V_j^a is the unidimensional transformation function for attribute j,
w_j^a are the weights for the attribute j with respect to the overall focus.

Though the unidimensional transformation function, $V_j^a(X_i)$, can be expressed by a single formula, it actually includes, in operation, two steps: (1) the transformation of the attribute levels into ratios of values in the pairwise comparison matrix of alternatives with respect to attribute j, and (2) the derivation of the transformed values by the eigenvector approach. In strict applications, the assignment of these ratios is based on Saaty's nine-integer scale. If the decisionmaker is able to express preferences in this way, then the pairwise comparison matrix for attribute j should be the ratios of the attribute values $V_j^m(X_i)$ as in MUT. The matrix shown below is said to be perfectly consistent if $c_{kl} = 1/c_{lk}$, and $c_{kl} = c_{ki}c_{il}$, where c_{kl} is the cell at row k and column l.

$$\begin{bmatrix} \dfrac{V_j^m(X_1)}{V_j^m(X_1)} & \dfrac{V_j^m(X_1)}{V_j^m(X_2)} & \cdots & \dfrac{V_j^m(X_1)}{V_j^m(X_m)} \\ \dfrac{V_j^m(X_2)}{V_j^m(X_1)} & \dfrac{V_j^m(X_2)}{V_j^m(X_2)} & \cdots & \dfrac{V_j^m(X_2)}{V_j^m(X_m)} \\ \vdots & \vdots & & \vdots \\ \dfrac{V_j^m(X_m)}{V_j^m(X_1)} & \dfrac{V_j^m(X_m)}{V_j^m(X_2)} & \cdots & \dfrac{V_j^m(X_m)}{V_j^m(X_m)} \end{bmatrix} . \tag{8}$$

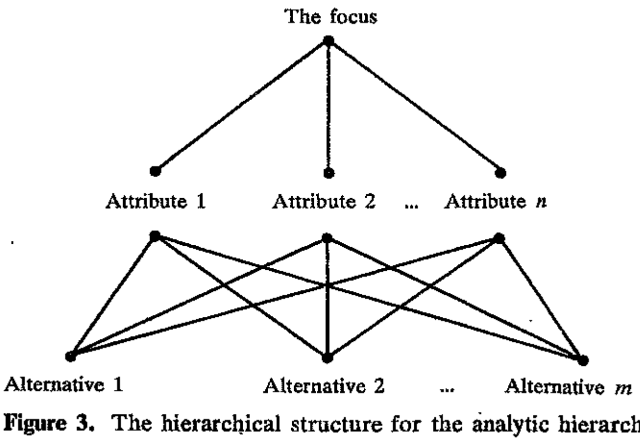

Figure 3. The hierarchical structure for the analytic hierarchy process.

The unique solution of the eigenvector in which the elements sum to 1 can be found by solving for the maximum eigenvalue for the square matrix. In the perfectly consistent matrix, the maximum eigenvalue is equal to the number of rows or columns of the matrix, which is m in this case, and the columns in the matrix are linearly dependent. With a consistent matrix we know a priori that any of the columns in the matrix is an eigenvector solution to the maximum eigenvalue problem. Inconsistency in judgments is discussed further in section 3. For a consistent matrix any column of the values can be standardized by dividing the elements of the column by the sum of the elements. That is,

$$V_j^a(X_i) = V_j^m(X_i) \Big/ \sum_{i=1}^{m} V_j^m(X_i) . \tag{9}$$

Let the reciprocal of the sum of the values of attribute j for the alternatives be denoted as k_j. Equation (9) can be rewritten as follows:

$$V_j^a(X_i) = k_j V_j^m(X_i) , \tag{10}$$

where

$$k_j = \left[\sum_{i=1}^{m} V_j^m(X_i) \right]^{-1} .$$

Similarly, the decisionmaker should estimate the weights for the attributes, the pairwise comparison matrix of the attributes with respect to the overall focus, to be the trade-offs among the attributes with respect to the standardized values $V_j^a(X_i)$. Let these weights be w_j^a. The $n \times n$ matrix is as follows:

