

SOME CRITICAL ISSUES IN MAKING DECISIONS WITH PAIRWISE COMPARISONS

EVANGELOS TRIANTAPHYLLOU¹ and STUART H. MANN²

¹ Assistant Professor, Department of Industrial and Manufacturing Systems Engineering, Louisiana State University, 3134C CEBA Building, Baton Rouge, LA 70803-6409, U.S.A.

E-mail: IETRIAN@LSUVM.SNCC.LSU.EDU

² Professor of Operations Research and Director of the School of Hotel, Restaurant and Recreation Management, The Pennsylvania State University, University Park, PA 16802, U.S.A.

Abstract: Pairwise comparisons are the core of the AHP and provide an intuitive and efficient method in eliciting information for multi-criteria decision-making (MCDM) applications. However, there are a number of critical issues related to their application in solving a MCDM problem. The main challenges are: (i) how to quantify them, (ii) how to process the resulted reciprocal matrices, and (iii) how to process the decision matrices. This paper presents some of the alternative approaches proposed to solve the previous problems and discuss their relative effectiveness and limitations.

Key words: Pairwise comparisons, AHP, scales, eigenvalue approach, multi-criteria decision-making.

1. Introduction.

The Analytic Hierarchy Process (AHP) is best illustrated by Saaty in [26] to [31]. The AHP has attracted the interest of many researchers mainly due to the nice mathematical properties of the method and the fact that the required input data are rather easy to obtain. The AHP is a decision support tool which can be used in solving complex decision problems. It uses a multi-level hierarchical structure of objectives, criteria, subcriteria, and alternatives. The pertinent data are derived by using a set of pairwise comparisons. These comparisons are used to obtain the weights of importance of the decision criteria, and the relative performance measures of the alternatives in terms of each individual decision criterion. If the comparisons are not perfectly consistent, then it provides a mechanism for improving consistency [44].

However, since the early days it became apparent that there are some problems with the way pairwise comparisons are used and the way the AHP evaluates alternatives. First, Belton and Gear in [2] observed that the AHP may reverse the ranking of the alternatives when an alternative identical to one of the already existing alternatives is introduced. In order to overcome this deficiency, Belton and Gear proposed that each column of the AHP decision matrix to be divided by the maximum entry of that column. Thus, they introduced a variant of the original AHP, called the revised-AHP. Besides the revised-AHP, other authors also introduced other variants of the original AHP (see, for instance, [16]). However, the AHP is the most widely accepted method and is considered by many as the most reliable MCDM method.

The fact that rank reversal also occurs in the AHP when near copies are considered, has also been studied by Dyer and Ravinder [4] and Dyer and Wendell [5]. Saaty [29, 30] provided some axioms and guidelines on how close a near copy can be to an original alternative without causing a rank reversal. He suggested that the decision maker has to eliminate alternatives from consideration that score within 10 percent of another alternative. This recommendation was later sharply criticized by Dyer [6].

Dyer [6, 7] demonstrated that the original AHP may result in arbitrary rankings, even when no identical alternatives are considered. He also says that in some cases the AHP may yield results which may be highly correlated with the true performances of a consistent decision maker, and even provide an accurate ranking of alternatives in other special cases. However, the same author asserts that, in general, this cannot be guaranteed a priori.

Dyer attributed this problem to the principle of hierarchic composition, which is one of the main assumptions in the AHP. This principle assumes that the weights of the criteria do not depend on the alternatives under consideration. As an alternative approach, Dyer suggested that AHP be combined with utility theory. That paper by Dyer stimulated some responses by Saaty [31] and Harker and Vargas [11]. They expressed many concerns whether traditional utility theory can be combined with the AHP.

As Winkler [42] remarked in a related technical note, utility theory is not a panacea. He cautions the reader to be careful in accepting alternative theories. Clearly, these problems are still controversial in decision theory. Some additional discussions on these truly important issues are due to Bell and Farquhar [1], Hogarth and Reder [35], Weber and Camerer [41], Munier [22], Fishburn and LaValle [10], and Sarin [25].

Many of the previous problems are not bound only in the AHP. They are present with any method which has to elicit information from pairwise comparisons. These problems can be divided into the following three categories:

- (i) *How to quantify the pairwise comparisons,*
- (ii) *How to process the resulted reciprocal matrices,*
- and (iii) *How to process the decision matrices.*

Next, we consider some of the main ideas related with pairwise comparisons (or PCs). In the sections that follow, we consider each one of the previous challenges, and discuss some remedies which have been proposed to cope with the related difficulties.

2. The Concept of the Pairwise Comparisons

One of the most crucial steps in many decision-making methods is the accurate estimation of the pertinent data. This is a problem not bound in the AHP method only, but it is crucial in many other methods which need to elicit qualitative information from the decision-maker. Very often qualitative data cannot be known in terms of absolute values. For instance, what is the worth of the i -th alternative in terms of a political impact criterion? Although information about questions like the previous one is vital in making the correct decision, it is very difficult, if not impossible, to quantify it correctly. Therefore, many decision-making methods attempt to determine the relative importance, or weight, of the alternatives in terms of each criterion involved in a given decision-making problem.

An approach based on pairwise comparisons which was proposed by Saaty (see, for instance, [26] and [28]) has long attracted the interest of many researchers. Pairwise comparisons are used to determine the relative importance of each alternative in terms of each criterion. In this approach a decision-maker has to express his opinion about the value of one single pairwise comparison at a time. Usually, the decision-maker has to choose his answer among 10-17 discrete choices. Each choice is a linguistic phrase. Some examples of such linguistic phrases are: "*A is more important than B*", or "*A is of the same importance as B*", or "*A is a little more important than B*", and so on. Our concern in this paper is not the wording of these linguistic statements, but, instead, the numerical values which should be associated with such statements.

