Published in:

Journal of Multi-Criteria Decision Analysis, **Vol. 3, No. 3, pp. 133-155, 1994.**

On the Evaluation and Application of Different Scales For Quantifying Pairwise Comparisons in Fuzzy Sets

E. Triantaphyllou[1], F.A. Lootsma[2], P.M. Pardalos[3], and S.H. Mann[4]

[1:] Dept. 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
Web: http://www.imse.lsu.edu/vangelis

[2:] Faculty of Technical Mathematics and Informatics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

[3:] Dept. of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, Gainesville, FL 32611, U.S.A.

[4:] Professor of Operations Research, Director of SHRIM, The Pennsylvania State University, 118 Henderson Building, University Park, PA 16802, U.S.A.

ABSTRACT: One of the most critical issues in many applications of fuzzy sets is the successful evaluation of membership values. A method based on pairwise comparisons provides an interesting way for evaluating membership values. That method was proposed by Saaty, almost 20 years ago, and since then it has captured the interest of many researchers around the world. However, recent investigations reveal that the original scale may cause severe inconsistencies in many decision-making problems. Furthermore, exponential scales seem to be more natural for humans to use in many decision-making problems. In this paper two evaluative criteria are used to examine a total of 78 scales which can be derived from two widely used scales. The findings in this paper reveal that there is no single scale that can outperform all the other scales. Furthermore, the same findings indicate that a few scales are very efficient under certain conditions. Therefore, for a successful application of a pairwise comparison based method the appropriate scale needs to be selected and applied.

KEY WORDS: Fuzzy sets, pairwise comparisons, membership values, multi-criteria decision-making.

1. Introduction.

One of the most crucial steps in many decision-making methods is the accurate estimation of the pertinent data. Very often these 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.

Consider the case of having a single decision criterion and a set of N alternatives, denoted as A_i ($i = 1,2,3,\ldots N$). The decision maker wants to determine the relative performance of the alternatives in terms of the single criterion. In a case like this, one may consider the N alternatives as the members of a fuzzy set. Then, the degree of membership of element (i.e., alternative) A_i expresses the degree that alternative Ai meets this criterion. That is, in the previous context the membership degrees can be viewed as the degree the members of a set of objects meet a single criterion. This is also the approach considered by Federov et al (1982), Saaty (1974) and (1978), and was also discussed by Chen and Hwang (1992).

An approach based on pairwise comparisons which was proposed by Saaty (1977), and (1980) has long attracted the interest of many researchers, because both of its easy applicability and interesting mathematical properties. Pairwise comparisons are used to determine the relative importance of each alternative in terms of each criterion. In this approach the decision maker(s) 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. When one focuses directly on the membership issue one may use linguistic statements such as "How much more does alternative A belong to the set S than alternative B?". The main focus in this paper is not the wording of these linguistic statements, but, instead, the numerical values which should be associated with such statements. The importance of evaluating the membership values in applications of fuzzy set theory in engineering and scientific fields is best illustrated in the 1,800 references given by Gupta et al (1979).

The main problem with the pairwise comparisons is **how to quantify** the linguistic choices selected by the decision maker during the evaluation of the pairwise comparisons. All the methods which use the pairwise comparisons approach eventually express the qualitative answers of a decision maker into some numbers. The present paper 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 fuzzy data.

Pairwise comparisons are quantified by using a **scale**. Such a scale is nothing but 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 Analytic Hierarchy Process (AHP). The second approach attempts was proposed by Lootsma (1988), (1990), and (1991) and determines **exponential scales**. Both approaches depart from some psychological theories and develop the numbers to be used based on these psychological theories.

The present paper is organized as follows. The second section illustrates the principals of the two classes of scales. The second section also presents some ways for generating even more scales based on Saaty's linear scale and on the exponential scales proposed by Lootsma. The third section discusses ways for evaluating the performance of various scales. This is achieved in terms of two evaluative criteria. The next section (section 4) describes the problem of selecting the appropriate scale (or scales) as a multi-criterion decision-making problem. Computational results presented in the fifth section reveal that under different conditions some scales are more efficient than others. These findings are presented in depth in the final section which is the conclusion section.

2. Background Information.

As it was mentioned in the previous section, two classes of scales are considered in this paper. The first class of scales is defined on the interval [9, 1/9] and is based on the original Saaty scale. The second class of scales is based on the exponential scales introduced by Lootsma (1988) and (1991).