$$\begin{bmatrix} \frac{w_1^a}{w_1^a} & \frac{w_1^a}{w_2^a} & \cdots & \frac{w_1^a}{w_n^a} \\ \frac{w_2^a}{w_1^a} & \frac{w_2^a}{w_2^a} & \cdots & \frac{w_2^a}{w_n^a} \\ \vdots & \vdots & & \vdots \\ \frac{w_n^a}{w_1^a} & \frac{w_n^a}{w_2^a} & \cdots & \frac{w_n^a}{w_n^a} \end{bmatrix} . \tag{11}$$

How should these weight ratios or trade-offs be determined so that the multidimensional transformation function in AHP is indicative of the preference structure? Analogous to WR, these weights can be derived from the transformations implied in the evaluation procedures. These transformations are made explicit by the unidimensional transformation functions $V_j^a(X_i)$ in equation (10). Substituting equation (10) into equation (7), we have

$$Z^a(X_i) = \sum_{j=1}^{n} w_j^a k_j V_j^m(X_i) , \qquad \text{for } i = 1, 2, ..., m . \tag{12}$$

The weights w_j^a must be determined so that the trade-offs among the attributes with respect to values are unchanged. In other words, the results of the partial derivatives of Z^a in equation (12) with respect to the unidimensional value functions $V_j^m(X_i)$ are equal to the weights w_j^m in MUT.

$$\begin{aligned} w_j^m &= \frac{\partial Z^a}{\partial V_j^m} = w_j^a \frac{\partial V_j^a}{\partial V_j^m} \\ &= w_j^a k_j ; \qquad \text{for } j = 1, 2, ..., n . \end{aligned} \tag{13}$$

Therefore,

$$w_j^a = \frac{1}{k_j} w_j^m = \sum_{i=1}^{m} w_j^m V_j^m(X_i) , \qquad \text{for } j = 1, 2, ..., n . \tag{14}$$

Because AHP requires that the weights also be standardized, which means that the sum of the weights is equal to 1, then

$$w_j^a = w_j^a \Bigg/ \sum_{k=1}^{n} w_k^a = \sum_{i=1}^{m} w_j^m V_j^m(X_i) \Bigg/ \sum_{i=1}^{m} \sum_{k=1}^{n} w_k^m V_k^m(X_i) , \qquad \text{for } j = 1, 2, ..., n . \tag{15}$$

The meaning of weights in AHP is, therefore, that one unit of standardized value in the multidimensional value function Z^a equals w_j^a times a unit of standardized value for unidimensional transformation function V_j^a. The standardized value of attribute j for alternative i is obtained by dividing the attribute value of that alternative by the total attribute value of a given set of alternatives. This is equivalent to saying that the scale of value in MUT is standardized in AHP and therefore the corresponding weights must also be transformed. The weights in AHP imply the trade-offs among the attributes with respect to standardized values.

In the apartment example, assume that the attribute values for the apartment (-10 min, 200 ft^2) are (0.7, 0.8) and the sums of the values for distance and floor area are 4.9 and 6.4. The standardized values of distance to work and total floor area for the apartment are $\frac{1}{7}$ and $\frac{1}{8}$ respectively. The weights w_d^a and w_f^a are the relative contributions of a unit, in terms of standardized value, of distance to work and that of total floor area to a unit of the overall standardized value for any of the apartments. If the scaling constants w_d^m and w_f^m are equal, the weights in AHP w_d^a and w_f^a can be deductively derived by dividing the total values of the two attributes, 4.9 and 6.4, by the sum of the total values, 11.3. Therefore, the weights in AHP for distance and floor area are 0.43 and 0.57 which are presumed to represent the relative importance of the two attributes.

