Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Remarks on the Analytic Hierarchy Process

James S. Dyer,

MANAGEMENT SCIENCE
Vol. 36, No. 3, March 1990
Printed in U.S.A.

REMARKS ON THE ANALYTIC HIERARCHY PROCESS*

JAMES S. DYER
Department of Management Science and Information Systems,
The University of Texas, Austin, Texas 78712

The analytic hierarchy process (AHP) is flawed as a procedure for ranking alternatives in that the rankings produced by this procedure are arbitrary. This paper provides a brief review of several areas of operational difficulty with the AHP, and then focuses on the arbitrary rankings that occur when the principle of hierarchic composition is assumed. This principle requires that the weights on the higher levels of a hierarchy can be determined independently of the weights on the lower levels. Virtually all of the published examples of the use of the AHP to evaluate alternatives relative to a set of criteria have assumed this principle. The key to correcting this flaw is the synthesis of the AHP with the concepts of multiattribute utility theory.
(DECISION ANALYSIS, MULTIATTRIBUTE UTILITY THEORY; ANALYTIC HIERARCHY PROCESS)

1. Introduction

Over a decade has elapsed since Saaty (1977) introduced the analytic hierarchy process (AHP). During this time a number of applications have been proposed, but a number of criticisms of this approach have also appeared. In a recent paper Harker and Vargas (1987), referred to henceforth as HV, respond to criticisms of the AHP, and conclude that ". . . the acceptance of this method has been slowed by what we believe to be (a) misunderstandings of its theoretical foundations, and (b) a reluctance to move away from traditional methods of analysis . . .". We disagree. The AHP is flawed as a procedure for ranking alternatives in that the rankings produced by this procedure are arbitrary. This flaw can be corrected, but not by moving away from traditional methods of analysis. The key to the proper use of the AHP relies on its synthesis with the concepts of multiattribute utility theory.

The defense of the AHP by HV ranges over a number of topics, and we take exception to many of their arguments and conclusions. A point-by-point response to these arguments would become tedious, and detract from the major point of our observations. However, a few issues associated with the implementation of the AHP will be briefly mentioned and identified as topics for further discussion. To avoid unnecessary duplication, we shall assume that the reader is familiar with HV.

HV defend the AHP against criticisms of the ambiguity of the questions that the decision maker must answer by claiming that ambiguity is inherent in all preference elicitation methods, including those of classical utility theory. This argument is misleading, since the elicitation questions posed in classical utility theory are well defined, and depend on a choice among alternatives by the subject rather than on a subjective response on a ratio scale. This distinction is important, since the biases noted in responses to the elicitation methods of classical utility theory are subtle in nature, and have generated new insights regarding how individuals respond to risky decisions (e.g., see the review by Weber and Camerer 1987).

The elicitation questions associated with AHP have much more in common with the questions used to determine a strength of preference function, which require a subjective estimate of strength of preference on a cardinal scale. The inherent difficulty with direct

* Accepted by Robert L. Winkler, former Departmental Editor; received May 31, 1988. This paper has been with the author $3\frac{1}{2}$ months for 1 revision.

0025-1909/90/3603/0249$01.25

subjective estimates has long been recognized, and has been a major reason that a preference theory based on the concept of strength of preference has not been in favor in the literature (Fishburn 1970, 1988, Dyer and Sarin 1981, Farquhar and Keller 1988).

The AHP elicitation questions suffer from even more ambiguity than those of strength of preference questions, however, since they require the decision maker to implicitly or explicitly determine a 0.0 reference point on a ratio scale. Suppose a thoughtful person hears the question, "How much better is A_i than A_j on a criterion?" His appropriate response would be, "Relative to what?" This latter question expresses intuitively the need for the definition of the reference point. We are willing to agree with HV that a thoughtful decision maker aided by a skillful analyst could provide meaningful responses to the AHP elicitation questions in some cases. However, this is a potential source of error in the AHP procedure, and it cannot be swept aside as easily as HV imply.

Two other aspects of the AHP procedure defended by HV are the use of a scale of one to nine to aid the required ratio judgments, and the use of the eigenvector approach to average inconsistent responses. The one to nine scale has some obvious shortcomings (e.g., see Dyer and Wendell 1985), but it has also been used to advantage in a number of empirical studies. This evidence suggests that there are some special cases where this scale may be a useful elicitation aid for a careful analyst, but its use must be tempered with judgment.

