Rank and the Controversy About the Axioms of Utility Theory
A Comparison of AHP and MAUT

Thomas L. Saaty
University of Pittsburgh

Abstract

A variety of authors have given real life examples that violate the utility theory axiom of independence from irrelevant alternatives. A theory that can only rank alternatives one at a time such as Multiattribute Utility Theory has difficulty showing how introducing new alternatives that do not add new criteria but affect the weights of the existing criteria as in problems of scarcity, can lead to rank reversal. The Analytic Hierarchy Process deals both with situations in which rank should be preserved and those when it should not. The paper also deals with problems raised against Multiattribute Utility by its own practitioners and with differences between it and the Analytic Hierarchy Process.

> The voice of the majority does not afford a
> proof of any value in truths a little
> difficult to discover, because such truths are
> much more likely to have been discovered by
> one man than by a nation.
>
> René Descartes
> Discourse on Method

1. Introduction

Independence from irrelevant alternatives is an axiom of utility theory, but what is independent and what is irrelevant? If one brings an additional alternative into a decision, should its presence ever affect the order of the old alternatives? The axiom says that it should not if it is "irrelevant". If it does, even though it is irrelevant, then we have a paradox because the axiom is contradicted. Utility theory followers probably prefer not to face such paradoxes but the matter needs to be brought into the open because there are other theories, for example, the Analytic Hierarchy Process, that do not have this axiom and do not suffer from these paradoxes. To utility practitioners, no decision

theory is valid if it does not subscribe to their axioms. In the Analytic Hierarchy Process irrelevant alternatives are not a part of the problem, nor can they be, because in making paired comparisons one would ordinarily compare them with existing relevant ones and assign them some importance, thus contradicting their irrelevance.

The problem discussed in this paper is : given that a set of alternatives have been ranked separately on each of several criteria and from these an overall ranking is derived. Suppose that a new alternative is added to the collection, can it affect and even reverse the old ranking? No one would argue that this can happen if a new criterion is introduced by the alternative, or the old ranking of the criteria themselves is changed. But what if none of these happens, can rank reversal occur? If no new criteria are introduced and alternatives are rated independently so that criteria weights established in advance are unaffected by the way the alternatives are evaluated, there can be no rank reversal at least in theory. In fact there are practical examples in which this happens and we need another way than ranking the alternatives one at a time to show this effect. One of the main objectives of this paper is to show that many people have given examples to show that rank reversal occurs naturally and is not accounted for in utility theory.

While surfacing the correct rank from judgment input data is important for any decision theory, it is also of capital importance that, given the input numbers, one can see why the ranks come out as they do. With the AHP, understanding preference of A over B as the dominance of A over B, one can derive the ranks of A and B from this dominance simply with the use of mathematics with no further assumptions.

Often decision making is not a matter of simply ranking alternatives once. Sometimes one must rank once for benefits and once for costs and then combine these two different rankings. Thus a decision may require deriving more than one set of rankings for

the same set of alternatives.

In general, rank is not an inherent property of objects. Rather, it is a subjective human construct. It is how we want to order the world either because we want it to conform to the standards we create from our personal experience, if we have a well defined one, or because we have the ability to make cogent relative comparisons of objects on properties they have in common based on observation without invoking the norms of our previous experience. These two ways can lead to different rankings. In spite of our desire to always preserve rank, our chaotic world is not set up to oblige us and in fact does not preserve rank in all real life situations. Given that this is the case, it is sophistry to insist that rank should always be preserved in the decision making process.

2. The Debate

The debate about rank is occurring because of differing assumptions of the decision theories, the Analytic Hierarchy Process (AHP) and Multiattribute Utility Theory (MAUT). In MAUT, the assumption is that decision making situations are always normative, that is, that somewhere in the background standards or norms always exist. The standards are uncovered by computing utility functions based on lottery comparisons. In other words, in choosing among restaurants, even though one may not be a connoisseur, a utility theorist would somehow ask that person the sort of questions that would enable him or her to construct utility functions as effectively as if they were designed by gourmets who make it their business to travel around sampling food and learning how to establish norms about the quality of food, the ambiance and all the factors necessary to make a normative decision. With MAUT the standards are thus set not only for the specific restaurants currently under consideration, but also for all others which may come along later, even though the decision maker may have no special expertise. The problem is that consideration of a new

restaurant may change the standards of one's preferences for restaurants especially if the decision maker has not been to many restaurants. However, utility theorists think long term experience is not needed to establish exacting norms, but can be replaced by asking sufficiently perceptive questions and computing functions from the answers.

The question one should ask is: How many restaurants can be added or deleted without changing the normative standards that have been derived? It is likely that with sufficient imagination one would say that even one restaurant can change the norms and the outcome of the decision. Even here rank preservation is illusory.

However, in the AHP, restaurants are compared in pairs according to the intensity with which one element of a pair possesses an attribute over another element thereby obtaining a ranking based on relative comparisons.

The AHP accepts as biologically given the proposition of cognitive psychologists that the mind has the dual capability to compare things with respect to a single property (criterion) against standards learned through experience one at a time, or to compare them against each other in pairs, again with respect to a single property, when there are no standards [3].