2.4 *Concordance analysis (CA)*

Concordance analysis is an adaptation for multiattribute evaluation of the traditional concordance methods used in attitude measurement where the task is, given a hypothesis of a 'true' scale, to estimate this scale by sampling among a set of judges (Hopkins, 1980; Moroney, 1951). In CA, each attribute is analogous to a judge and weights are assigned to these attributes. In general, CA requires the decisionmaker to assign weights of relative importance for the attributes. Then, by application of some kind of transformation function, the differences among the alternatives are synthesized by a set of indices. The final ranking of the alternatives is determined by transforming the difference indices into a unidimensional scale. Different transformation functions can be applied and result in different meanings of the weights. The analogy with agreement among judges does not justify the method, because there is no equivalent hypothesis in multiattribute evaluation that the attributes (judges) should agree.

Voogd (1983) proposed a generalized concordance analysis in which the alternatives can be compared in pairs. The weights are used to calculate the concordance and discordance indices for all possible pairs of alternatives. Then, two thresholds of concordance and discordance indices are chosen by the decision-maker as distinguishing 'qualified alternatives'. The final evaluation is based on the total number of concordance and discordance indices of an alternative within the prespecified qualification boundaries. The concordance and discordance index functions, C and D, are defined as follows:

$$\mathrm{C}(X_i, X_{i'}) = \left[\sum_{j \in C_{ii'}} (w_j^c)^\alpha \bigg/ \sum_{k=1}^{n} (w_k^c)^\alpha \right]^{1/\alpha}, \qquad \text{for } i = 1, 2, \ldots, m, \tag{16}$$

$$\mathrm{D}(X_i, X_{i'}) = \left[\sum_{j \in D_{ii'}} (w_j^c |e_{ij} - e_{i'j}|)^\alpha \bigg/ \sum_{k=1}^{n} (w_k^c |e_{ik} - e_{i'k})^\alpha \right]^{1/\alpha}, \qquad \text{for } i = 1, 2, \ldots, m \tag{17}$$

where
$C_{ii'}$ is the set of attributes on which alternative i is superior or equivalent to i',
$D_{ii'}$ is the set of attributes on which alternative i is strictly inferior to alternative i',
e_{ij} is the standardized score of attribute j in CA for alternative i;
w_j^c is the weight of relative importance for attribute j defined in CA;
α is the scaling parameter decided by the decisionmaker to vary the importance of small weights and small divergence between the scores.
The final evaluation of the alternatives depends on an overall score, s_i:

$$s_i = \sum_{\substack{i'=1 \\ i' \neq i}}^{m} z_{ii'} - \sum_{\substack{i'=1 \\ i' \neq i}}^{m} z_{i'i}, \tag{18}$$

where

$$z_{ii'} = \begin{cases} 1, & \text{if } \mathrm{C}(X_i, X_{i'}) > \eta \text{ and } \mathrm{D}(X_i, X_{i'}) < \mu, \text{ where } \eta \text{ and } \mu \text{ are thresholds,} \\ 0, & \text{otherwise.} \end{cases}$$

In another procedure introduced by Massam and Wolfe (1979) a cardinal relation among the alternatives is constructed by a multidimensional scaling technique. In this procedure, a dissimilarity matrix is created to indicate the degree of dissimilarity among the alternatives. Using a multidimensional scaling technique, the decisionmaker is able to search, on the basis of iterative algorithms, for the best-fit configuration in a multidimensional space in which the order of the distances is consistent with that of dissimilarity indices among the alternatives. The alternatives are thus ranked according to the resulting multidimensional scale.