The use of the eigenvector approach is not a major issue with respect to the more controversial aspects of the AHP. The objective of this part of the procedure is to average the inconsistencies in the responses, and there are a number of approaches that might be considered. As HV indicate, the eigenvector approach does have some advantages in this role, although Barzilai, Cook, and Golany (1987) have also made a strong case for the use of the geometric mean.

Each of the issues highlighted above represents an area where the AHP is still subject to legitimate criticisms, despite the strong defense by HV. However, they are issues that are primarily operational in nature, and do not represent a flaw in the basic methodology. We now turn to more substantive issues.

§2 contains a brief review of the axiomatic foundations of the AHP. In contrast to those of classical utility theory, these axioms do not offer an intuitively appealing foundation for the methodology that is subject to empirical tests. The major focus of this evaluation of the AHP is in §3 where we argue that the AHP, as it is traditionally applied to the evaluation of alternatives, generates rank orderings that are not meaningful with respect to the underlying preferences of the decision maker. A symptom of this deficiency is the phenomenon of rank reversal, which has been discussed in the literature for almost a decade. §4 presents several suggestions for modifications in the AHP, including some motivated by insights from multiattribute utility theory. This paper concludes in §5 with a brief summary.

2. The Axioms of the Analytic Hierarchy Process

HV claim that the AHP has been criticized because it lacks an axiomatic foundation, but that Saaty (1986) has now provided the necessary axioms to counter this deficiency. The axioms developed by Saaty (1986) were a significant contribution to the theory of AHP. However, HV miss an important point that can be explained by a comparison of Saaty's axioms with those of expected utility theory.

Several systems of axioms have been developed for expected utility theory that all produce, for practical purposes, the von Neumann-Morgenstern (1947) expected utility model (for a recent review, see Fishburn 1982). Most of these systems are variations on the three axioms due to Jensen (1967) which we present below for purposes of comparison.

Let $\succ$ be a binary relation on a nonempty mixture set of lotteries X, Y, and Z, interpreted as strict preference. The indifference relation and the preference indifference relation can be defined from $\succ$ as follows: $X \sim Y$ if neither $X \succ Y$ nor $Y \succ X$, and $X \gtrsim Y$ if either $X \succ Y$ or $X \sim Y$.

AXIOM 1. Complete ordering. *For any two lotteries X and Y, either* $X \gtrsim Y$ *or* $Y \gtrsim X$. *For any three lotteries X, Y, and Z, if* $X \gtrsim Y$ *and* $Y \gtrsim Z$, *then* $X \gtrsim Z$.

AXIOM 2. Continuity. *Given the lotteries X, Y, and Z such that* $X \succ Y \succ Z$, *there exists a probability p in* (0, 1) *such that* $Y \sim pX + (1 - p)Z$.

AXIOM 3. Independence. *If* $X \succ Y$, *then for any p in* (0, 1] *and any Z*, $pX + (1 - p)Z \succ pY + (1 - p)Z$.

If a decision maker's preferences over lotteries are consistent with these axioms, then they can be represented by expected utility.

Two observations are particularly important regarding this axiom system for expected utility theory. First, each of these axioms has a clear and obvious meaning as a description of choice behavior. Therefore, each axiom can be debated on the basis of its appeal as a normative descriptor of rationality, and each axiom can also be subjected to empirical testing. Many of the recent advances in expected utility theory have occurred as the result of efforts to weaken the requirement of Axiom 3, which has a strong normative appeal but is occasionally violated by decision makers in practice (see Fishburn 1988 for an elaboration).

The second observation is that the axioms apply only to the binary relation $\succ$, and the existence of a real-valued von Neumann-Morgenstern utility function is derived from this axiom set. Thus, it is not necessary to assume the existence of the preference function as a primitive notion in this theory. In order to extend the expected utility theory into a multiple attribute context and to obtain an additive functional form decomposible by attribute, one additional independence condition is required that again is subject to empirical testing by reference to the simple choice behavior of the decision maker (e.g., Keeney and Raiffa 1976).