In the AHP, comparing alternatives one at a time with respect to standards or norms is called absolute measurement. It is the normative approach of the AHP in describing what some people would do if they have the knowledge to impose norms. Absolute measurement requires knowledge and experience to compare the intensities (excellent, very good, good, average, poor; or high, medium, low, and the like) with which a criterion is perceived to occur in an alternative. Otherwise establishing weights for intensities could become an artificial undertaking. Here rank is preserved because adding a new alternative does not affect the individual ranking of the old alternatives, since standards have already been set.

When one does not have knowledge and experience to establish

norms, one can, instead of setting standards, use the relative measurement mode to compare an alternative with another alternative with respect to an attribute to determine the degree to which it possesses that attribute according to a scale of dominance values. If it is decided in advance not to revise the weights of the criteria, then by using an ideal alternative (dividing by the weights of the highest ranked one for each criterion) there would be no rank reversal. Otherwise, adding new alternatives can logically, not randomly, create a new ranking. This is essentially the so-called rank reversal that is of concern to critics of the AHP. This reversal may happen even though neither new criteria are introduced nor are the weights of the old criteria revised directly, but rescaled by the alternatives and hence allowed to change. This is useful for dealing with scarcity and other situations such as phantom alternatives.

This type of rank reversal is not limited to AHP. It happens in other areas such as linear programming (LP). Adding variables (alternatives) in LP may lead to completely different solutions that are not extensions of the previous solutions, that is, the part of the solution corresponding to the old problem is not necessarily the same. Rank reversal is a constant companion of linear programming and goal programming, particularly of concern to people who use them in multicriteria decision making. Adding variables changes the dimensionality of a problem and its consequent optimality of vertices. In linear programming, any variable can have an effect on the solution, no matter whether one calls it relevant or irrelevant.

A number of people use linear and goal programming as single and multicriteria decision making methods. By extending the variables and constraints, there is what corresponds to rank reversal in these methods. My colleague Professor Jerrold May gives the following example:

Maximize $x_1 + 2x_2$, subject to $x_1 + x_2 \leq 1$, $x_1 \geq 0$, $x_2 \geq 0$. There are three vertices to consider $(0,0),(1,0)$ and $(0,1)$. The optimal solution is at $(0,1)$.

Now augment the problem with a new variable so that it becomes :

Maximize $x_1 + 2x_2 + bx_3$, $b > 0$, subject to $x_1 + x_2 + x_3 \leq 1$, $x_i \geq 0$, $i = 1,\ldots,3$. The feasible set of vertices is $(0,0,0)$, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. The optimum solution is at $(0,0,1)$ if $b > 2$ which when restricted to the old problem gives $(0,0)$ and not $(0,1)$ and thus adding a variable does not preserve the old optimal solution.

This example is easy to generalize to any number of variables. When linear programming is used in multicriteria decisions as in parametric programming, the same thing applies in going from one efficient solution (extreme point) to another, with rare exceptions at convex combinations on the efficient surface.

To see how important relative measurement is, one can give various examples to show that for the same problem, absolute and relative comparisons are both compelling, and the decision maker will decide whether he knows enough to establish a priori standards or whether he should use relative measurement to rank the alternatives. This can lead to different results depending on who is the decision maker. Here is an example. The Federal Drug Administration (FDA) with its experience and guidelines would apply standards to rate experimental drugs to be tried on terminal AIDS patients. It would be mostly concerned with the overall benefit to society at large and ban the use of a toxic drug though some few patients might be saved by using it. A terminal patient, for lack of knowledge about drugs, would not have standards as the FDA does to rate the drugs one at a time, and in that case he or she would prefer to compare the drugs against each other in a relative framework due to the perceived urgency of the situation. The individual might rank a more toxic drug first because of its greater effectiveness. I believe that people behave this way and the situation described above is a good example of how the situation determines the approach.

3. Rank Reversal in MAUT - Irrelevant Alternatives Questioned by Utility Experts

MAUT has a single procedure to use to rank alternatives, no matter what, the intention being to always preserve rank. Even if the procedure were to succeed in preserving rank all the time, the question would arise as to whether it correctly ranks alternatives in all situations. The answer is no. Long ago, counterexamples

92

showed that there are situations where rank should naturally be allowed to reverse. While it is desirable to have a procedure that can force rank preservation, not all situations should be treated by the same procedure. There must be a mathematical way to allow for either rank preservation or rank reversal without too much talk and justification on the side. After all, is not one of the purposes of using mathematics, is to organize and clarify thinking on difficult subjects? Rank reversal that occurs in the practice of MAUT cannot be pushed aside by philosophical arguments.

Granted that in some cases rank reversal is considered legitimate because adding a new alternative may introduce a new criterion or give rise to a change in the criteria weights but what if there is no new information to change the criteria and if the mathematical approach does not show how the weights of the criteria change carefully with the new alternatives? Are there examples in MAUT that lead to rank reversal? Some of the examples below are of this kind, by the admission of utility theorists.

Marketing literature is full of consumer preference rank reversals. An example used by my colleagues (teachers who know nothing about AHP but know MAUT) in the classroom is the case of a product that sells for example at $300, but is considered too expensive and is less preferred than one that sells for $150. A strategy that is often implemented is to introduce a product similar to the $300 one, but priced much higher, e.g., $1,000. It is well documented that the formerly most expensive $300 product then looks like a much better value and is preferred.