For the procedures in CA to be consistent with the preference structure, the ranking of alternatives based on equations (18) must be the same as that in MUT. This implies that s_i must be monotonically transformed from the functional value of the multidimensional value function $\mathrm{Z}^m(X_i)$. Put simply,

$$\mathrm{Z}^m(X_i) < \mathrm{Z}^m(X_{i'}) \Rightarrow s_i < s_{i'}. \tag{19}$$

The relation between w_j^c and w_j^m could be derived as in previous methods, given the scaling parameter α and the thresholds η and μ, by solving the following equations:

$$\mathrm{Z}^m(X_i) - \mathrm{Z}^m(X_{i'}) = \mathrm{F}(s_i - s_{i'}) \tag{20}$$

or

$$\sum_{j=1}^{n} w_j^m [\mathrm{V}_j^m(X_i) - \mathrm{V}_j^m(X_{i'})] = \mathrm{F}\left[\sum_{\substack{i''=1 \\ i'' \neq i}}^{m} (z_{ii''} - z_{i''i}) - \sum_{\substack{i''=1 \\ i'' \neq i'}}^{m} (z_{i'i''} - z_{i''i'}) \right], \tag{21}$$

where F is a monotonical transformation function.

It is difficult, therefore, to define a concrete meaning for the weights in CA because that meaning depends on the complicated condition that must be fulfilled in equation (21). The meaning of the weights in CA is so abstracted from preference concepts that it is inconceivable that decisionmakers can think about these weights in terms of what they actually mean as indicated in the above equations. The weights on concordance [equation (16)] are the trade-offs of each alternative being preferred to each other alternative for a given attribute compared to other attributes, given that the weights are transformed by an exponential parameter α, also chosen by the decisionmaker. For the discordance indices [equation (17)] the weights apply to the differences for each pair of alternatives between standardized scores of attribute level of an attribute, given a transformation by an exponential parameter. All of the above is contingent on preference resulting from equation (18) based on the difference between the number of alternatives to which a given alternative is preferred (meets thresholds) and the number of alternatives which are preferred to that alternative.

In the apartment example, the weights for distance and floor area depend, therefore, on the determination of the alternatives to be compared, the scaling parameter, the thresholds, and the monotonic transformation function. In most applications, the weights are simply referred to as the relative importance among attributes, but they are weights on very abstracted, dimensionless scales.

2.5 *Computation of equivalent alternatives (CEA)*

In computation of equivalent alternatives, the alternatives are directly compared in terms of attribute levels without any explicit transformation functions (Stokey and Zeckhauser, 1978). One alternative is first picked and transformed into a preferentially equivalent alternative by the decisionmaker making trade-off judgments among the attributes in terms of attribute levels. The transformation is an attempt to make at least one alternative dominant over at least one other, which means that all the attribute levels in the dominant alternative are greater than or equal to those in the dominated alternative. The dominated alternatives are discarded and another alternative is then selected from the remaining set. The process continues until all but one alternative has been eliminated. The remaining alternative is the best choice. When the numbers of alternatives and attributes are large, the selection of the attributes based on which the trade-off judgements are made and the sequence of alternatives to be compared become important factors of consideration for efficiency.

Suppose that the decisionmaker has selected two alternatives, X_i and $X_{i'}$, to be compared and that X_i is chosen as the promising alternative, which the decisionmaker thinks has the greater probability of being dominant. To test the dominance of the promising alternative, the decisionmaker has to transform the promising alternative by trading off attributes, raising attributes that are inferior compared with the other alternative, and compensating by lowering other attributes that are superior, thus creating a new alternative, indifferent in value terms from the original. If at least one of the attributes in the transformed promising alternative is preferred and all others are at least equal, the promising alternative dominates the other. Otherwise, additional trade-off judgments, using other attributes, must be made until one alternative dominates the other.

Mathematically, let the attributes in the promising alternative be partitioned into three subsets: preferred (x_{ij}^p), inferior (x_{ik}^q), and tied (x_{il}^r) attributes compared with the other alternative in the pair. Rearrange the order of the attributes as follows:

$$X_i = (x_{i1}^q, ..., x_{iq}^q, x_{i1}^p, ..., x_{ip}^p, x_{i1}^r, ..., x_{ir}^r) .$$

Consequently, the other alternative can be expressed according to the same terms,

$$X_{i'} = (x_{i'1}^q, ..., x_{i'p}^q, x_{i'1}^p ..., x_{i'q}^p, x_{i'1}^r, ..., x_{i'r}^r) ,$$

where $p+q+r = n$, the total number of attributes.