Similar remarks apply to the axioms of utility theory for the case of certainty, where lotteries are not explicitly considered. That is, the axioms have an intuitive meaning that can be subjected to empirical tests, and the preference function is derived from these axioms rather than assumed as a primitive notion. Keeney and Raiffa (1976, Chapter 3) and Dyer and Sarin (1979) provide a review of these concepts.

In contrast, the axioms provided by Saaty (1986) fail to be motivated by testable descriptions of behavior. He begins with three "primitive notions" which include the assumption of a fundamental scale. Suppose $\mathcal{A}$ is a finite set of alternatives and $\mathcal{C}$ is a finite set of attributes, with elements A and C respectively. This primitive notion is stated by HV as follows:

FUNDAMENTAL SCALE. *Let P denote the set of mappings from* $\mathcal{A} \times \mathcal{A}$ *to* $R^+, f: \mathcal{C} \to P$, *and* $P_C \in f(C)$ *for* $C \in \mathcal{C}$. *Thus, every pair* $(A_i, A_j) \in \mathcal{A} \times \mathcal{A}$ *can be assigned a positive real number* $P_C(A_i, A_j) = a_{ij}$ *that represents the relative intensity with which an individual perceives a property* $C \in \mathcal{C}$ *in an element* $A_i \in \mathcal{A}$ *in relation to other* $A_j \in \mathcal{A}$:

$$A_i \succ_C A_j \quad \text{if and only if} \quad P_C(A_i, A_j) > 1,$$

$$A_i \sim_C A_j \quad \text{if and only if} \quad P_C(A_i, A_j) = 1.$$

Four axioms then follow that are primarily structural or definitional in nature. Thus, this axiom system is based on a "primitive notion" that essentially assumes the existence of the ratio scale, rather than deriving the existence of this scale from a set of axioms that are descriptive of rational behavior and subject to empirical tests. As a result, this axiomatization is reminiscent of Bernoullian utility theory which influenced economic theory in the second half of the nineteenth century (see the discussion by Fishburn 1988, Chapter 1).

The appeal of AHP would be strengthened by an effort to link its theoretical basis to that of classical preference theory, and by providing a more fundamental set of axioms descriptive of behavior that allow one to derive the existence of the ratio scale P_C. A natural link may be the relationship between difference and ratio measures that has been explored by Krantz et al. (1971, Section 4.4.3). Since the relationship between difference measurement and expected utility theory is now understood (see Dyer and Sarin 1979 and Sarin 1982), research in this direction could provide the basis for a synthesis of these methodologies.

3. Rank Reversal and Arbitrary Rankings

We now turn to a consideration of perhaps the most controversial aspect of AHP, the phenomenon of rank reversal. The difficulty can be simply stated as follows: The ranking of alternatives determined by the AHP may be altered by the addition of another alternative for consideration. This characteristic of the methodology has been well known for years, and has been discussed in a number of articles by critics and by proponents of the AHP (e.g., see Belton and Gear 1983, Dyer and Ravinder 1983, Kamenetzky 1982, Saaty and Vargas 1984A, B, Saaty, Vargas, and Wendell 1983, Watson and Freeling 1982, 1983, Saaty 1987, and Dyer and Wendell 1985). The real issue, however, is not the phenomenon of rank reversal per se. Rather, rank reversal is a symptom of a much more profound problem with the AHP: *the rankings provided by the methodology are arbitrary.*

3.1. *The Example of Belton and Gear*

HV present a discussion of the example of rank reversal that was developed by Belton and Gear (1983). In their example, three alternatives A, B, and C are compared against three criteria C_1, C_2, and C_3. The three alternatives are the following:

	C_1	C_2	C_3
A	1	9	8
B	9	1	9
C	1	1	1

When these three alternatives are evaluated by the AHP using the principle of hierarchic composition and assuming equal weights on the attributes, the rankings in order are B, A, and C (the computations are available in HV and are not repeated here).