Once the pairwise comparisons are determined by using a scale, they are processed in order to derive the final values. These values are estimates of the relative magnitudes of the membership values. Saaty (1980) proposes the use of a method which is based on eigenvalues. Another method, which is based on a logarithmic regression model, is proposed by Lootsma (1988) and (1991). For a critical discussion of the eigenvalue approach, along with some other approaches, see (Triantaphyllou, Pardalos and Mann, 1990a) and (Triantaphyllou and Mann, 1993).

Also, an approach which uses **differences** instead of ratios is presented in (Triantaphyllou, (1993)). That approach describes how similarity relations among a group of entities can be estimated by using an efficient quadratic programming formulation. All these approaches are capable of estimating **relative** magnitudes. Therefore, final values could only be derived, if at least one of them were known apriori. However, this is not possible in real applications, thus only relative magnitudes are derivable by using pairwise comparisons.

It should be stated here that when pairwise comparisons are used the entire process may become impractical when the number of elements becomes large. If N is the number of elements, then the number of comparisons is N(N-1)/2. For instance, for N = 100 the decision maker would have to make 4,950 pairwise comparisons. Nor is the approach applicable to the elicitation of membership functions on real intervals unless the domain is discretized.

2.1. 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, (1980)). 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, (1956)). 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 instructions depicted in Table 1 (Saaty, (1980)). 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 (1991). 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 in a scale 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:

Table 1.

Scale of Relative Importances (according to Saaty (1980))

Intensity of Importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
7	Demonstrated importance	An activity is strongly favored and its dominance demonstrated in practice
9	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2,4,6,8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals of above nonzero	If activity i has one of the above nonzero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i.	

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

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.

2.2. Exponential Scales.

A class of exponential scales has been introduced by Lootsma (1988) and (1991). 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 (1979), the difference $e_{n+1} - e_n$ must be greater than or equal to the smallest perceptible difference, which is proportional to e_n. The permitable choices by the decision maker are summarized in Table 2. As a result of Robert's observation the numerical equivalents of these linguistics choices need to satisfy the following relations:

$$e_{n+1} - e_n = \epsilon\, e_n, \quad (\text{where } \epsilon > 0) \quad \text{or:}$$

$$e_{n+1} = (1 + \epsilon)\, e_n = (1 + \epsilon)^2\, e_{n-1} = \ldots$$

$$\ldots = (1 + \epsilon)^{n+1}\, e_o, \quad (\text{where: } e_o = 1) \qquad \text{or:} \quad e_n = e^{\gamma * n}$$

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). Table 3 presents the values of two exponential scales that correspond to two different values of the γ parameter. Apparently, different exponential scales can be generated by assigning different values to the γ parameter.

Another difference between exponential scales and the Saaty scale is on 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. In the following section we present some evidence that human beings follow exponential scales when they categorize an interval. More on these examples can be found in Lootsma (1990) and (1991).

Table 2.
Scale of Relative Importances (According to Lootsma (1988))

Intensity of Importance	Definition
e_0	Indifference between A_i and A_j
e_1	Indifference threshold towards A_i
e_2	Weak preference for A_i
e_3	Commitment threshold towards A_i
e_4	Strong preference for A_i
e_5	Dominance threshold towards A_i
e_6	Very strong preference for A_i
Reciprocals of above nonzero	If member i has one of the above nonzero numbers assigned to it when compared with member j, then j has the reciprocal value when compared with i.

Table 3.
Two Exponential Scales

NORMAL ($\gamma = 1/2$)	STRETCHED ($\gamma = 1$)	Definition
$e^{\gamma*0}$ = 1.00	= 1.00	e_0
$e^{\gamma*1}$ = 1.65	= 2.72	e_1
$e^{\gamma*2}$ = 2.72	= 7.39	e_2
$e^{\gamma*3}$ = 4.48	= 20.09	e_3
$e^{\gamma*4}$ = 7.39	= 54.60	e_4
$e^{\gamma*5}$ = 12.18	= 148.41	e_5
$e^{\gamma*6}$ = 20.09	= 403.43	e_6

2.3. Examples of the Use of Exponential Scales.

It is surprising to see how consistently humans categorize certain intervals of interest in totally unrelated areas. In this section we present some examples to show, for instance, how human subjects partition certain ranges on the time, sound, and light intensities.

a) **Historical periods.** The written history of Europe, from 3000 BC until today, is subdivided into a small number of major periods. Looking backwards from 1989, the year when the Berlin Wall was opened, we distinguish the following turning points marking off the start of a characteristic development:

1947	42	years before 1989	beginning of cold war and decolonization,
1815	170	years before 1989	beginning of industrial and colonial dominance,
1500	500	years before 1989	beginning of world-wide trade and modern science,
450	1550	years before 1989	beginning of middle ages,
-3000	5000	years before 1989	beginning of ancient history.