A second example deals with the introduction of a phantom alternative. Such an alternative might be priced slightly lower but is not made available. The consumer will perceive the existing alternative as an overpriced one despite its better quality and will select a cheaper one that is also available. The experts (who are not AHP practitioners) believe that normative utility theory fails to account for such behavior. However these can be explained with relative measurement.

A good example of rank reversal with no change of criteria or their weights is provided by Forman [10]. At a company, Susan excels over Jack in every quality except computer skills, where he is much needed and thus he is overall preferred to Susan. When the company hires John, who also has computer skills but is not as good as Jack, Susan becomes the preferred employee and there is reversal in Jack and Susan's rank. One can see that if one additional computer literate employee is not sufficient to reverse Jack and Susan's rank, surely when enough people are hired with computer skills, at some point such rank reversal would happen.

Regularity is a condition of choice theory that has to do with rank preservation. R. Corbin and A. Marley [5] provide an example that, "concerns a lady in a small town, who wishes to buy a hat. She enters the only hat store in town, and finds two hats, a and b_1, that she likes equally well, and so might be considered equally likely to buy. However, now suppose that the sales clerk discovers a third hat, b_2, identical to b_1. Then the lady may well choose hat a for sure (rather than risk the possibility of seeing someone wearing a hat just like hers), a result that contradicts regularity." We show how relative measurement of the AHP explains this utility theory paradox.

Here is a plausible set of judgments in the case of the two hats A and B, with A preferred to B. Adding C that is a copy of A, changes the preference to B over A with respect to rareness and thus makes it the more desired choice overall.

Style (.4)		A	B		Rareness (.6)		A	B		Style (.4)		A	B	C		Rareness (.6)		A	B	C	
A		1	3	.75	A		1	1	.5	A		1	3	1	.42	A		1	1/6	1	.125
B		1/3	1	.25	B		1	1	.5	B		1/3	1	1/3	.16	B		6	1	6	.75
										C		1	3	1	.42	C		1	1/6	1	.125

A = .75 x .4 + .5 x .6 = .6   A = .42 x .4 + .125 x .6 = .238
B = .25 x .4 + .5 x .6 = .4   B = .16 x .4 + .75 x .6 = .514
                                     C = .42 x .4 + .125 x .6 = .238

Again let us account for the effect of phantom products. Two products A and B are evaluated according to quality (Q) and price

(P). A is 3 times more expensive than B (or simply B has a more desirable price) but it is of considerably better quality than B.

Q	A	B
A	1	5
B	1/5	1

P	A	B
A	1	1/3
B	3	1

We assume that Q and P are equally important, and obtain the following priorities :

$$W_A = .542$$
$$W_B = .458$$

Let C be a product similar to A in quality and on price dominates B six time more. The addition of C makes A look overpriced and a consumer may decide that B is the best choice.

Q	A	B	C
A	1	5	1
B	1/5	1	1/5
C	1	5	1

P	A	B	C
A	1	1/3	2
B	3	1	6
C	1/2	1/6	1

and the overall priorities are given by:

$$W_A = .338$$
$$W_B = .379$$
$$W_C = .283$$

When alternatives are rated one at a time, we do not have a good way to measure how they affect the weights of the criteria when that is the way needed to account for rand reversals. After adding several copies of the same alternative, one is unable to make the minor distinction to change them, and yet as the number of copies of alternatives increases, rank does change. The number itself becomes a controlling criterion. Sometimes one needs to examine the effect of an increase and change in this number to see what can happen to rank on an ongoing basis.

The next question to ask is whether instead of copies, similar alternatives can affect a decision. Further and more generally, the number of any set of alternatives can affect the outcome of a decision. In the hat example the one change in number caused the lady to revise her judgment and choose the previously less preferred hat. Relative measurement would account for this [23] and for what would happen if the number continued to change. A

method that always preserves rank is oblivious to such happenings, justifying them not with a mathematical but with a philosophical theory and even then not with a fully and exhaustively worked out argument to satisfy all those who have counterexamples. No mathematician would support an axiom that is contradicted by counterexamples. That is where the confusion stems from. In behavioral systems, for any law one attempts to establish, one often discovers that its contrary can also be true. One soon learns that absolutes are almost nowhere! There is no point to being normative in all situations.

Zeleney [43] gives an example which contradicts the utility theory axiom, "if an alternative is nonoptimal it cannot be made optimal by adding a new alternative to the problem". The following table of lotteries with two possible outcomes gives the payoffs with probabilities p_1 and p_2, corresponding to five alternatives A_i.

	p_1	p_2
A_1	4	2
A_2	6	4
A_3	4	3
A_4	6	0
A_5	3.3	3.3

Consider the alternatives A_1 and A_2 only. A_1 dominates A_2 for all possible combinations of outcomes because

$$4p_1 + 2(1 - p_1) > 6p_1 + 4(1 - p_1).$$

A_3 and A_4 dominate each other for different values of p_1 and are equal at $p_1 = .6$.

A_3 dominates A_4 for $0 \leq p_1 < .6$ and

A_4 dominates A_3 for $.6 < p_1 \leq 1$.

If p_1 is chosen at random, A_3 is the winner 60% of the time and would be chosen over A_4. Consider now A_3, A_4 and A_5. A_5 is independent of p_1 because it gets the same return from p_2. However A_5 and A_3 dominate each other for different values of p_1 between 0 and .6. Thus A_3 is no longer dominant 60% of the time which makes A_4 which is dominant 40% of the time and is independent of A_5 the optimal alternative!