Belton and Gear then add a fourth alternative, D, which is an exact copy of alternative B, and obtain the new ranking A, B and D (tie), and C. HV criticize this example because alternative D is a copy of B. They argue that Axiom 4 of the theory of the AHP developed by Saaty (1986) explicitly excludes copies from consideration. To quote from their discussion of this point,

> For example, were we comparing alternative A (e.g., a blue Mercedes) and alternative B (e.g., a Chrysler) and a copy of A with respect to the given criteria (e.g., a red Mercedes where color is *not* a criterion), this copy should be removed since the preferences of A versus B will automatically give us the preferences of B versus all copies of A by the indifference relationship $\sim_C$; i.e., A must be indifferent to all its copies.

They then conclude, "Thus, Belton and Gear's counterexample is vacuous when these facts are recognized."

The defense of the AHP on the grounds that copies should not be allowed as alternatives is without foundation, and cannot be supported on intuitive or on technical grounds. First, consider how one would expect a procedure for evaluating and ranking alternatives to behave. If a set of alternatives have been ranked by a procedure and a copy of one of them is added to the set, we would expect the procedure to rank this copy exactly the same as its matching alternative is ranked (which the AHP does), and to assign both the same rank that its matching alternative was assigned originally (which the AHP does not do). The only exceptions to this expectation would occur if the criteria include a concern about the "uniqueness" of an alternative, and the copy reduces the rating of both on this criterion, or if the copy provides additional information that changes the perception of the decision maker regarding the alternatives. In the example provided by Belton and Gear, neither of these considerations explains the reversal in the rankings. Rather, the reason for this reversal is inherent in the logic of the AHP.

To add further emphasis to this point, suppose we do agree that if we compare a Mercedes and a Chrysler using the AHP, then it is not appropriate to add another Mercedes to the set of alternatives. However, suppose we add a BMW which is "close" to a Mercedes on most criteria, but not a perfect copy. Certainly, we would not expect the addition of the BMW to reverse the rankings of the Mercedes and the Chrysler. But this is exactly what will often occur when a new alternative similar to another ranked previously by the AHP is added to the set.

To illustrate this phenomenon with the example provided by Belton and Gear, suppose we add alternative D that is similar to alternative B, but with scores of 8, 1, and 8 on criteria C_1, C_2, and C_3, respectively, rather than 9, 1, and 9 as used originally by Belton and Gear. The judgment matrices are

C_1	A	B	C	D	w
A	1	1/9	1	1/8	1/19
B	9	1	9	9/8	9/19
C	1	1/9	1	1/8	1/19
D	8	9	8	1	8/19

C_2	A	B	C	D	w
A	1	9	9	9	9/12
B	1/9	1	1	1	1/12
C	1/9	1	1	1	1/12
D	1/9	1	1	1	1/12

C_3	A	B	C	D	w
A	1	8/9	8	1	8/26
B	9/8	1	9	9/8	9/26
C	1/8	1/9	1	1/8	1/26
D	1	8/9	8	1	8/26

which yield the scores

$$w_A = 0.37, \quad w_B = 0.30, \quad w_C = 0.06 \quad \text{and} \quad w_D = 0.27$$

reversing rank between A and B as in the Belton and Gear example but without relying on a copy of B.

The fact that rank reversal also occurs with "near copies" has been recognized earlier (see Dyer and Ravinder 1983), and Saaty (1987) provides results that indicate how close a near copy can be to an original observation without causing a rank reversal. Rather than recognizing this phenomenon as an indicator of a potential source of error in the AHP, he suggests that we eliminate alternatives from consideration that score within 10 percent of another alternative! Thus, not only can we not compare two Mercedes against a Chrysler with the AHP, we find that we cannot compare a Mercedes and a BMW against a Chrysler (assuming that they are within 10 percent of one another).

3.2. *The Example of Dyer and Wendell*

HV next consider the example of rank reversal provided by Dyer and Wendell (1985), and shown below:

	Criteria			
Alternatives	C_1	C_2	C_3	C_4
A_1	1	9	1	3
A_2	9	1	9	1
A_3	8	1	4	5
A_4	4	1	8	5

Assuming that the four criteria are judged to be equally important, the rankings determined by the AHP for the first three alternatives are given by

	C_1	C_2	C_3	C_4	Score	Rank
A_1	1/18	9/11	1/14	3/9	0.320	3
A_2	9/18	1/11	9/14	1/9	0.336	2
A_3	8/18	1/11	4/14	5/9	0.344	1

and for the four alternatives by

	C_1	C_2	C_3	C_4	Score	Rank
A_1	1/22	9/12	1/22	3/14	0.264	1
A_2	9/22	1/12	9/22	1/14	0.243	4
A_3	8/22	1/12	4/22	5/14	0.246	2
A_4	4/22	1/12	8/22	5/14	0.246	2

where the alternatives A_1 and A_3 have reversed rankings.