The main problem with the pairwise comparisons is how to quantify the linguistic choices selected by the decision maker during their evaluation. All the methods which use the pairwise comparisons approach eventually express the qualitative answers of a decision maker into some numbers which, most of the time, are ratios of integers. A case in which pairwise comparisons are expressed as differences (instead of ratios) was used to define similarity relations and is described by Triantaphyllou in [40]. The next section examines the issue of quantifying pairwise comparisons. Since pairwise comparisons are the keystone of these decision-making processes, correctly quantifying them is the most crucial step in multi-criteria decision-making methods which use qualitative data.

3. Problem #1: On the Quantification of Pairwise Comparisons

Pairwise comparisons are quantified by using a scale. Such a scale is an one-to-one mapping between the set of discrete linguistic choices available to the decision maker and a discrete set of numbers which represent the

importance, or weight, of the previous linguistic choices. There are two major approaches in developing such scales. The first approach is based on the linear scale proposed by Saaty (1980) as part of the AHP. The second approach was proposed by Lootsma in [13], [14], and [15] and determines exponential scales. Both approaches depart from some psychological theories and develop the numbers to be used based on these psychological theories.

3.1. Scales Defined on the Interval [9, 1/9]

In 1846 Weber stated his law regarding a stimulus of measurable magnitude. According to his law a change in sensation is noticed if the stimulus is increased by a constant percentage of the stimulus itself (Saaty, [28]). That is, people are unable to make choices from an infinite set. For example, people cannot distinguish between two very close values of importance, say 3.00 and 3.02. Psychological experiments have also shown that individuals cannot simultaneously compare more than seven objects (plus or minus two) (Miller, [21]). This is the main reasoning used by Saaty to establish 9 as the upper limit of his scale, 1 as the lower limit and a unit difference between successive scale values.

The values of the pairwise comparisons are determined according to the scale introduced by Saaty in [28]. According to this scale (which we call Scale1), the available values for the pairwise comparisons are members of the set: {9, 8, 7, 6, 5, 4, 3, 2, 1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9}. The above numbers illustrate that the values for the pairwise comparisons can be grouped into the two intervals [9, 1] and [1, 1/9]. As it was stated above, the values in the interval [9, 1] are evenly distributed, while the values in the interval [1, 1/9] are skewed to the right end of this interval.

There is no good reason why for a scale defined on the interval [9, 1/9] the values on the sub-interval [9, 1] should be evenly distributed. An alternative scale could have the values evenly distributed in the interval [1, 1/9], while the values in the interval [9, 1] could be simply the reciprocals of the values in the interval [1, 1/9]. This consideration leads to the scale (which we call Scale2) with the following values: {9, 9/2, 9/3, 9/4, 9/5, 9/6, 9/7, 9/8, 1, 8/9, 7/9, 6/9, 5/9, 4/9, 3/9, 2/9, 1/9}. This scale was originally presented by Ma and Zheng in [17]. In the second scale each successive value on the interval [1, 1/9] is $(1 - 1/9) / 8 = 1/9$ units apart. In this way, the values in the interval [1, 1/9] are evenly distributed, while the values in [9, 1] are simply the reciprocals of the values in [1, 1/9]. It should be stated here that the notion of having a scale with a group of values evenly distributed, is followed in order to be in agreement with the same characteristic of the original Saaty scale. As it will be seen in the next section, other scales can be defined without having evenly distributed values.

Besides the second scale, many other scales can be generated. One way to generate new scales is to consider weighted versions between the previous two scales. That is, for the interval [1, 1/9] the values can be calculated using the formula:

$$\text{NewValue} = \text{Value}(\text{Scale1}) + (\text{Value}(\text{Scale2}) - \text{Value}(\text{Scale1})) * (\alpha / 100).$$

where α can range from 0 to 100. Then, the values in the interval [9, 1] are the reciprocals of the above values. For $\alpha = 0$ Scale1 is derived, while for $\alpha = 100$ Scale2 is derived.

3.2. Exponential Scales

A class of exponential scales has been introduced by Lootsma in [13] to [15]. The development of these scales is based on an observation in psychology about stimulus perception (denoted as e_i). According to that observation, due to Roberts [23], the difference $e_{n+1} - e_n$ must be greater than or equal to the smallest perceptible difference, which is proportional to e_n . As a result of Robert's observation the numerical equivalents of these linguistics choices need to satisfy the following relations:

$$\begin{aligned} e_{n+1} - e_n &= \epsilon e_n, \quad (\text{where } \epsilon > 0) \quad \text{or:} \\ e_{n+1} &= (1 + \epsilon) e_n = (1 + \epsilon)^2 e_{n-1} = \dots \\ \dots &= (1 + \epsilon)^{n+1} e_0, \quad (\text{where: } e_0 = 1) \quad \text{or: } e_n = e^{\gamma * n}. \end{aligned}$$

In the previous expressions the parameter γ is unknown (or, equivalently, ϵ is unknown), since $\gamma = \ln(1 + \epsilon)$, and e is the basis of the natural logarithms (please note that e_i is just the notation of a variable).

Another difference between exponential scales and the Saaty scale is the number of categories allowed by the exponential scales. There are only four major linguistically distinct categories, plus three so-called threshold categories between them. The threshold categories can be used if the decision maker hesitates between the main categories. For a more detailed documentation on psychophysics we refer the reader to Marks [18], Michon et al.