These major echelons, measured by the number of years before 1989, constitute a geometric sequence with the progression factor 3.3. We obtain a more refined subdivision when we introduce the years:

1914	75	years before 1989	beginning of world wars,
1700	300	years before 1989	modern science established,
1100	900	years before 1989	beginning of high middle ages,
-800	2800	years before 1989	beginning of Greek/Roman history.

With these turning points interpolated between the major ones, we find a geometric sequence of echelons with progression factor 1.8.

b) **Planning horizons.** In industrial planning activities, we usually observe a hierarchy of planning cycles where decisions under higher degrees of uncertainty and with more important consequences for the company are prepared at increasingly higher management levels. The planning horizons constitute a geometric sequence, as it shown in the following list:

1 week		weekly production scheduling,
1 month	4 weeks	monthly production scheduling,
4 months	16 weeks	ABC planning of tools and labor,
1 year	52 weeks	capacity adjustment,
4 years	200 weeks	production planning,
10 years	500 weeks	strategic planning of company structure.

The progression factor of these major horizons is 3.5. In practice there are no planning horizons between these major ones.

c) **Size of nations.** The above categorization is not only found on the time axis, but also in spatial dimensions when we categorize the nations on the basis of the size of their population. Omitting the very small nations with less than one million inhabitants, we have:

small nations	4 million	DK, N,
medium-size nations	15 million	NL, DDR,
large nations	60 million	D, F, GB, I
very large nations	200 million	USA, RUSSIA,
giant nations	1000 million	China, India.

We find again a geometric sequence, with progression factor 4.0. Furthermore, it seems reasonable to interpolate the following threshold echelons:

small/medium size	8 million	A, B, GR,
medium size/large	30 million	E, PL,
large/very large	110 million	Japan,

because the respective nations fall typically between the major echelons. The refined sequence of echelons has the progression factor 2.0.

d) **Loudness of sounds.** The range of audible sounds can roughly be categorized as follows:

40 dB	very quiet; whispering,
60 dB	quiet; conversation,
80 dB	moderately loud; electric mower and food blenders,
100 dB	very loud; farm tractors and motorcycles,
120 dB	uncomfortable loud; jets during take-off.

Although the precision should be taken with a grain of salt because we have a mixture of sound frequencies at each of these major echelons, we obviously find here a geometric sequence of subjective sound intensities with the progression factor 4.

e) **Brightness of light.** Physically, the perception of light and sound proceed in different ways, but these sensory systems follow a similar pattern. The range of visible light intensities can roughly be categorized as follows:

30 dB	star light,
50 dB	full moon,
70 dB	street lighting,
90 dB	office space lighting,
110 dB	sunlight in summer.

Under the precaution that the precision should not be taken too seriously because we have at each of these major echelons a mixture of wave lengths, we observe that the subjective light intensities also constitute a geometric sequence with the progression factor 4.

In the previous paragraphs we have used 5 examples to demonstrate that exponential scales are common in human comparative judgment when dealing with historical periods, planning horizons, size of nations, perception of light and sound intensities. Therefore, these examples make exponential scales only plausible. Lootsma (1991) has studied the scale sensitivity of the resulting scores when exponential scales are used. He observed that the rank order of the scores is not affected by variations of the scale parameter; the numerical values of the calculated scores are weakly dependent on that parameter.

For a more detailed documentation on psychophysics we refer the reader to Marks (1974), Michon et al (1976), Roberts (1979), Zwicker (1982), and Stevens and Hallowell Davis (1983). The reader will find that the sensory systems for the perception of tastes, smells, and touches follow the power law with exponents in the vicinity of 1.

3. Evaluating Different Scales.

In order different scales to be evaluated, two evaluative criteria are developed. Furthermore, a special class of pairwise matrices is developed in the next section. These special matrices are then used in conjunction with the two evaluative criteria in order to investigate some stability properties of different scales.