Because the preferred attributes in X_i are exactly the inferior attributes in $X_{i'}$, the two sets are interchanged in the two alternatives. That is, x_{i1}^q and $x_{i'1}^p$ refer to the same attribute. Suppose that the decisionmaker increases attribute x_{ij}^q by t_{ij} so that the resulting attribute level is equal to that in alternative $X_{i'}$ for that attribute, and so that the increase is compensated by the decrease of t_{ik} in x_{ik}^p, where $1 \leqslant j \leqslant q$, and $1 \leqslant k \leqslant p$. Assume that x_{ik}^p is selected so that the resulting attribute level $x_{ik}^p - t_{ik}$ is still preferred compared with that in alternative $X_{i'}$. Let the resulting transformed promising alternative after the first trade-off be denoted as X_i^1. The attribute vector of the transformed promising alternative is thus as follows:

$$X_i \sim X_i^1 = (x_{i1}^q, ..., x_{ij-1}^q, x_{ij+1}^q, ..., x_{iq}^q, x_{i1}^p, ..., x_{ik}^p - t_{ik}, ..., x_{ip}^p, x_{i1}^r, ..., x_{ir}^r, x_{ir+1}^r) ,$$

where the notation $\sim$ means the alternatives on both sides are preferentially equivalent and x_{ir+1}^r is equal to $x_{ij}^q + t_{ij}$.

Note that the attribute x_{ij}^q is tied with that in $X_{i'}$ after the increment t_{ij} and therefore is located in the partition of the tied attributes. This implies, in terms of value in the preference structure, that

$$Z^m(X_i) = Z^m(X_i^1) , \tag{22}$$

and, therefore,

$$w_j^m V_j^m(x_{ij}^q + t_{ij}) + w_k^m V_k^m(x_{ik}^p - t_{ik}) = w_j^m V_j^m(x_{ij}^q) + w_k^m V_k^m(x_{ik}^p) . \tag{23}$$

From equation (23), the trade-off value t_{ik} can easily be solved as follows:

$$t_{ik} = V_k^{m^{-1}} \left\{ \frac{w_j^m}{w_k^m} [V_j^m(x_{ij}^q + t_{ij}) - V_j^m(x_{ij}^q)] \right\} . \tag{24}$$

The decisionmaker then selects the next inferior attribute in the transformed promising alternative and trades it off with one of the preferred attributes until all the attributes, if possible, at least tie with those in the other alternative.

This procedure should lead to the best alternative that is also the top ranked by MUT because the decisionmaker makes trade-off judgments so that the overall values of these transformed alternatives X_i^n are identical. The main difference between the results of CEA and MUT is that CEA searches for the dominant alternative whereas MUT evaluates all the alternatives. CEA could rank all alternatives by repeating the process for each remaining subset.

Note that in our derivations all multidimensional value functions are additive, which implies that the attributes are preferentially independent of each other. That is, when making trade-off judgments for a certain pair of attributes, the decisionmaker does not consider the attribute levels of other attributes. We will discuss the case of preferential dependence in section 3. It is clear that CEA does not require the decisionmaker to assign weights directly to the attributes, but to make trade-off judgments among attributes in terms of attribute levels. Weights are implied in this process and these derived weights can be compared with those in MUT. The decisionmaker is asked to make explicit trade-offs that are unambiguous, but the empirical question remains as to whether a decisionmaker is able to do so. The implied weights are the willingness to trade off in terms of the original measures of attribute levels. These weights can be directly computed from MUT weights as in equation (2); in the nonlinear case the weights must be computed for a particular combination of levels of the attributes. It is, therefore, easy to determine what a decisionmaker who gives a set of MUT weights should imply as CEA weights.
In the apartment example, though the trade-offs on the original attribute scales of walking time, say 3 min and 2 min, for a certain amount of living space, say 20 ft^2,

are different at combinations of (−13 min, 150 ft^2) and (−10 min, 200 ft^2), respectively, the implied weights in terms of MUT value functions remain unchanged.