[19], Roberts [23], Zwicker [45], and Stevens and Hallowell Davis [24]. The reader will find that the sensory systems for the perception of tastes, smells, and touches follow the power law with exponents near 1.

3.3. Evaluating Different Scales

In order for different scales to be evaluated, two evaluative criteria were developed by Triantaphyllou et al. in [38]. Furthermore, a special class of pairwise matrices was also developed. These special matrices were then used in conjunction with the two evaluative criteria in order to investigate some stability properties of different scales.

The first matrix is called the Real Continuous Pairwise matrix, or the RCP matrix. This matrix is configured by the actual pairwise comparisons of a given test problem. Next, a scale is used and the hypothesis that the decision maker is as accurate as possible is made. As result, a RCP matrix can be approximated by, what in [38] is called, the Closest Discrete Pairwise matrix or the CDP matrix. The CDP matrix may not be perfectly consistent. That is, the consistency index (CI) values of CDP matrices are not necessarily equal to zero (see also figure 1). More on this inconsistency, and some other important theoretical issues, can be found in Triantaphyllou et al. [36] and [38]. It is important to observe here that the CDP matrices are the reciprocal matrices with pairwise comparisons that a decision maker will construct if we assume that each of his pairwise comparisons is the closest possible to its actual real value.

Recall that the decision maker is limited by the discrete values (i.e., the values from the set Θ provided to him by a scale). He can never know the actual values of his pairwise comparisons. He simply attempts to approximate them. In other words, we assume here that these approximations are the closest possible. Clearly this is a highly favorable assumption when one attempts to investigate the effectiveness of various scales. In [38] the previous two classes of matrices with pairwise comparisons were used in conjunction with two evaluative criteria to study the effectiveness of the previous scales.

Two kinds of ranking inconsistency were examined. The first kind is "ranking reversal". The second kind is "ranking indiscrimination". In order to examine the effectiveness of various scales the concept of the CDP matrices can be used. That is, the ranking implied by a CDP matrix (which, as mentioned in the previous section, represents the best decisions that a decision maker can make) has to be identical with the actual ranking indicated by the corresponding original RCP matrix. Therefore, the following two evaluative criteria were introduced to investigate the effectiveness of any scale which attempts to quantify pairwise comparisons:

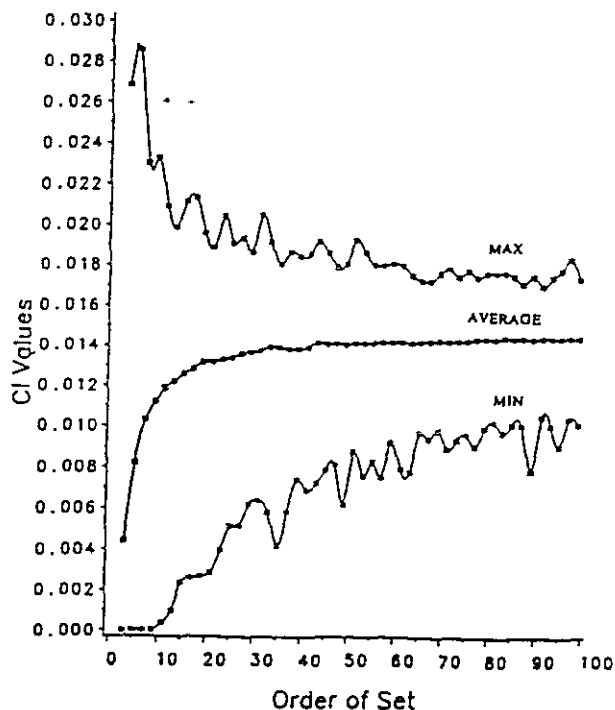


Figure 1. (from Triantaphyllou et al. [36]).
Maximum, Average, and Minimum CI Values of Random CDP Matrices When the Original Saaty Scale is Used.

CRITERION 1:

Let A be a random RCP matrix with the actual values of the pairwise comparisons of N alternatives. Let B be the corresponding CDP matrix when some scale is applied. Then, the ranking yielded when the CDP matrix is used should not demonstrate any ranking inversions when the CDP ranking is compared with the ranking derived from the RCP matrix.

CRITERION 2:

Let A be a random RCP matrix with the actual values of the pairwise comparisons of N alternatives. Let B be the corresponding CDP matrix when some scale is applied. Then, the ranking yielded when the CDP matrix is used should not demonstrate any ranking indiscriminations when the CDP ranking is compared with ranking derived from the RCP matrix.

Since the previous two ranking anomalies are independent of the scale under consideration or the method used to process matrices with pairwise comparisons, the previous two evaluative criteria can be used to evaluate any scale and method.

In Triantaphyllou et al. [38] the previous two evaluative criteria were applied as follows: First, a random RCP matrix was determined. Given that RCP matrix and a scale, the corresponding CDP matrix was found. Next, any ranking inversions or indiscriminations were recorded. This was done for matrices of different sizes. The resultant inversion and indiscrimination rates lead to the formulation of a decision problem as follows. Since there were a total of 78 different scales for which relative performance data in terms of the two evaluative criteria were obtained, it was concluded that this was a classical multi-criteria decision-making problem. That is, the 78 scales can be treated as the alternatives in this decision-making problem. The only difficulty in this consideration is how to assess the weights for the two evaluative criteria. Which criterion is the most important one? Which is the less important? Apparently these type of questions cannot be answered in a universal manner.