3.1. The Concepts of the RCP and CDP Matrices.

As it was mentioned earlier, reciprocal matrices with pairwise comparisons were introduced by Saaty (1977) as a tool for extracting all the pertinent information from a decision maker. The same author also proposed a scale which results in matrices with entries from the set Θ, where Θ is the set of integers 1,2,3,...,9 and their reciprocals (see also Table 1). If a different scale is to be used, then Θ will be the finite set of discrete values which represent that scale. Each entry in these matrices represents numerically the value of a pairwise comparison between two alternatives with respect to a single criterion. These matrices are constructed as to be an effective way of capturing the necessary information (Saaty, (1980)).

The Saaty matrices have received wide acceptance as being an effective way of evaluating membership values in real-world problems (see, for example, Chu et al (1979), Federov et al (1982), Hihn and Johnson (1988), Khurgin and Polyakov (1986), Lootsma et al (1990), and Vargas (1982)).

The analyses in Triantaphyllou et al (1990b) were based on the assumption that in the real world the membership values in a fuzzy set take on continuous values. Let $\omega_1, \omega_2, \omega_3, \ldots, \omega_n$ be the real (and thus unknown) membership values of a fuzzy set with n members. Each of the ω_i values is assumed to be in the interval [1,0]. If the decision maker knew the above real values then, he would be able to have constructed a matrix with the real pairwise comparisons. In this matrix, say matrix A, the entry $\alpha_{ij} = \omega_i/\omega_j$.

That is, the entry α_{ij} represents the real (and thus unknown) value of the comparison when the i-th member is compared with the j-th member. We call this matrix the **Real Continuous Pairwise** matrix, or the RCP matrix. Since in the real world the ω_i's are unknown, so are the entries α_{ij} of the previous matrix. However, we will assume here that the decision maker, instead of an unknown entry α_{ij} is able to determine the closest value taken from the set Θ of the numerical values provided by a scale. In other words, instead of the real (and thus unknown) value α_{ij} one is able to determine the value a_{ij} such that:

the difference * α_{ij} - a_{ij} * is minimum and $a_{ij} \in \Theta$.

Therefore, judgments about the values of the pairwise comparison of the i-th element when it is compared with the j-th element are **assumed** to be so accurate that they are closest (in absolute value terms) to the true or real values one is supposed to estimate when a scale with the discrete values Θ is used.

It should be stated at this point that other norms, alternative to the previous one, are also possible to be assumed as the way a decision maker best approximates real (and thus unknown) pairwise comparisons. For instance, such an alternative norm is the following:

the difference * $\alpha_{ij}/(1 + \alpha_{ij})$ - $a_{ij}/(1 + a_{ij})$ * is minimum and $a_{ij} \in \Theta$.

However, any **norm** which attempts to approximate the real (and thus unknown) ratios with ratios taken from a finite and discrete set of values, ***will always allow for the possibility that some real ratios (which are close enough to each other) will be mapped to the same discrete value from the current scale.*** The last statement indicates that Theorem 1 (stated later in section 3.2.) will still be valid if alternative norms are considered (however, its present proof assumes that the first norm is used).

The matrix with the entries a_{ij} that we assume the decision maker is able to construct has entries from the discrete and finite set Θ. We call this matrix the **Closest Discrete Pairwise** matrix or the CDP matrix. The CDP matrix may not be perfectly consistent. That is, the consistency index (CI) values (see the next section for an exact definition of CI) of CDP matrices are not necessarily equal to zero. More on this inconsistency issue will be discussed in the following section. 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. The following example illustrates further the concepts of the RCP and CDP matrices.

An Example. Let us assume that the real (and hence unknown) membership values, after normalization, of a fuzzy set with three members are ω_1 = 0.77348, ω_2 = 0.23804, and ω_3 = 0.23848. Then, the RCP matrix with the real values of the pairwise comparisons is:

$$RCP = \begin{bmatrix} 1 & 3.24938 & 3.24342 \\ 0.30775 & 1 & 0.99817 \\ 0.30832 & 1.00183 & 1 \end{bmatrix}$$

This is true because 0.30775 = (0.23804/0.77348), 0.30832 = (0.23848/0.77348), and so on. If, for instance, the original Saaty scale is to be used (as it is depicted in Table 1) then, it can be verified with a simple exhaustive enumeration that the corresponding CDP matrix is:

$$CDP = \begin{bmatrix} 1 & 3 & 3 \\ \frac{1}{3} & 1 & 1 \\ \frac{1}{3} & 1 & 1 \end{bmatrix}$$

To see this consider the (1,2) entry of the previous RCP matrix. For this entry we have α_{12} = 3.24938. Therefore, when the values in Table 1 are to be used in order to quantify the (1,2) pairwise comparison then, the α_{12} entry is approximated by the value 3. The value 3 is the closest one to the value 3.24938 when the values in Table 1 are used. Clearly, this is an assumption which is made here in order to study different scales. A similar explanation holds for the rest of the entries in the previous CDP matrix.