These numbers are reproduced here because a careful inspection of the results will provide some insights into the flaw in the AHP when the principle of hierarchic composition is assumed (i.e., the weights on the criteria do not depend on the alternatives under consideration). In order to obtain the scores shown above, the numbers in the columns are each multiplied by 0.25, reflecting the assumption that the criteria are equally important, and summed across the rows.

The normalized eigenvectors determined in the AHP essentially allocate the weights on the criteria to the alternatives. For example, when only the three alternatives A_1, A_2, and A_3 are considered, 8/18ths of the weight on criterion C_1 is allocated to alternative A_3 by the first term in the calculation of the AHP score for A_3, and 9/11ths of the weight on criterion C_2 is allocated to A_1 by the second term in its AHP score calculation. Alternative A_3 has higher scores than A_1 on criteria C_3 and C_4 which are enough to give it a higher ranking than A_1.

Now, consider what happens when alternative A_4 is introduced. Alternative A_4 does moderate to well on criteria C_1, C_3, and C_4 which is where A_3 gained most of its allocated score in the case of three alternatives. Therefore, it dilutes the allocation of the scores of these criteria. Since alternative A_1 performed rather poorly on these criteria, it did not suffer significantly because it had such a small proportion of this weight initially. However, A_4 has poor performance on criterion C_2 where A_1 excels, so the fraction of the weight of C_2 allocated to A_1 falls from 9/11 to only 9/12. As a result, alternative A_3 suffers from the introduction of alternative A_4 much more than A_1, and a preference reversal occurs.

Clearly this example is not explained away by the argument that A_4 is a copy or "near copy" of another of the original alternatives. Further, Dyer and Wendell provide an interpretation of this example which leads to the conclusion that the appropriate ranking of the alternatives is obtained by simply summing the criterion scores, which produces a ranking in the order A_2, A_3 and A_4 (tie), and A_1. This ranking was not given by either of the applications of the AHP illustrated above.

This example does assume the principle of hierarchic composition, defined by Saaty (1980) to be applied in situations where the weights on the higher levels of a hierarchy can be determined independently of the weights on the lower levels. It is important to emphasize that virtually all of the published examples of the use of the AHP to evaluate alternatives relative to a set of criteria have assumed this principle. The principle of hierarchic composition is also assumed in the implementation of the AHP in the decision support software Expert Choice. Thus, we are led to the following conclusion:

When the principle of hierarchic composition is assumed, the results produced by the AHP are arbitrary.

While the results from the AHP may be highly correlated with the true preferences of a consistent decision maker in some cases, and even provide an accurate ranking of alternatives in other special cases, in general this cannot be guaranteed *a priori.*

4. Resolution

Arbitrary rankings are produced by the AHP when the principle of hierarchic composition is assumed because the weights on the criteria may not be appropriate given the normalization procedure used on the scores of the alternatives. Here we consider some proposed approaches to resolving this problem and show, in particular, how the method can be modified (using ideas from multiattribute utility theory) to correct this flaw.

4.1. *Altering the Weights*

It is possible to alter the weights of the attributes in the AHP in order to avoid rank reversal. However, this means that the questions of paired comparisons involving the criteria must depend on the specific alternatives being considered.

HV attribute the example of rank reversal provided by Dyer and Wendell to a misuse of the theory of the AHP, and claim that the weights in this example should depend on the alternatives. As a result of this insight, the principle of hierarchic composition is violated, and the system with feedback approach of the AHP must be used which requires the construction of a "super matrix" W. This matrix is composed of the relative weights of alternatives according to the criteria, and of the relative weights of criteria according to the alternatives. The final set of weights for the criteria and scores for the alternatives is given by $\lim_{k\to\infty} W^{2k+1}$.