The weights for these criteria depend on the specific application under consideration. However, one may argue that, in general, ranking indiscrimination is less severe than ranking reversal. Depending on how more critical ranking reversals are, one may want to assign a higher weight to the ranking reversal criterion. If both ranking reversal and ranking indiscrimination are equally severe then the weights of the two criteria are equal (i.e., they are set equal to 0.50).

For the above reasons, the previous decision-making problem was solved for all possible weights of the two criteria. Criterion 1 was assigned weight W_1 , while criterion 2 was assigned weight $W_2 = 1.00 - W_1$ (where $1.00 \geq W_1 \geq 0.00$).

For each of these combinations of the weights of the two evaluative criteria, the decision-making problem was solved by using the revised Analytic Hierarchy Process (introduced by Belton and Gear [2]). In Triantaphyllou and Mann [34] the revised Analytic Hierarchy Process was found to perform better when it was compared with other multi-criteria decision-making methods. For each of the above decision-making problems the best and the worst alternative (i.e. scale) was recorded.

The results regarding the best scales are depicted in figure 2. Similarly, the results regarding the worst scales are depicted in figure 3. In both cases the best or worst scales are given for different values of the weight for the first criterion (or equivalently the second criterion) and the size of the set. The computational results demonstrate that only very few scales can be classified either as the best or the worst scales. It is possible the same scale (for instance, scale 78) to be classified as one of the best scales for some values of the weight W_1 , and also as the worst scale for other values of the weight W_1 . Probably, the most important observation is that the results illustrate very clearly that there is no single scale which is the best scale for all cases. Similarly, the results illustrate that there is no single scale which is the worst scale for all cases. However, according to these computational results, the best (or worst) scale can be determined only if the number N is known and the relative importance of the weights of the two evaluative criteria has been assessed.

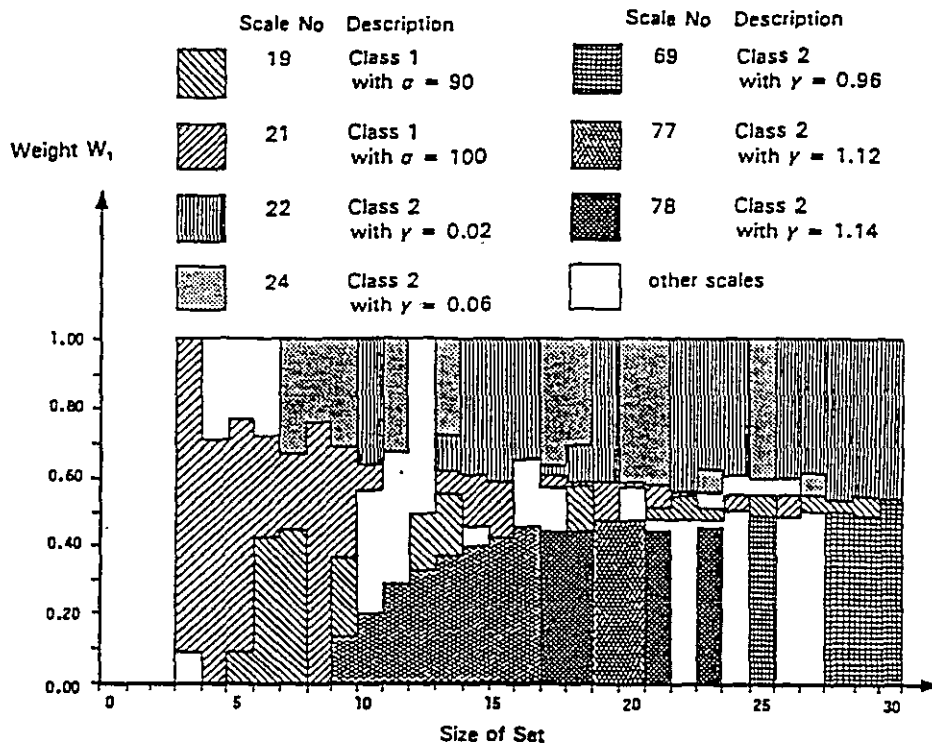


Figure 2.
The Best Scales (according to Triantaphyllou et al., [38]).