3 Discussion

Multiattribute techniques are usually presented by their advocates with a focus on their peculiarities, rather than their commonalities. Although each of the five methods presented here has particular characteristics that we have slighted, we have illustrated how a commonality of meaning can be achieved. In this section, we will discuss some important issues which the decisionmaker would encounter in using these methods, including linear unidimensional value functions, preferential dependence among attributes, and inconsistency of repeated judgments.

In the derivation of weights for WR, we have demonstrated that w_j^w are invariant across the alternatives only when the unidimensional value functions V_j^w are linear with respect to the unidimensional transformation functions T_j. If the unidimensional value functions V_j^m and the unidimensional transformation functions T_j are both linear, these weights are necessarily invariant across the alternatives. Consider equation (5). If V_j^m and T_j are both linear with respect to the attribute levels, V_j^w must necessarily be linear with respect to the transformed attribute levels from T_j.

The additive multidimensional functions Z^m, S, Z^a, and D in MUT, WR, AHP, and CA, respectively, assume that the attributes are mutually preferentially independent of each other. The procedure in CEA also implies preferential independence among the attributes. That is, for any attribute j, if x_{ij} is preferred to $x_{i'j}$, given certain attribute levels of other attributes, x_{ij} is always preferred to $x_{i'j}$, given any arbitrary attribute levels of the attributes. Mathematically, let Y' and Y'' be the attribute vectors of the complementary attribute set of attribute j, for all $x_{ij'}$, Y', Y'',

$$[(x_{ij}, Y') \gtrsim (x_{i'j}, Y')] \Rightarrow [(x_{ij}, Y'') \gtrsim (x_{i'j} Y'')] , \tag{25}$$

where $\gtrsim$ means that the alternative on the left side is preferred or indifferent to the alternative on the right side.

The additive multidimensional value function is the necessary and sufficient condition for mutually preferential independence among attributes (Keeney and Raiffa, 1976). This theorem immediately restricts the applications of these methods because the attributes must be shown to be or assumed to be mutually preferentially independent before the methods can be applied. In general, if attributes are not mutually preferentially independent, the best strategy is to define new surrogate measures of the attributes so that they are (Keeney and Raiffa, 1976).

We have assumed in the beginning the invariant and unique preference structure. In practice, this assumption is broken because the value judgments from decisionmakers are said to be labile (Fischhoff et al, 1980). There is a need, therefore, for a consistency check in the methods. Only MUT and AHP provide such devices. The simplest consistency check procedure in MUT is to ask the decisionmaker repeatedly to specify the preferences among alternatives to see whether the preferences are consistent with the results from the value functions (Keeney and Raiffa, 1976). In AHP, the consistency check is conducted by a consistency index indicating the deviation of the maximum eigenvalue derived from a pairwise comparison matrix from the ideal eigenvalue for that matrix, which is the number of the rows or columns of that matrix.

4 Empirical comparisons

The types of information entered into the methods also affect the quality of results. MUT and CEA ask trade-off questions, whereas WR, AHP, and CA ask about relative importance among the attributes. In each of the two sets, the purposes of the information expected by the methods are also different. The types of trade-off questions asked in MUT and CEA are similar but for different purposes. In MUT, the decisionmaker is asked to specify an imaginary alternative with certain given attribute levels to which he or she is indifferent compared with another alternative. In MUT the trade-off information is used to estimate the weights of the attributes in the multidimensional value function. In CEA, the decisionmaker has control in making the trade-off judgments in the sense of selecting any real alternative from the alternative set and any attribute to trade with. The trade-off information is used directly to rank the alternatives.

In asking questions about relative importance, the deductive meanings of the weights in WR, CA, and AHP are different. There could, therefore, be two possibilities of error. First, the decisionmaker may not respond correctly to the questions which the methods ask to elicit the weights. Second, the methods themselves may not ask the correct questions which, if asked appropriately, should result in the responses as in the deductive comparison. For example, simply asking about relative importance among attributes seems to be totally irrelevant to the complex definition of the weights in CA derived in section 2. As to WR and AHP, the question of relative importance is also too vague if the weights thus elicited are said appropriately to reflect the meanings as defined in the deductive comparison.