3.2. On The Consistency of CDP Matrices.

If all the pairwise comparisons are perfectly consistent with each other then, the following relation should always be true among any three comparisons $a_{i,k}$, $a_{k,j}$, and $a_{i,j}$ (Saaty (1980)):

$$a_{i,k} * a_{k,j} = a_{i,j} \qquad \text{for any } 1 \leq i,j,k \leq N.$$

Saaty expresses the inconsistency of a pairwise comparison matrix in terms of the consistency index (CI). CI is defined as follows:

$$CI = \frac{\lambda_{\max} - N}{N-1},$$

where λ_{max} is the maximum eigenvalue of the matrix with the pairwise comparisons and N is the order of that matrix.

In the following paragraphs we will show that CDP matrices can be **inconsistent** regardless of the scale used to quantify the actual pairwise comparisons. This is stated in terms of the following theorem:

THEOREM 1:

Regardless of the scale that is used to quantify the pairwise comparisons of N (N $\geq$ 3) entities, the corresponding CDP matrices may be inconsistent.

PROOF:

Without loss of generality, suppose that A_1, A_2, and A_3 are three items of a collection of N (N $\geq$ 3) items that we need to compare in terms of some criterion. Let the current scale be defined on the following (2k + 1) **discrete** values (where: k $\geq$ 1) :

$$[1/V_k, 1/V_{k-1}, 1/V_{k-2},\ldots, 1/V_2, 1/V_1, 1, V_1, V_2,\ldots, V_{k-2}, V_{k-1}, V_k],$$

where $V_i > 0$ for any i = 1,2,3,...,k.

In this proof it will be shown that, when the previous scale is used, then it is possible for the three comparisons a_{12}, a_{13}, a_{32} made by the decision maker **not to satisfy** the consistency requirement:

$$a_{12} = a_{13} * a_{32}.$$

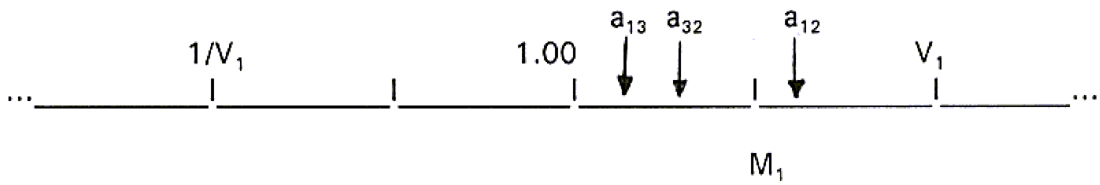
Suppose that the **actual values** of the pairwise comparisons that correspond to the previous three items A_1, A_2, and A_3 are as follows:

$$\frac{A_1}{A_2} = \alpha_{12} = \frac{3V_1 + 1}{4}$$

$$\frac{A_1}{A_3} = \alpha_{13} = \frac{V_1 + 3}{4}$$

$$\text{and} \quad \frac{A_3}{A_2} = \frac{(A_1/A_2)}{(A_1/A_3)} = \alpha_{32} = \frac{\alpha_{12}}{\alpha_{13}} = \frac{3V_1 + 1}{V_1 + 3}.$$

Using the above relations it can be easily verified (since V1 > 1) that the following conditions (I) are true (see also figure 1):



$$\frac{V_1 + 1}{2} > \alpha_{32} > 1.$$

NOTE: $M_1 = (V_1 + 1)/2$, is the middle point of the interval $[1.00, V_1]$.

Figure 1.

Actual Comparison Values.

From figure 1, or conditions (I), it follows that in the corresponding CDP matrix the decision maker will assign the following three values a_{12}, a_{13}, a_{32} (taken from the current scale) to the previous three pairwise comparisons:

$a_{12} = V_1$

$a_{13} = 1.00$

$a_{32} = 1.00.$

Clearly, the consistency requirement **does not hold** for these three values because:

$a_{12} \neq a_{13} * a_{32}.$

In other words, the entire CDP matrix is **inconsistent**.