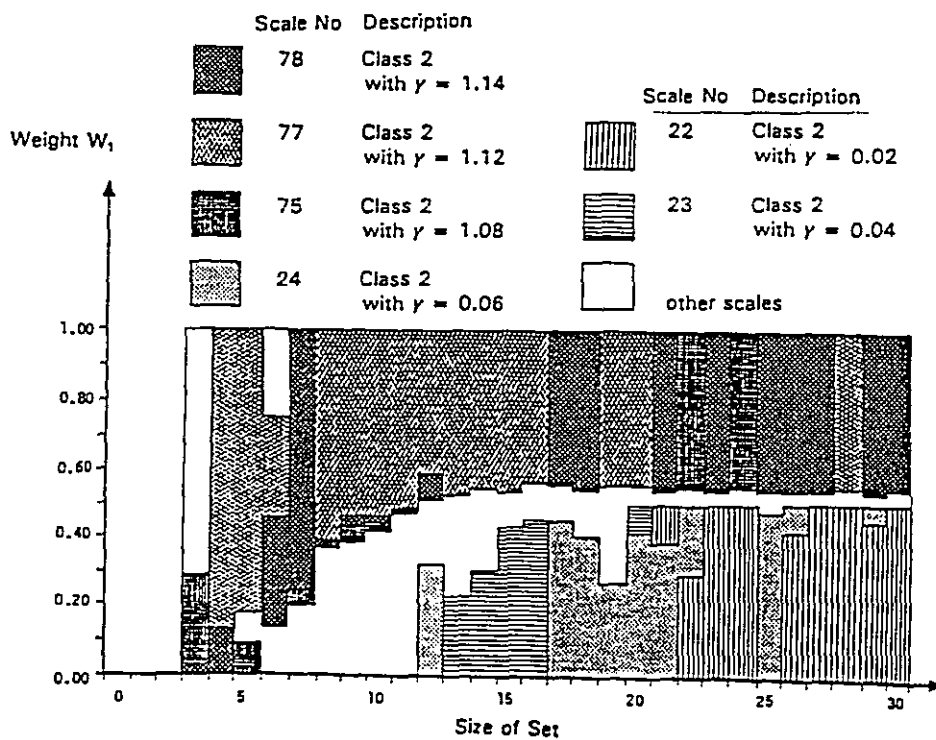


Figure 3.
The Worst Scales (according to Triantaphyllou et al., [38]).

4. Problem #2: Processing Reciprocal Matrices with PCs

At this point it assumed that the decision-maker has determined the values of all the pairwise comparisons. That is, available are the values a_{ij} (for $i, j = 1, 2, 3, \dots, N$), where a_{ij} represents the relative performance of alternative A_i when it is compared with alternative A_j in terms of a single criterion. Given these values, the decision maker needs to determine the relative weights, say W_i ($i=1, 2, 3, \dots, N$), of the alternatives in terms of the single criterion. Saaty [28] has proposed a method which asserts that the desired weights are the elements of the right principal eigenvector of the matrix with the pairwise comparisons. This method has been evaluated under a continuity assumption by Triantaphyllou and Mann in [37]. Moreover, other authors have proposed alternative approaches.

For instance, Chu, et al. in [3] observed that, given the data a_{ij} , the values W_i to be estimated are desired to have the following property:

$$a_{ij} \approx W_i/W_j.$$

This is reasonable, since a_{ij} is meant to be the estimate of the ratio W_i/W_j . Then, in order to get the estimates for the W_i given the data a_{ij} , they proposed the following constrained optimization problem:

$$\begin{aligned} \text{minimize } S &= \sum_{i=1}^N \sum_{j=1}^N (a_{ij}W_j - W_i)^2 \\ \text{subject to: } &\sum_{i=1}^N W_i = 1, \\ \text{and } W_i > 0, &\text{ for any } i = 1, 2, 3, \dots, N. \end{aligned}$$

They also gave an alternative expression S_1 that is more difficult to solve numerically. Where:

$$\text{minimize } S_1 = \sum_{i=1}^N \sum_{j=1}^N (a_{ij} - W_j/W_i)^2.$$

In Federov et al. [8], a variation of the previous least-squares formulation was proposed. For the case of only one decision maker the authors recommended to use the following models:

$$\begin{aligned} \log a_{ij} &= \log W_i - \log W_j + \Psi_1(W_i, W_j) \varepsilon_{ij}, \\ \text{and } a_{ij} &= W_i/W_j + \Psi_2(W_i, W_j) \varepsilon_{ij}, \end{aligned}$$

where W_i and W_j are the true (and thus unknown) weights; $\Psi_1(X, Z)$ and $\Psi_2(X, Z)$ are given positive functions (where $X, Z > 0$). The random errors ε_{ij} are assumed to be independent with zero mean and unit variance. However, they fail to give a way of selecting the appropriate two previous positive functions.

In the following paragraphs we present the main idea which was originally described in Triantaphyllou et al., [34]. In that treatment the assumption of the human rationality is made. According to that assumption the decision maker is a rational person. Rational persons are defined here as individuals who try to minimize their regret [32], to minimize losses, or to maximize profit [43]. In the present context, minimization of regret of losses, or maximization of profit could be interpreted as the effort of the decision maker to minimize the errors involved in the pairwise comparisons

As it was stated in the previous paragraphs, in the inconsistent case, the entry a_{ij} of the matrix A is an estimate of the real ratio W_i/W_j . Since it is an estimate, the following is true:

$$a_{ij} = (W_i/W_j)d_{ij}, \text{ for } i, j = 1, 2, 3, \dots, N. \quad (1)$$

In the previous relation, d_{ij} denotes the deviation of a_{ij} from being a perfectly accurate judgment. Obviously, if $d_{ij} = 1$, the a_{ij} value was perfectly estimated. From the previous formulation, we conclude that the errors involved in these pairwise comparisons are given by

$$\varepsilon_{ij} = d_{ij} - 1.00,$$

or by using (1) above,

$$\varepsilon_{ij} = a_{ij} (W_j/W_i) - 1.00. \quad (2)$$

When the set of alternatives (or criteria) contains N elements, then $N(N-1)/2$ total pairwise comparisons need

to be estimated. The corresponding $N(N-1)/2$ errors are (after using relations (1) and (2)):

$$e_{ij} = a_{ij} (W_j/W_i) - 1.00, \text{ for } i, j = 1, 2, 3, \dots, N, \text{ and } j > i. \quad (3)$$

Since the W_i 's are relative weights which (in most cases) have to add up to 1.00, the following relation should also be satisfied:

$$\sum_{i=1}^N W_i = 1.00, \text{ and } W_i > 0, \text{ for } i = 1, 2, 3, \dots, N. \quad (4)$$