An obvious concern is the number of ratio comparisons required by this supermatrix. Suppose n alternatives are evaluated on m criteria. Then the decision maker would be required to make $m(n^2 - n)/2 + n(m^2 - m)/2$ comparisons. A case with seven alternatives and five criteria, which is not a particularly large problem for a real application, would require 175 ratio comparisons! Therefore, the number of comparisons required by this procedure would place a restrictive limitation on its actual use, although it may be possible to reduce this number in practice by using only a subset of the ratio comparisons ordinarily assessed.

The problem of the large number of required ratio comparisons is compounded by the difficulty of responding to the questions requiring that the criteria be compared with respect to the alternatives. HV refer to a trivial case where this question can be answered because the criteria are all measured in monetary returns. However, in an automobile selection problem with cost and appearance as two of the criteria, the decision maker would be required to respond to questions such as this, "Does the Mercedes perform better on cost or on appearance, and by how much?"

Responses to these types of questions may not be meaningful, even when the decision maker is aided by a skillful analyst. Recall that the questions required by the AHP have been criticized as being ambiguous when two alternatives are compared on the same criteria. At least in these cases the alternatives are measured on the same metric. Ratio comparisons of the performance of an alternative across criteria would require much more complex cognitive comparisons by the decision maker (see the discussion by Kleinmuntz and Schkade 1988).

HV state that the supermatrix approach should be used in cases where the principle of hierarchic composition is not valid, and so the criteria weights are dependent on the alternatives. However, the AHP theory does not include any "independence conditions" that can be tested empirically based on the responses of the decision maker to determine *a priori* when the principle of hierarchic composition is not valid. Rather, HV seem to conclude that the principle must have been violated in the example provided by Dyer and Wendell because the addition of another alternative caused a rank reversal *and* the new alternative was not a copy of another alternative.

When the AHP is applied to the problem of evaluating alternatives on multiple criteria, the principle of hierarchic composition is *always violated.* The supermatrix approach requires responses from the decision maker that are numerous and ambiguous. Therefore, this does not seem to be a fruitful approach to resolving the difficulties inherent in applying the AHP at this time, although further research may provide the basis for overcoming these limitations.

4.2. *Using Absolute Measurement*

Saaty (1987) provides an example of the use of absolute measurement to compare alternatives which has the advantage that the rankings are not affected by the addition or the deletion of alternatives. In this approach, the AHP is used to assign scores to ratings on the criteria, such as "high", "average", and "low", and then alternatives are evaluated by assigning a rating to the performance of the alternatives on each criteria. This approach is illustrated with an example from the college admissions decision. The two criteria are grades and entrance exams, and the ratings on each are high, medium, and low.

Unfortunately, this approach is based on the assumption of the principle of hierarchic composition, which implies in this case that the weights on the two criteria are independent of the ratings used to measure performance on them. In general, we have seen that this is not true. Therefore, the rankings produced by this approach will be arbitrary, even though these rankings will not change when new alternatives are added or deleted. However, if a new rating category is introduced on one criterion, such as "above average", and the AHP is used to assign new scores to the four possible ratings on this criterion, then the rankings of the alternatives are likely to change even if none of them are assigned this new rating of "above average" and no other ratings are changed. The problem of a reversal in rankings is only a symptom of the more fundamental problem that the rankings produced by the AHP are arbitrary.

4.3. *Rescaling the Attribute Ratings*

A remarkable observation regarding this controversy about the phenomenon of rank reversal in the AHP is that a simple solution does exist. The key to resolving the problem is by analogy to multiattribute utility theory. This solution is appropriate as long as the criteria used in the application of the AHP satisfy the property of difference independence (Dyer and Sarin 1979), which can be tested *a priori*.