4. MAUT's Problems with Rank Reversal

To understand the commitment that people who subscribe to MAUT have against rank reversal, it is important to understand some of the problems of MAUT. The problems of MAUT are more ingrained in its parent utility theory. M. McCord and R. de Neufville of MIT, the latter a well known practitioner of MAUT, write [21] :

> "The experimental data leads to two immediate conclusions: (1) Large descriptive differences exist between the functions resulting from the various methods of assessment.*   (2) Individuals preferred to deviate in a fairly consistent fashion from the behavior presumed by the axioms. .... Our assessments were carried out by skilled, experienced practitioners of decision analysis. .... the conclusion is that the justification of the practical use of expected utility decision analysis as it is known today is weak. Specifically, the whole computational side of the method, which seeks to prescribe a normatively best choice by means of a calculus based on a description of a person's utility, does not appear valid. Either we do not know how to encode this vNM [von Neumann Morgenstern] utility or it does not exist."

> * In the spirit of (1) above, I would like to draw the attention of the reader to the proposal that AHP practitioners use MAUT peoples' various ranking methods for the alternatives and what the outcomes might be.

M. Bunge [4] writes,

> "in practice DT (utility theory, which he calls Decision Theory) is too unrealistic to be of any help in policy design and decision making".
> "In sum, DT has not been validated as a normative theory on the basis of extensive analyses of real life decision trees. Moreover, some of its assumptions are seldom satisfied in real life".

> "DT has been refuted experimentally: most people just do not behave the way predicted by the theory". He uses the works of Kahneman and Tversky to make this point.

> "These results contradict DT. Worse, they show that, though people have preferences, they have no utility functions".

> "Defenders have not bothered to put DT to the empirical test, to find out whether successful decision making does satisfy the assumptions of DT. Critics have done this and found DT inadequate as a prescriptive theory".

W. Edwards [8] in writing about expected utility maximization said, "In 1954 it was already clear that it too does not fit the facts." P.H. Farquhar, K.M. Freeman, and A.R. Pratkanis [9] in their experiments on phantom alternatives observe that their findings, "violate the basic axioms of classical choice theory". K. Lancaster [15] notes that, "A theory which can make no use of so much information is a remarkably empty one. Even the technique of supposing the existence of a utility function for all possible goods, including those not yet invented, and regarding the prices of nonexistent goods as infinite--an incredible stretching of the consumers' powers of imagination--has no predictive value."

The economist M. A. Machina [20] says, "Today choice under uncertainty is a field in flux: the standard theory is being challenged on several grounds from both within and outside economics." P. Schoemaker [37] also wrote on the limitations of expected utility theory. A. Tversky [39] has written, "A critique of expected utility theory: descriptive and normative considerations". M. Zeleny [43] writes after he gives an example of rank reversal, "A nonoptimal alternative A has been made optimal by adding nonoptimal E to the feasible set. This result is in violation of a fundamental axiom which reflects the course of development of traditional decision analysis: If an alternative A is nonoptimal, it cannot be made optimal by adding a new alternative to the problem." He also shows "the fallacy of the axiom with an example in a probabilistic setting of theoretical utility theory."

5. Support for AHP

Dyer [7], a critic of the AHP, writes about relative measurement in the AHP as follows : "The AHP is flawed as a procedure for ranking alternatives in that the rankings produced by this procedure are arbitrary." What he means by arbitrary is not clear. Dyer does not give a concrete example to illustrate exactly what he means. Instead he plays with numbers. There have been many successful applications of the AHP. Some time ago, at a NATO

sponsored meeting on multicriteria decisions in Porto, Portugal, Lawrence D. Phillips of the London School of Economics assigned the attendees to four groups, including a small AHP group and a MAUT group. They were given the task of ranking 19 toasters used in Britain. They had less than two hours to rank these toasters. It was noted that the group that used the AHP was the only one to complete the task, with four out of its five top toasters ranked the same as in a British consumer magazine called 'Which', revealed after the exercise. The AHP does not promote arbitrary rankings.

From my days at the Arms Control and Disarmament Agency when I was a practicing utility theorist [24] and supported with millions of dollars research and applications of game and utility theory, I have found negotiators, political types and also decision makers reluctant to abdicate their internal convictions to a process based on lotteries that is exacting and technical to learn. The AHP is more accessible, and the use of hierarchic structures makes it possible to include in the decision what people have on their minds. Decision makers are not constrained by many mathematical assumptions. We have successfully predicted the outcome of every U.S. presidential election since the beginnings of the AHP in the 1970's; see Prediction, Projection and Forecasting [25]. We and others have dealt very successfully with high return stocks, leading to investments for nearly twenty years; see Logic of Priorities [26]. We have done high level planning with corporations using the AHP, and the results are invariably well received; see Analytical Planning [27]. I have also participated in the process of negotiating arms reduction and conflict resolution, the latter studied with benefits and costs; see Conflict Resolution [28].

To the best of our knowledge, there is no other established theory that deals with hierarchic decisions as does the AHP. The AHP emphasizes consistency, but can tolerate a modicum of inconsistency. If inconsistency is high, the decision is re-examined or rejected as unreliable. Furthermore, the AHP is based on ratio scales, essential for dealing with hierarchic structures.