When the data (e.g. the pairwise comparisons) are perfectly consistent, then relations (3) and (4) can be written as follows:

$$B W = b. \quad (5)$$

The vector b has zero entries everywhere, except that the last entry is equal to 1.00; the matrix B has the following form (blank entries represent zeros):

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \dots & n-1 & n \\ -1 & a_{1,2} & & & & & & & & 1 \\ -1 & & a_{1,3} & & & & & & & 2 \\ -1 & & & a_{1,4} & & & & & & 3 \\ \cdot & & & & \cdot & & & & & \cdot \\ \cdot & & & & & \cdot & & & & \cdot \\ \cdot & & & & & & \cdot & & & \cdot \\ \cdot & & & & & & & \cdot & & \cdot \\ \cdot & & & & & & & & a_{1,n-1} & \cdot \\ -1 & & & & & & & & a_{1,n} & n-1 \\ -1 & a_{2,3} & & & & & & & & 1 \\ -1 & & a_{2,4} & & & & & & & 2 \\ -1 & & & a_{2,5} & & & & & & 3 \\ \cdot & & & & \cdot & & & & & \cdot \\ \cdot & & & & & \cdot & & & & \cdot \\ \cdot & & & & & & \cdot & & & \cdot \\ \cdot & & & & & & & \cdot & & \cdot \\ \cdot & & & & & & & & a_{2,n-1} & \cdot \\ -1 & & & & & & & & a_{2,n} & n-2 \\ & & & & & & & & & \cdot \\ & & & & & & & & & \cdot \\ & & & & & & & & & \cdot \\ & & & & & & & & & \cdot \\ & & & & & & & & -1 & a_{n-1,n} & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix}$$

The error minimization issue is interpreted in many cases (for instance, in regression analysis and in the linear least-square problem) as the minimization of the sum of squares of the residual vector $r = b - BW$ [33]. In terms of the previous formulation (5), this means that, in a real-life situation (i.e., when errors are not zero any more), the real intention of the decision maker is to minimize the following expression

$$f^2(x) = \|b - BW\|_2^2, \quad (6)$$

which, apparently, expresses a typical linear least-square problem.

In Triantaphyllou et al. [35] all the previous methods were tested in terms of an example originally presented by Saaty in [26] and also later used by other authors (e.g., [3] and [8]). In that test it was found that the proposed human rationality approach results in much smaller residuals. Moreover, in the same study it was found, on thousands of randomly generated test problems, that the eigenvalue approach may result in considerably higher residual values than the proposed least-squares approach that uses the previous human rationality assumption.

5. Problem #3: Processing the Decision Matrices

In Triantaphyllou and Mann [34] the AHP, revised AHP, Weighted Sum Model (WSM) [9], and the Weighted Product Model (WPM) [20] were examined in terms of two evaluative criteria. That study focused on the last step of any MCDM method which involves the processing of the final decision matrix. That is, given the weights of relative performance of the decision criteria, and the performance of the alternatives in terms of each one of the decision criteria, then determine what is the ranking (or relative priorities) of the alternatives.

As it was shown in Triantaphyllou and Mann [34], however, these methods can give different answers to the same problem. Since the truly best alternative is the same regardless of the method chosen, an estimation of the accuracy of each method is highly desirable. The most difficult problem that arises here is how one can evaluate a multi-dimensional decision-making method when the true best alternative is not known. Two evaluative criteria were introduced in [34] for the above purpose.

The first evaluative criterion has to do with the premise that *a method which is accurate in multi-dimensional problems should also be accurate in single-dimensional problems. There is no reason for an accurate multi-dimensional method to fail in giving accurate results in single-dimensional problems; since single-dimensional problems are special cases of multi-dimensional ones.* Because the first method, the WSM, gives the most acceptable results for the majority of single-dimensional problems, the result of the WSM was used as the standard for evaluating the other three methods in this context.

The second evaluative criterion considers the premise that *a desirable method should not change the indication of the best alternative when an alternative (not the best) is replaced by another worse alternative (given that the importance of each criterion remains unchanged).*

In Triantaphyllou and Mann [34] the previous two evaluative criteria were applied on random test problems with the numbers of decision criteria and alternatives taking the values 3, 5, 7, ..., 21. In those experiments it was found that all the previous four MCDM methods were inaccurate. Furthermore, these results were used to form a decision problem in which the four methods themselves were the alternatives. The decision criteria were derived by considering the two evaluative criteria. The numerical results (% of failure rates) are presented in the following table and are also depicted in figure 3. To our greatest surprise, one method would recommend another, rival method, as being the best method! However, the final results seemed to suggest that the revised AHP was the most efficient MCDM method of the ones examined. This was reported in Triantaphyllou and Mann [34] as a decision making paradox. Finally, a different approach of evaluating the performance of the AHP and the revised AHP, by using the previous concept of the RCP and CDP matrices, is described by Triantaphyllou and Mann in [39]. In that treatment it was found that these two methods may yield dramatically inaccurate results (more than 80% of the time).

Table 1.
Summary of the Computational Results (from [34]).