EOP (End Of Proof).

The previous theorem states that under the favorable assumption that the decision maker is capable of determining only the **closest** values of the pairwise comparisons, the resulting CDP matrices may be inconsistent. The following paragraphs of this section discuss the issue of the **maximum** consistency, denoted as CI_{max}, of CDP matrices. The following lemma provides an interesting result regarding the maximum error δ_{max} associated with the pairwise comparisons of a CDP matrix. The maximum error δ_{max} is defined as:

$$\delta_{max} = MAX(e_{ij} - 1) \text{ where } e_{ij} = a_{ij}(W_j / W_i),$$

for any i,j = 1,2,3,..., N.

The a_{ij}'s are the entries of a pairwise matrix and W_i, W_j are the **real** weights of the items i and j, respectively.

LEMMA 1:

Let a scale for quantifying pairwise comparisons be defined on the following (2k+1) discrete values:

$$[\ 1/V_k,\ 1/V_{k-1},\ 1/V_{k-2},\dots,\ 1/V_2,\ 1/V_1,\ 1,\ V_1,\ V_2,\dots,\ V_{k-2},\ V_{k-1},\ V_k\ ],$$

where $V_i > 0$, for any $i=1,2,3,\dots, k$.

Then, the maximum error, δ_{max}, of the pairwise comparisons in a CDP matrix is given by the formula:

$$\delta_{max} = MAX \left\{ \frac{V_j - V_{j-1}}{V_j + V_{j-1}} \right\} \quad for\ j=1,2,3,...,k, \qquad and\ V_0 = 1.$$

PROOF:

Suppose that a pairwise comparison has **actual** (and hence unknown) value equal to α, where: $1/V_j \geq \alpha \geq 1/V_{j-1}$ for some $k \geq j \geq 1$. Let M be the middle point of the interval $[1/V_j,\ 1/V_{j-1}]$. That is:

$$M = \frac{1/V_{j-1} - 1/V_j}{2} + 1/V_j = \frac{V_j + V_{j-1}}{2\ V_j\ V_{J-1}}.$$

Then, the largest δ value for this particular pairwise comparison occurs when the value of α coincides with the middle point M. This is true because in this case the closest value from the values permitted by the current scale has the largest distance from α. That is, under the assumption that the decision maker will choose the **closest value,** the value of this pairwise comparison will become equal either to

$1/V_j$ or $1/V_{j-1}$. In the first case the corresponding δ, we call it δ_1, becomes:

$$\delta_1 = \frac{1/V_j}{M} - 1 = \frac{\frac{1}{V_j}}{\frac{V_j + V_{j-1}}{2\ V_j\ V_{j-1}}} - 1 = \frac{V_{j-1} - V_j}{V_{j-1} + V_j}.$$

Similarly, in the second case the value of δ, we call it δ_2, becomes:

$$\delta_2 = \frac{V_j - V_{j-1}}{V_{j-1} + V_j}, \quad that\ is:\ |\delta_1| = |\delta_2|.$$

Since, in general, it is assumed that: $1/V_k \leq \alpha \leq V_k$, it is derived that the maximum value of δ, δ_{max}, is given by the following formula:

$$_{ax} = MAX\left\{\frac{V_j - V_{j-1}}{V_j + V_{j-1}}\right\}, \qquad for\ j=1,2,3,...,k, \quad and\ V_o=1.$$

Finally, it is worth mentioning that both the expressions δ_1 and δ_2 remain the same if the values V_j and V_{j-1} are replaced by their reciprocals.

EOP.

In the previous considerations, and throughout this paper, it is assumed that the real values of the pairwise comparisons are within the interval $[1/V_k,\ V_k]$. If, instead, the real ratios were allowed to be from the range zero to infinity, then the associated errors could be infinitely large. In other words, the real ratios are assumed to take values according to the scale under consideration.

Although this may appear to be restrictive, it eliminates the possibility of having infinitely large errors when the decision maker attempts to approximate pairwise comparisons by using a discrete and finite scale. Furthermore, this is a plausible assumption since, most of the time, the elements in a fuzzy set are assumed to be somehow closely associated (i.e., similar) with each other and do not allow for extreme cases. Therefore, it makes sense not to permit to have infinitely large errors in the estimation process.

Next, lemma 1 is used to prove theorem 2 which deals with the value of CI_{max} of random CDP matrices.