In weighting from level to level of the hierarchy, one can weight ratio scales by ratio scales and get ratio scales. However if one multiplies interval scales by interval scales one does not get a properly defined scale. Some MAUT practitioners have suggested using paired comparisons and hierarchic composition to establish priorities for the criteria and subcriteria. But at the bottom level of the hierarchy these people want to convert to using interval scales because they prevent rank reversal as they rate alternatives on a normative utility scale which forces rank preservation. But why should one make this sudden change in scale?

There are many reasons why this shift from ratio to interval scales is undesirable and leads to difficulties that would show up in works involving hierarchic structures reflecting dependence in the real world between criteria and other criteria and between criteria, subcriteria, and alternatives. I am reluctant to encourage people to use interval scales 'in making decisions for many practical and technical reasons already known to AHP practitioners. I believe the future will bear me out in this observation.

The idea for interval scales is due to the fact that MAUT generalized the interval scale approach of the much older single attribute utility theory. In MAUT one avoids the problem of composition with interval scales by using only a two level hierarchy consisting of a goal and attributes. The alternatives are then ranked according to the attributes. The weights of the attributes are considered as scaling constants and not a measure of the importance of the criteria (Keeney and Raiffa [14], page 272). This is the reason why with such scales, MAUT does not use hierarchies in which those scales are combined as does the AHP.

6. How AHP and MAUT Differ

A basic difference between AHP and utility theory is that in utility theory, to be able to compare two alternatives, one must know the value of the alternatives according to the criterion and thus the criterion needs to be defined numerically in a given

range. This assumes that a criterion is already measurable or a scale is somehow defined for it. In other words if the criterion is being a good politician, one cannot decide which of two people is the better politician unless numbers or qualitative descriptions are assigned to being a good politician. Similarly, to decide on the redder of two apples, redness needs to be measured or classified in ranges. The reason for the ranges is that MAUT needs to score the alternatives on a criterion to be able to assign utilities to the score. In the AHP this operation is summarized through the paired comparisons. There is no need to define ranges for the criteria because the alternatives are compared in relative terms with respect to each criterion and the priority scale is derived for these comparisons rather than assumed. In MAUT a basic and the only additional assumption to the axioms of utility theory, is that the criteria must be mutually utility independent. This leads to the result that utility functions always exist mathematically. Yet the utility theorist Fred Roberts wrote a chapter [22] whose title is "What if Utility Functions Do Not Exist?", and attributed the possible nonexistence of utility functions to sufficiently inconsistent judgments by the decision maker. Because of its underlying mathematics, the AHP is well equipped to deal with the question of the consistency of judgments and its effect on the reliability of a decision. A useful part of the theory is how to measure and use inconsistency as a guide in decision making.

In the AHP the criteria may or may not be independent. Their independence is a consequence of the framework of the problem and does not determine the existence of a solution. In general, in a hierarchy, they are assumed to be independent. But there is a way to determine dependence which involves answering questions about the strength of the relation of a criterion to other criteria with respect to higher level goals, by individuals experienced with the interactions between criteria in the given problem as in the dependence of industries on each other in input-output analysis [36]. When there are several levels, the abstractions give MAUT

practitioners difficulty in coping with the issue of criteria independence and thus also with the question of the existence of a utility function. Dyer [7] who said that the AHP produces arbitrary rankings, surprisingly suggested the remedy for MAUT that "the weights on the criteria can be obtained in the traditional AHP manner and tested to ensure that the weights of the criteria and the scores of alternatives on the criteria are normalized with respect to the same range of alternative values". This is a strange recipe, which if one understands it correctly, would use the AHP and ratio scales to derive priorities in a hierarchy of more than two levels, but would then use MAUT and interval scales to rate the alternatives at the bottom of that hierarchy! But this would contradict Keeney and Raiffa's thinking [14]. Utility theory, on which MAUT was built, has little to say about several criteria and how to measure and combine them. Thus until the arrival of MAUT, there was inadequate mathematical attention given to how to rate the criteria on ratio scales. It is indeed ironic that some MAUT practitioners have proposed the use of the AHP to rate the criteria in their business. The idea does not tally well with what Keeney and Raiffa wrote on page 272 of their book, "If we have assessed k_y = .75 and k_z = .25, we cannot say that Y is three times as important as Z. In fact, we cannot conclude that attribute Y is more important than Z. Going one step further, it is not clear how we would precisely define the concept that one attribute is more important than another." MAUT users end up with painfully constructed interval scales, but with no such elegantly constructed ways to weight the (much more important) criteria on a ratio scale.

7. Some Rules for Rank Preservation and Reversal

We consider two types of decisions, those with separate benefit and cost rankings leading to a third ranking and those with a single ranking. When benefit/cost analysis is used, rank reversal will inevitably occur whether absolute or relative measurement are used. We have found it to be a realistic and satisfactory approach to a large number of decisions. .