The key is to ensure that both the weights on the criteria and the scores of the alternatives on the criteria are normalized with respect to the same range of alternative values (e.g., see Keeney and Raiffa 1976, p. 273). This can be accomplished in one of two ways:

1. The weights on the criteria can be obtained in the traditional AHP manner. Then the decision maker can be asked to specify the ranges over which he assumed the alternatives vary on each criterion. For example, if he is trying to choose an automobile, he might indicate that he assumed cost could vary from $10,000 to $25,000, or appearance from beautiful to ugly. If the actual alternatives to be considered do not have performance measures that cover this entire range, "dummy" alternatives should be generated and added to the set under consideration to span these ranges. For example, a "best" alternative could be created with the most preferred performance on each criterion, and a "worst" alternative could be created with the least preferred performance on each criterion. The evaluation of the alternatives could then proceed in the usual AHP fashion.
2. As an alternative, the decision maker could be told the ranges over which the alternatives under consideration actually vary. Then he could be asked to answer the pairwise comparisons regarding the importance of the criteria by considering the relative importance of a change from the least preferred to the most preferred values for criterion i compared to a similar change for criterion j.

The eigenvectors for the criteria determined by the AHP should then be scaled by subtracting the smallest component in each eigenvector from all components in the eigenvector, and then dividing this modified eigenvector by its largest component. An example of the use of this approach is provided by Dyer and Wendell (1985). This is essentially the same strategy suggested by Kamenetzky (1982), and it resolves the flaw in the AHP of arbitrary rankings and its symptom, rank reversal. In addition, the rankings obtained by the AHP from the responses of a consistent decision maker will then be the same as those obtained by using an additive measurable multiattribute value function (Dyer and Sarin 1979), an additive multiattribute value function (Keeney and Raiffa 1976, Chapter 3), and an additive multiattribute utility function (Keeney and Raiffa 1976, Chapter 6) when the appropriate independence conditions are satisfied.

5. Conclusions

Applications of the AHP based on the assumption of the principle of hierarchic composition produce rankings based on the consistent responses of a decision maker that cannot be shown to be consistent with his or her preferences. The defense of this basic flaw by HV and by Saaty (1987) does not provide an effective resolution of this problem.

As noted above, the actual solution to this problem is relatively simple, and is based on a synthesis of the AHP assessment methodologies with the theory of multiattribute utility theory (MAUT). Further, these ideas have been published previously, but have not been adopted by AHP researchers. The primary hindrance to the adoption of ideas from the AHP methodology by the OR/MS community has been the effort to maintain the AHP as a separate and distinct approach to the evaluation of alternatives. To quote from HV, "One should also not view the AHP as a subset of or perturbation to the methods of decision analysis such as MAUT. The main controversy in the decision analysis community has arisen by not recognizing that the AHP is based on an entirely different set of axioms." This same attitude also motivated the following comment by a reviewer of an earlier draft of this paper, "There is no doubt in my mind that we are wasting time for reasons other than science. Could utility theory be in such need as to seek synthesis with the AHP? The question posed by many: Does the AHP, as a ratio scale approach, need utility theory and for what purpose? I have not heard AHP practitioners voice such an opinion, and many of them are very knowledgeable about utility theory." We consider these to be most unfortunate positions.

At a higher level of abstraction we assume that researchers in "utility theory" and in AHP methodologies are attempting to model the preferences of a decision maker so that the rankings of alternatives produced by these approaches reflect these preferences. Therefore, it seems reasonable to conclude that both fields would benefit from efforts to synthesize these two approaches to the same problem. As an example, the problem of assessing a preference function in practice is challenging, and the OR/MS community would benefit from having alternative approaches available to deal with different situations and different individuals, and for consistency checks. Farquhar (1984) has surveyed lottery-based approaches for assessing risky preference functions, and Farquhar and Keller (1988) have provided a similar survey of strength-of-preference methods for assessing riskless value functions. The work by Dyer and Sarin (1979) and Sarin (1982) provides the conditions under which lottery-based and strength-of-preference assessment techniques may be used interchangeably, greatly expanding the flexibility of an analyst in selecting an assessment procedure and in verifying the result.

A synthesis of utility theory and the AHP would provide similar benefits, perhaps suggesting ratio scale assessment procedures that could be used within the context of traditional utility theory applications. A recent paper by Vargas (1986) provides a contribution to this direction of research. MAUT concepts provide a simple solution to the rank reversal problem in the AHP, and may also provide the insights necessary to allow the AHP assessment techniques to be applied to nonadditive multiattribute evaluation models, and to problems explicitly involving risk in the form of probability assessments.

We conclude that much more is to be gained from a synthesis of AHP and MAUT than from efforts to maintain them as separate areas of research and application.