Criterion	1		2		3		4		5	
	1	2	3	4	1	2	3	4	1	2
Cases with										
Number of alter	3	3	3	...	21	3	3	3	...	21
Number of criter.	3	5	7	...	21	3	3	7	...	21
Subcriterion	1	2	3	...	100	101	102	103	...	200
Method										
WPM	14.7	12.7	12.1	...	25.7	0	0	0	...	0
AHP	8.2	10.5	12.3	...	10.9	0.16	0.26	0.26	...	0.02
Rev. AHP	7.4	8.2	8.8	...	3.0	0.50	0.50	0.50	...	0.60

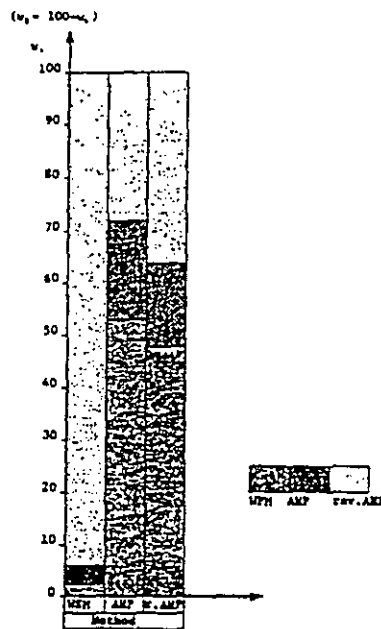


Figure 4. (from [34])

Indication of the Best MCDM Method According to Different Approaches.

6. Conclusions

There is no doubt that pairwise comparisons offer a unique opportunity to effectively elicit information regarding qualitative data in MCDM problems. Although they were first proposed (in the way described in this paper) as part of the AHP method, they can be used to quantify data in many other MCDM methods. However, there are some inherent difficulties regarding the application of pairwise comparisons to MCDM methods.

The present paper presented an overview of some of the problems studied by the authors in the past. These problems can be divided into three classes: (i) problems in quantifying pairwise comparisons, (ii) problems in processing reciprocal matrices with pairwise comparisons, and (iii) problems regarding the processing of the data in the final decision matrix.

Regarding the first class of problems, it was demonstrated by the authors and their associates in Triantaphyllou et al. [34] that there is no optimal way for quantifying pairwise comparisons. That is, there is no single scale which is always perfect. No matter what scale a decision maker uses, there is always potential for introducing inaccuracies. Of course, the effect of these inaccuracies on the quality of the final decision, depends on their magnitude.

The second class of problems reveals that there is not a uniquely best way for extracting the relative priorities from a complete set of pairwise comparisons. Different assumptions on what the real intentions of the decision maker were (or ought to be) may lead to totally different relative priorities. A least squares method proposed by the authors in Triantaphyllou et al. [35] gave robust results (under an assumption of human rationality) but it is rather complicated to be easily applied in practice.

Probably the most paramount conclusion is related to class (iii) of problems. Research by the authors has revealed that the AHP, revised AHP, WSM, and WPM methods may lead to incorrect conclusions, even if problems (i) and (ii) are not present. In Triantaphyllou and Mann [34] it was demonstrated that all the previous four methods may reach the wrong decision.

The results presented in this paper should not be interpreted as a proof that pairwise comparisons are not an efficient and effective way for eliciting information about qualitative data. Instead, these results must be viewed as evidence to the contrary of what has been sometimes claimed in the literature. To find the truly best solution to a MCDM problem may never be humanly possible. The conclusions of the solution should be taken lightly and used only as indications to what may be the best answer. Although the search for finding the best MCDM method may never end, research in this area of decision making is still critical and very valuable.

REFERENCES

1. Bell, B.E., and Farquhar, P.H., "Perspectives on Utility Theory", *Operations Research*, 34 (1986) 179-183.
2. Belton, V., and Gear, T., "On a short-coming of Saaty's method of analytic hierarchies", *Omega*, (1983) 228-230.
3. Chu, A.T.W., Kalaba, R.E., and Spingarn, K., "A comparison of two methods for determining the weights of belonging to fuzzy sets", *Journal of Optimization Theory and Applications*, 27 (1979) 531-538.
4. Dyer, J.S. and Ravinder, H.V., "Irrelevant alternatives and the Analytic Hierarchy Process", *Working Paper*, Department of Management, The University of Texas at Austin, (1983).
5. Dyer, J.S., and Wendell, R.E., "A critique of the Analytic Hierarchy Process", *Working Paper* 84/85-4-24, Department of Management, The University of Texas at Austin, (1985).
6. Dyer, J.S., "Remarks on the Analytic Hierarchy Process", *Management Science*, 36/3 (1990) 249-258.
7. Dyer, J.S., "A clarification of 'Remarks on the Analytic Hierarchy Process'", *Management Science*, 36/3 (1990) 274-275.
8. Federov, V.V., Kuz'min, V.B., and Vereskov, A.I., "Membership degrees determination from Saaty matrix totalities", Institute for System Studies, Moscow, USSR. Paper appeared in: *Approximate Reasoning in Decision Analysis*, M.M. Gupta, and E. Sanchez (editors), North-Holland Publishing Company, (1982) 23-30.
9. Fishburn, P.C., *Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments*, Operations Research, 1967.
10. Fishburn, P.C. and LaValle, I.H., (Eds.). *Annals of Operations Research. Volume 19: Choice Under Uncertainty*. J.C. Baltzer, Basel, Switzerland, 1989.
11. Harker, P.T., and Vargas, L.G., "Reply to 'Remarks on the Analytic Hierarchy Process'", *Management Science*, 36/3 (1990) 269-273.
12. Hogarth, R.M., and Reder, M.W., (Eds.) *Rational Choice: The Contrast Between Economics and Psychology*, University of Chicago Press, Chicago, Illinois, 1986.
13. Lootsma, F.A., "Numerical scaling of human judgment in pairwise-comparison methods for fuzzy multi-criteria decision analysis", *Mathematical Models for Decision Support*. NATO ASI Series F, Computer and System Sciences, Springer, Berlin, 48 (1988) 57-88.
14. Lootsma, F.A., Mensch, T.C.A. and Vos, F.A., "Multi-criteria analysis and budget reallocation in long-term research planning", *European Journal of Operational Research*, 47 (1990) 293-305.
15. Lootsma, F.A., "The French and the American school in multi-criteria decision analysis", *Recherche Operationnelle / Operations Research*, 24/3 (1990) 263-285.
16. Lootsma, F.A., "Scale sensitivity and rank preservation in a multiplicative variant of the AHP and SMART". *Report 91-67, Faculty of Technical Mathematics and Informatics*, Delft University of Technology, Delft, The Netherlands, 1991.
17. Ma, D. and Zheng, X., "9/9-9/1 scale method of the AHP", *Proceedings of the 2nd International Symposium on the AHP*, Vol. 1, Pittsburgh, PA, (1991) 197-202.
18. Marks, L.E., *Sensory Processes, The New Psychophysics*, Academic Press, New York, 1974.
19. Michon, J.A., Eijkman, E.G.J., and de Klerk, L.F.W., *Handboek der Psychonomie*, (in Dutch), Van Loghum Slaterus, Deventer, The Netherlands, 1976.
20. Miller, D.W., and Starr, M.K., *Executive Decisions and Operations Research*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1969.
21. Miller, G.A., "The magical number seven plus or minus two: some limits on our capacity for processing information", *Psychological Review*, 63 (1956) 81-97.
22. B.R. Munier (Ed.). *Risk, Decision and Rationality*. D. Reidel, Dordrecht, Holland, 1988.
23. Roberts, F.S., *Measurement Theory*, Addison-Wesley, Reading, Mass, 1979.
24. Stevens, S.S. and Hallowell Davis, M.D., *Hearing, its Psychology and Physiology*, American Institute of Physics, New York, 1983.
25. R.K. Sarin. Analytical Issues in Decision Methodology. In I. Horowitz (Ed.), *Organization and Decision*