Let us now consider the case of a single ranking and a sequential ranking of alternatives. There are the following factors to consider: criteria, criteria weights, whether criteria can always be assumed to be independent from the alternatives and whether the alternatives (which always depend on the criteria and on their weights) depend on each other in their ranking; in other words when to use absolute measurement (the rank preserving mode) and when to use relative measurement (the rank reversal or adjustment mode). Note that to say that criteria are independent of the alternatives may not be always true. The importance of a criterion can certainly change on examining a particular set of alternatives and the question is whether old knowledge of the importance of the criteria should not be modified by examining the new set of alternatives of a decision. One needs a method to deal with decision situations in which the criteria depend on the alternatives and also on the number of alternatives.

Alternatives are comparable objects sharing the same properties in varying degrees. An added alternative is either a copy or a perturbation of the others and hence it can be thought·of as similar to the others. The alternatives of one problem can become criteria for another as in policies and consequences of policies. Ratio scales are the only kind of numbers that can be used to multiply and aggregate weights in a hierarchy. One cannot use ordinals or interval scales for that purpose.

There are counterexamples in the literature for nearly every situation where one might want to impose a general condition to preserve rank. Statements made about how the rank of an alternative should or should not affect the ranks of other alternatives, point to the need for a way to compare the alternatives and to relative measurement. With that in mind, let us see when one might justifiably say that rank can and should be preserved and when one might not.

1 - Rank reversal is legitimate if new criteria are introduced by adding an alternative.

2 - Rank reversal is legitimate if the weights of the old

criteria are changed by adding an alternative.

3 - When relative measurement is used to compare all the alternatives and the judgments are consistent, rank reversal is legitimate and it can occur according to several criteria but not just according to a single criterion.

4 - Rank should be preserved by using normative or descriptive absolute measurement, with the condition imposed that the criteria be independent of the alternatives and the weights of the criteria set in advance never to be affected by examining new alternatives. This is applicable to an open system of alternatives in which new alternatives can be added at any time (sequential decision making).

5 - In problems of scarcity, the number of alternatives in an open system changes the weights of the criteria and can lead to rank reversal. Here one cannot assume that the criteria are independent of the alternatives. Relative measurement can be used to account for the effect of the alternatives on the weights of the criteria.

Appendix : Three Paradoxes of Utility Theory

In single attribute utility theory, if a new alternative is introduced, a completely new problem is defined and there need not be a relationship between the two rankings, new and old. Introducing new information requires such an assumption because only one criterion, utility, is allowed in that theory. Thus rank reversal is acceptable, but the problem of rating alternatives is not solved in a satisfactory way by using just a single criterion as suggested by some of the paradoxes.

There are three kinds of paradoxes in utility theory. A major concern is that these paradoxes occur frequently and are of serious concern for the healthfulness of the theory. The first (typified by Allais' paradox) are examples contradicting axiom 4 of transitivity (and expected utility which is based on lotteries.) The second are examples contradicting axiom 6 of nonoptimal acts. The third are examples contradicting axiom 7 of independence from

irrelevant alternatives. Here is an example of each type.

1. Luce and Raiffa [18] (pp. 22,4) point out in their book that preferences must precede the numerical characterization of them. Then they say :

> If we add more gambles to the collection and try to assign utilities as we have done, it is clear that to be successful the subject's preferences will have to satisfy some consistency requirements. For example, if he prefers A to B, B to C, and a lottery which yields A with probability 2/3 and C with probability 1/3 to a lottery which yields B with probability 3/4 and C with probability 1/4, then we are in trouble. Or if he prefers A to B, B to C, and B to any lottery involving A and C as prizes so long as it is a bona fide gamble, i.e., $p \neq 1$, we are again in trouble.

To correct the situation they propose imposing conditions (through axiomatization) on the way people should be thinking in rating alternatives and on the nature of the alternatives themselves.
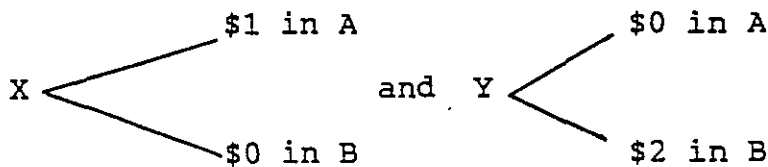
Two Attribute Rank Reversal

The following example was developed by the author's colleague Anil Makhija.

Suppose there are two states of the world, A and B, with probabilities, p(A) = .5 and p(B) = .5, respectively. Dollars in 'A' and 'B' provide the following utility:

U ($x in A) = x^2

and

U ($x in B) = x.

Then, we can compare preferences between two lotteries,

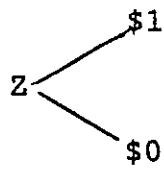$1 in A

X

$0 in B

and   Y

$0 in A

$2 in B

$$E[u(X)] = .5(1^2) + .5(0) \quad ; \quad E[u(Y)] = .5(0) + .5(2)$$

$$= 0.50 \qquad\qquad\qquad = 1$$

i.e. $Y > X$

Now suppose, we already own a lottery, Z

then

$$E[u(XZ)] = .5(2^2) + .5(0) \quad ; \quad E[u(YZ)] = .5(1^2) + .5(2)$$
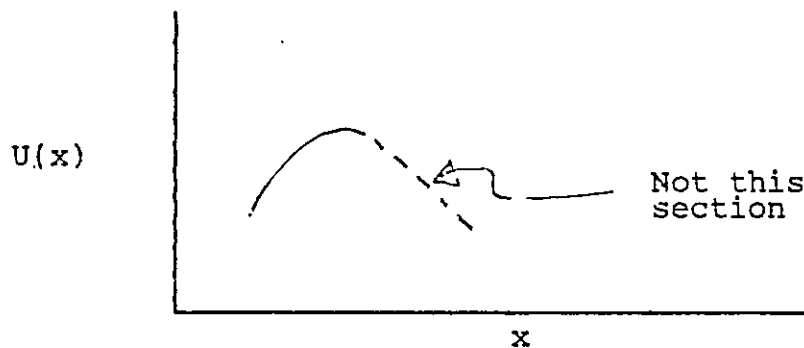
$$= 2 \qquad\qquad\qquad = 1.5$$