References

BARZILAI, J., W. D. COOK AND B. GOLANY, "Consistent Weights for Judgements Matrices of the Relative Importance of Alternatives," *Oper. Res. Lett.*, 6, 3 (1987), 131–134.

BELTON, V. AND T. GEAR, "On a Short-coming of Saaty's Method of Analytic Hierarchies," *Omega*, 11, 3 (1983), 228–230.

——— AND ———, "The Legitimacy of Rank Reversal—A Comment," *Omega*, 13, 3 (1985), 143–144.

DYER, J. S. AND H. V. RAVINDER, "Irrelevant Alternatives and the Analytic Hierarchy Process," Working Paper, The University of Texas at Austin, 1983.

——— AND R. SARIN, "Measurable Multiattribute Value Functions," *Oper. Res.*, 27, 4 (1979), 810–822.

——— AND ———, "Relative Risk Aversion," *Management Sci.*, 28 (1981), 875–886.

——— AND R. E. WENDELL, "A Critique of the Analytic Hierarchy Process," Working Paper 84/85-4-24, Department of Management, The University of Texas at Austin, 1985.

FARQUHAR, P., "Utility Assessment Methods," *Management Sci.*, 30, 11 (1984), 1283–1300.

——— AND R. KELLER, "Preference Intensity Measurement," Decision Research Program, Technical Report 88-2, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA, 1988.

FISHBURN, P., *Utility Theory for Decision Making*, Wiley, New York, 1970.

———, *The Foundations of Expected Utility*, Reidel, Dordrecht, 1982.

———, *Nonlinear Preference and Utility Theory*, The Johns Hopkins University Press, Baltimore, 1988.

JENSEN, N. E., "An Introduction to Bernoullian Utility Theory. I. Utility Functions," *Swedish J. Economics*, 69, (1967), 163–83.

HARKER, P. T. AND L. G. VARGAS, "The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process," *Management Sci.*, 33, 11 (1987), 1383–1403.

KAMENETZKY, R. D., "The Relationship Between the Analytic Hierarchy Process and the Additive Value Function," *Decision Sci.*, 13 (1982), 702–713.

KEENEY, R. AND H. RAIFFA, *Decisions with Multiple Objectives*, John Wiley and Sons, New York, 1976.

KLEINMUNTZ, D. N. AND D. A. SCHKADE, "The Cognitive Implications of Information Displays in Computer-Supported Decision Making," Working Paper, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 1988.

KRANTZ, D. H., R. D. LUCE, P. SUPPES AND A. TVERSKY, *Foundations of Measurement*. Vol. I, Academic Press, New York, 1971.

SAATY, T. L., "A Scaling Method for Priorities in Hierarchical Structures," *J. Math. Psychology*, 15 (1977), 234–281.

———, *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.

———, "Axiomatic Foundations of the Analytic Hierarchy Process," *Management Sci.*, 32 (1986), 841–855.

———, "Rank Generation, Preservation, and Reversal in the Analytic Hierarchy Process," *Decision Sci.*, 18 (1987), 157–177.

——— AND L. G. VARGAS, "Inconsistency and Rank Preservation," *J. Math. Psychology*, 18 (1984a), 205–214.

——— AND ———, "The Legitimacy of Rank Reversal," *Omega*, 12 (1984b), 513.

———, ——— AND R. WENDELL, "Assessing Attribute Weights by Ratios," *Omega*, 11, 1 (1983), 9–13.

SARIN, R., "Strength of Preference and Risky Choice," *Oper. Res.*, 30 (1982), 982–997.

VARGAS, L. G., "Utility Theory and Reciprocal Pairwise Comparisons: The Eigenvector Approach," *Socio-Economic Planning Sci.*, 20, 6 (1986), 387–391.

WATSON, S. R. AND A. N. S. FREELING, "Assessing Attribute Weights," *Omega*, 10, 6 (1982), 582–585.

——— AND ———, "Comment on: Assessing Attribute Weights by Ratios," *Omega*, 11, 1 (1983), 13.

VON NEUMANN, J. AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, N.J., 1947.

WEBER, M. AND C. CAMERER, "Recent Developments in Modelling Preferences Under Risk," *OR Spektrum*, 9 (1987), 129–151.