- Theory*, Kluwer- Nijhoff, Dordrecht, Holland, (1989) 156-172.
26. Saaty, T.L., "A scaling method for priorities in hierarchical structures", *Journal of Mathematical Psychology*, 15, (1977) 57-68.
 27. Saaty, T.L., "Exploring the interface between hierarchies, multiple objects and fuzzy sets", *Journal of Fuzzy Sets and Systems*, 1, (1978) 57-68.
 28. Saaty, T.L., *The Analytic Hierarchy Process*, McGraw-Hill International, New York, NY, 1980.
 29. Saaty, T.L., "Axiomatic foundations of the Analytic Hierarchy Process", *Management Science*, 32/2 (1983), 841-855.
 30. Saaty, T.L., "Rank generation, preservation, and reversal in the Analytic Hierarchy Process", *Decision Sciences*, 18, (1987) 157-177.
 31. Saaty, T.L., "An exposition of the AHP in reply to the Paper 'Remarks on the Analytic Hierarchy Process'", *Management Science*, 36/3 (1990) 259-268.
 32. Simon, H.A., *Models of Man*, 2nd Edition, John Wiley and Sons, New York, NY, 1961.
 33. Stewart, S.M., *Introduction to Matrix Computations*, Academic Press, New York, NY, 1973.
 34. Triantaphyllou, E., and Mann, S.H., "An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox", *International Journal of Decision Support Systems*, 5, 303-312 (1989).
 35. Triantaphyllou, E., Pardalos, P.M., and Mann, S.H., "A minimization approach to membership evaluation in fuzzy sets and error analysis", *Journal of Optimization Theory and Applications*, 66(2), 275-287 (1990).
 36. Triantaphyllou, E., Pardalos, P.M., and Mann, S.H., "The Problem of Evaluating Membership Values in Real World Situations", *Operations Research and Artificial Intelligence: The Integration of Problem Solving Strategies*, (D.E. Brown and C.C. White III, editors), Kluwer Academic Publishers, 197-214 (1990).
 37. Triantaphyllou, E., and Mann, S.H., "An evaluation of the eigenvalue approach for determining the membership values in fuzzy sets", *Fuzzy Sets and Systems*. 35/3 (1990) 295-301.
 38. Triantaphyllou, E., Lootsma, F.A., Pardalos, P.M., and Mann, S.H., "On the Evaluation and Application of Different Scales For Quantifying Pairwise Comparisons in Fuzzy Sets", *Multi-Criteria Decision Analysis*, Vol. 3 (1994) 1-23.
 39. Triantaphyllou, E., and Mann, S.H., "An evaluation of the AHP and the revised AHP when the eigenvalue method is used under a continuity assumption", *Computers and Industrial Engineering*, 24 pages, to appear, Summer of 1994.
 40. Triantaphyllou, E., "A quadratic programming approach in estimating similarity relations", *IEEE Transactions on Fuzzy Systems*, 1/2 (1993) 138-145.
 41. Weber W. and Camerer, C., "Recent developments in modeling preferences under risk", *OR Spectrum*, 9 (1987) 129-151.
 42. Winkler, R.L., "Decision Modeling and Rational Choice: AHP and Utility Theory" *Management Science*, 36/3 (1990) 247-248.
 43. Write, C., and Tate, M.D., *Economics and Systems Analysis: Introduction for Public Managers*, Addison-Wesley, Reading, MA, 1973.
 44. Vargas, L.G., "Reciprocal Matrices with Random Coefficients", *Mathematical Modeling*, 3, 69-81 (1982).
 45. Zwicker, E., *Psychoakustic*, Springer, Berlin, Germany, 1982.