i.e. $XZ > YZ$

With the ownership of Z, $X > Y$

Note that our utlity function does not satisfy the usual requirements $\delta u/\delta x > 0$ and $\delta^2 u/\delta x^2 < 0$. So, we change to a utility function which does, say a quadratic function.

But we must ensure that we use the appropriate section of that quadratic function.

$$U(x) = a + bx + cx^2$$

For example, consider :

$$U (x \text{ in } A) = -100 + 160x - 16x^2, \text{ and}$$

$$U (x \text{ in } B) = -25 + 50x - 4x^2 .$$

Redoing our previous example we have:

$$E[u(X)] = (0.5)(-100+160-16) + (.5)(0) = 22$$

$$E[u(Y)] = (.5)(0) + (0.5)(-25+100-16) = 29.5$$

Compared by themselves : Y > X

Assuming that we already own Z :

$$E[u(XZ)] = (0.5)(-100+320-64) + (.5)(0) = 78$$

$$E[u(YZ)] = (0.5)(-100+160-16) + (0.5)(-25+100-16) = 51.5$$

We are better off in both cases, but now XZ > YZ. Thus if we own Z, then X > Y .

Checking our utility functions $\delta U(x$ in $A)/\delta x = 160-32x > 0$ when $x < 5$ and $\delta^2 U(\bullet)/\delta x^2 < 0$.

Similarly, for $\delta U(x$ in $B)/\delta x = 50 - 8x > 0$ or $x < 6.25$ and $\delta^2 U(\bullet)/\delta x^2 < 0$.

In our cases : x in either A or B is at most 2 and the restrictions are met.

2. Violation of axiom 6 that by adding alternatives that are dominated or are equivalent to old alternative does not change the rank (optimality) of the old ones, an illustrated by an example from Luce and Raiffa [18] page 288.

"A gentleman wandering in a strange city at dinner time chances upon a modest restaurant which he enters uncertainly. The waiter informs him that there is no menu, but that this evening he may have either broiled salmon at $2.50 or steak at $4.00. In a first-rate restaurant his choice would have been steak, but considering his unknown surroundings and the different prices he elects the salmon. Soon after the waiter returns from the kitchen, apologizes profusely, blaming the uncommunicative chef for omitting to tell him that fried snails and frog's legs are also on the bill

of fare at $4.50 each. It so happens that our hero detests them both and would always select salmon in preference to either, yet his response is 'Spendid, I'll change my order to steak.' Clearly, this violates the seemingly plausible axiom 6. Yet can we really argue that he is acting unreasonably? He, like most of us, has concluded from previous experience that only "good" restaurants are likely to serve snails and frog's legs, and so the risk of a bad steak is lessened in his eyes."

3. A practical example which violates axiom 7, that the rank of a low ranking alternative cannot be improved (or diminished against old acts) by adding new alternatives (whether dominated or not) is illustrated by the lady shopping for hats mentioned earlier.

A BIBLIOGRAPHY of Questions and Answers and Criticisms of AHP

1. Barzilai, J., W.D. Cook and B. Golany, "Consistent Weights for Judgements Matrices of the Relative Importance of Alternatives", Oper. Res. Lett. 6, 3, 131-134, 1987.

2. Belton, V. and T. Gear, "On a short-coming of Saaty's method of analytic hierarchies", Omega 11, 3, 228-230, 1983.

3. Blumenthal, Arthur L., The Process of Cognition, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1977.

4. Bunge, Mario, Treatise on Basic Philosophy Volume 7, D. Reidel Publishing Company, Boston, 1985.

5. Corbin, Ruth, and A.A.J. Marley, "Random Utility Models with Equality: An apparent, but Not Actual, Generalization of Random Utility Models", Journal of Mathematical Psychology, 11, 274-293, 1974.

6. Crawford, G.B., "The geometric mean procedure for estimating the scale of a judgment matrix", <u>International Journal of Math Modeling</u>, 9,7-5, 1987.

7. Dyer, J.S., "Remarks on the Analytic Hierarchy Process", <u>Management Science</u>, 36, 3, 249-258, 1990.

8. Edwards, W., "Behavioral decision theory", <u>Annual Review of Psychology</u> (Edited by P. Farnsworth, O. McNeman and Q. McNemar), Palo Alto, California, 1961.

9. Farquhar, P.H., K.M. Freeman and A.R. Pratkanis, <u>Phantoms As Psychological Motivation: Evidence For Compliance And Reactance Processes</u>, Carnegie Mellon University, Pittsburgh, PA, 1990.