Given the clarity on how what the decisionmaker intends, or is presumed to intend, is manifested in each method, it is now possible to devise experiments to assess whether methods work in practice. The following experiments appear worthwhile. For realistic attributes and alternatives and particular types of problems, are the assumptions required likely to be met to a close enough approximation that the methods will be robust? Which methods will in practice be more robust than others? These questions can only be answered by addressing a large sample of real or clearly realistic problems with the different methods.

In the experiments, rather than testing decomposed multiattribute methods by their ability to replicate holistic judgments as has frequently been done (for example, by Schoemaker and Waid, 1982), it is more reasonable to test them by their ability to predict a decomposed judgment derived from neutral procedures. One way to obtain a neutral standard is to elicit judgments by means of one method (for example, AHP), then another (for example, MUT), and iterate between the two until agreement is reached. Agreement can be tested with the deductive transformations described in this paper. Given this neutral standard, experiments can be conducted to compare either the two methods in the iteration or a third method. Comparisons are, of course, among different attempts at judgment by the same subject using different methods. The judgments derived can also be tested by transforming them into the original attribute space and asking if the decisionmaker is satisfied with the implied trade-offs. This test and verification procedure must be based on the transformations in the deductive comparison which can be operationalized for the experiments, such as equations (6), (15), and (24).

5 Conclusions

The descriptions presented show not only common meanings, but gaps in stated assumptions that must be filled before meaningful experiments can occur. For example, is the decisionmaker in WR supposed to think about transformed attributes when estimating weights? A decisionmaker asked to assess inherent weights of importance will not necessarily provide the same response as one who is asked to weight scaled differences among the best and worst of the attribute levels for the given alternatives. The manipulations of preferences are clearly more transparent to the decisionmaker in some methods than in others. On the face of it one might expect that such methods would be easier for the decisionmaker to use effectively. No firm conclusions about techniques can be drawn, however, until experiments have been conducted. The framework presented here provides the transformation capability among the methods that is necessary in order to conduct valid experiments.

References

Belton V, 1986, "A comparison of the analytic hierarchy process and a simple multi-attribute value function" *European Journal of Operational Research* **26** 7-21

Fischer G W, 1977, "Convergent validation of decomposed multi-attribute utility assessment procedures for risky and riskless decisions" *Organizational Behavior and Human Performance* **18** 295-315

Fischhoff B, Slovic P, Lichtenstein S, 1980, "Knowing what you want: measuring labile values", in *Cognitive Processes in Choice and Decision Behavior* Ed. T Wallsten (Lawrence Erlbaum Associates, Hillsdale, NJ) pp 64-85

Hopkins L D, 1980, "Review of *Spatial Search: Applications to Planning Problems in the Public Sector*" *Regional Studies* **15** 67-68

Keeney R L, Raiffa H, 1976 *Decisions with Multiple Objectives* (John Wiley, New York)

Massam B H, Wolfe R A, 1979, "The multicriteria choice problem and plan evaluation: a test of a method" *Canadian Journal of Regional Science* **1** 11-24

Miller D, 1980, "Project location analysis using the goals achievement method of evaluation" *Journal of American Planning Association* **46** 195-208

Moroney M J, 1951 *Facts from Figures* (Cox and Wyman, London)

Saaty T L, 1980 *The Analytic Hierarchy Process* (McGraw-Hill, New York)

Schoemaker P J H, Waid C C, 1982, "An experimental comparison of different approaches to determining weights in additive utility models" *Management Science* **28** 182-196

Stokey E, Zeckhauser R, 1978 *A Primer for Policy Analysis* (W W Norton, New York)

Voogd H, 1983 *Multicriteria Evaluation for Urban and Regional Planning* (Pion, London)