10. Forman, E.H. "Rank Adjustment/Rank Reversal - and How to Evaluate a Large Number of Alternatives with AHP", George Washington University, Washington, DC, forthcoming paper

11. Harker, P.T. and L.G. Vargas, "Reply to 'Remarks on the Analytic Hierarchy Process' by J.S. Dyer", <u>Management Science</u> 36, 3, 269-273, 1990.

12. Holder, R.D., "Some Comments of the Analytic Hierarchy Process", <u>J. Opl. Res. Soc.</u> 41, 1073-1076, 1990

13. Kahneman, D. and A. Tversky, "Prospect theory: An analysis of decision under risk". <u>Econometrica</u> 47, 263-291, 1979.

14. Keeney, R.L. and H. Raiffa, <u>Decisions with Multiple Objectives: Preference and Value Tradeoffs</u>, Wiley, New York, 1976.

15. Lancaster, Kelvin J., "A New Approach to Consumer Theory", <u>Journal of Political Economy</u>, LXXIV, 1966.

16. Lootsma, F.A., "Conflict resolution via pairwise comparison of concessions", <u>Eur. J. Opl Res.</u> 40, 109-116, 1989.

17. Lootsma, F.A., "Letter to the Editor", <u>ORiON</u>, 5, 1, 52-54, 1989.

18. Luce, R.D. and H. Raiffa, <u>Games and Decisions</u>, Wiley, New York, 1957.

19. Luce, R. D., <u>Individual choice behavior: A theoretical analysis</u>. New York, Wiley, 1959.

20. Machina, Mark A., "Choice Under Uncertainty: Problems Solved and Unsolved", <u>Economic Perspectives</u>, vol.1, no.1, 121-154, Summer 1987

21. McCord, M. and R. de Neufville, "Empirical Demonstration That Expected Utility Decision Analysis is not Operational", Chapter (pp 181-200) in foundation of Utility and Risk Theory with Applications", Stigun Wenstop (ed.), D. Reidel Publishing Company, Boston, 1983.

22. Roberts, F.S., "What if Utility Functions Do Not Exist?", <u>Theory and Decision</u>, 3, 126-139, D. Reidel Publishing Company, Dordrecht-Holland, 1972.

23. Saaty, T.L., "Rank Generation, Preservation, and Reversal in the Analytic Hierarchy Decision Process", <u>Decision Sciences</u>, Vol. 18, No. 2, Spring 1987.

24. Saaty, T.L., <u>Mathematical Models of Arms Control and Disarmament</u>, Wiley, New York, 1968.

25. Saaty, T.L. and L.G. Vargas, <u>Prediction, Projection and Forecasting</u>, Kluwer Academic Publishers, Boston, 1991.

26. Saaty, T.L. and L.G. Vargas, <u>The Logic of Priorities</u>, Kluwer-Nijhoff Publishing, Boston, 1982.

27. Saaty,T.L. and K.P. Kearns, <u>Analytical Planning</u>, Pergamon Press, New York, 1985.

28. Saaty, T.L. and J.M. Alexander, <u>Conflict Resolution</u>, Praeger, New York, 1989.

29. Saaty, T.L. and L.G. Vargas, "The Legitimacy of Rank Reversal", <u>OMEGA</u>, 12, 5, 513-516, 1984.

30. Saaty, T.L., "Response From: Thomas L. Saaty, University of Pittsburgh", <u>ORION</u>, 5, 1, 55-57, 1989

31. Saaty, T.L., "Eigenvector and logarithmic least squares", <u>Eur. Journ. Opl Resrch.</u> 48, 156-160, 1990.

32. Saaty, T.L., "Rank According to Perron: A New Insight", <u>Mathematics Magazine</u>, 60, 4, 211, 1987.

33. Saaty, T.L. and L.G. Vargas, "Inconsistency and Rank Preservation", <u>Journal of Mathematical Psych.</u> 28, 2, 205, 1984

34. Saaty, T.L., "An Exposition of the AHP in reply to the paper 'Remarks on the Analytic Hierarchy Process'", <u>Management Science</u> 36, 3, 259-268, 1990.

35. Saaty, T.L., "Reply to Holder's Comments On The Analytic Hierarchy Process", forthcoming, 1991

36. Saaty, T.L. and L.G. Vargas, "A Note of Estimating Technological Coefficients by Hierarchical Measurement:, Socio-Econ. Plan. Sci. 13, 6, 333-336, 1979.

37. Schoemaker, P.J.H., "The Expected Utility Model : Its Variants, Purposes, Evidence, and Limitations." J. Econ. Lit. 20, 529-563, 1982.

38. Schoner, B. and W. Wedley, "Ambiguous criteria weights in AHP: consequences and solutions", Decis. Scie. 20, 462-475, 1989.

39. Tversky, A., "A critique of expected utility theory: descriptive and normative considerations". Erkenntnis 9, 163-173, 1975.

40. Tversky, A., "Intransitivity of Preferences", Psychol. Rev., 76, 31-48, 1969.

41. Viet, C.T. "Ratio and Subtractive Processes in Psychophysical Judgment" Journal of Experimental Psychology:General 107, 1, 81-107, 1978.

42. Watson, S.R. and A. Freeling, "Assessing Attribute Weights", Omega 10, 6, 582-583, 1982.

43. Zeleny, M., Multiple Criteria Decision Making, McGraw Hill, New York, 1982.