THEOREM 2:

Let a scale for quantifying pairwise comparisons be defined on the following (2k + 1) discrete values:

$$[\ 1/V_k,\ 1/V_{k-1},\ 1/V_{k-2},\ldots,\ 1/V_2,\ 1/V_1,\ 1,\ V_1,\ V_2,\ldots,\ V_{k-2},\ V_{k-1},\ V_k\ ],$$

where $V_i > 0$ for any $i = 1,2,3,\ldots, k$.

Then an upper bound of the maximum consistency index, CI_{max}, of the resulting CDP matrices is given by the following relation:

$$CI_{\max} \leq \frac{\delta^2_{\max}}{2},$$

$$\text{where: } \delta_{\max} = MAX \left\{ \frac{V_j - V_{j-1}}{V_j + V_{j-1}} \right\},$$

$$\text{for } j = 1,2,3,\ldots,k, \quad \text{and } V \quad 1.$$

PROOF:

The proof of this theorem is based on theorem 7-16, stated in Saaty (1980). According to that theorem the following relation is always true:

$$\lambda_{max} - N \leq (N - 1)/2\ \delta^2_{max}, \tag{1}$$

where δ_{max} is defined as:

$$\delta_{max} = MAX(e_{ij} - 1), \text{ and } e_{ij} = a_{ij}(W_j / W_i), \quad \text{for any } i,j = 1,2,3,\ldots, N.$$

The a_{ij}'s are the entries of the pairwise matrix and W_i, W_j are the **real** weights of items i and j, respectively. From relation (1), above, we get:

$$\frac{\lambda_{\max} - N}{N - 1} \le \frac{\delta_{\max}^2}{2} \quad or:$$

$$CI_{\max} \le \frac{\delta_{\max}^2}{2}. \tag{2}$$

For the case of CDP matrices the value of the maximum δ, denoted as δ_{max}, can be determined as follows (see also lemma 1):

$$\delta_{\max} = MAX \left\{ \frac{V_j - V_{j-1}}{V_j + V_{j-1}} \right\}, \qquad for\ j = 1, 2, 3, ..., k, and\ V_o = 1. \tag{3}$$

Therefore, the maximum consistency index, CI_{max}, of CDP matrices satisfies the relation:

$$CI_{\max} \le \frac{\delta_{\max}^2}{2},$$

where δ_{max} is given by (3), above.

EOP.

In the original Saaty scale a pairwise comparison takes on values from the discrete set: Θ = {1/9, 1/8, 1/7, ..., 1/3, 1/2, 1, 2, 3, ..., 7, 8, 9}. Therefore, it can be verified easily that the following corollary 1 is true when the original Saaty scale is used.

COROLLARY 1:

When the original Saaty scale is used, an upper bound of the maximum consistency index, CI_{max}, of the corresponding CDP matrices is:

$$CI_{\max} \le \frac{(1/3)^2}{2} = \frac{1}{18}.$$

Figure 2 depicts the maximum, average, and minimum consistency indexes of randomly generated CDP matrices which were based on the original Saaty scale. That is, first a RCP matrix was randomly generated. Next, the corresponding CDP matrix was derived and its CI value was calculated and recorded (see also Triantaphyllou et al (1990)). This experiment was performed 1,000 times for each value of N equal to 3, 4, 5, ..., 100. It is interesting to observe that the curves which correspond to the maximum and minimum CI values of samples of 1,000 randomly generated CDP matrices, are rather irregular. This was anticipated since it is very likely to find one extreme case from a sample of 1,000 CI value of random CDP matrices. One the other hand, however, the middle curve, which depicts the **average** CI values of random CDP matrices, is very regular. This was also anticipated because the impact of a few extreme CI values diminishes when a large sample (i.e., 1,000) of random CDP matrices is considered. Moreover, the same results indicate that the average CI value approaches the number 0.0145 when the value of N is greater than 20. More on the CI values of random Saaty matrices (i.e., not necessarily CDP matrices) can be found in Donegan and Dodd (1991).

The results in the current section reveal that CDP matrices (which are assumed to be the result of a highly effective elicitation of the pertinent pairwise comparisons) are very unlikely to be perfectly consistent. That is, some small inconsistency may be better than no inconsistency at all! (since no CDP matrix with CI = 0 was found when sets with more than five elements were considered). This is kind of a paradoxical phenomenon which is, however, explained why it occurs theoretically by the lemmas and theorems in this section.

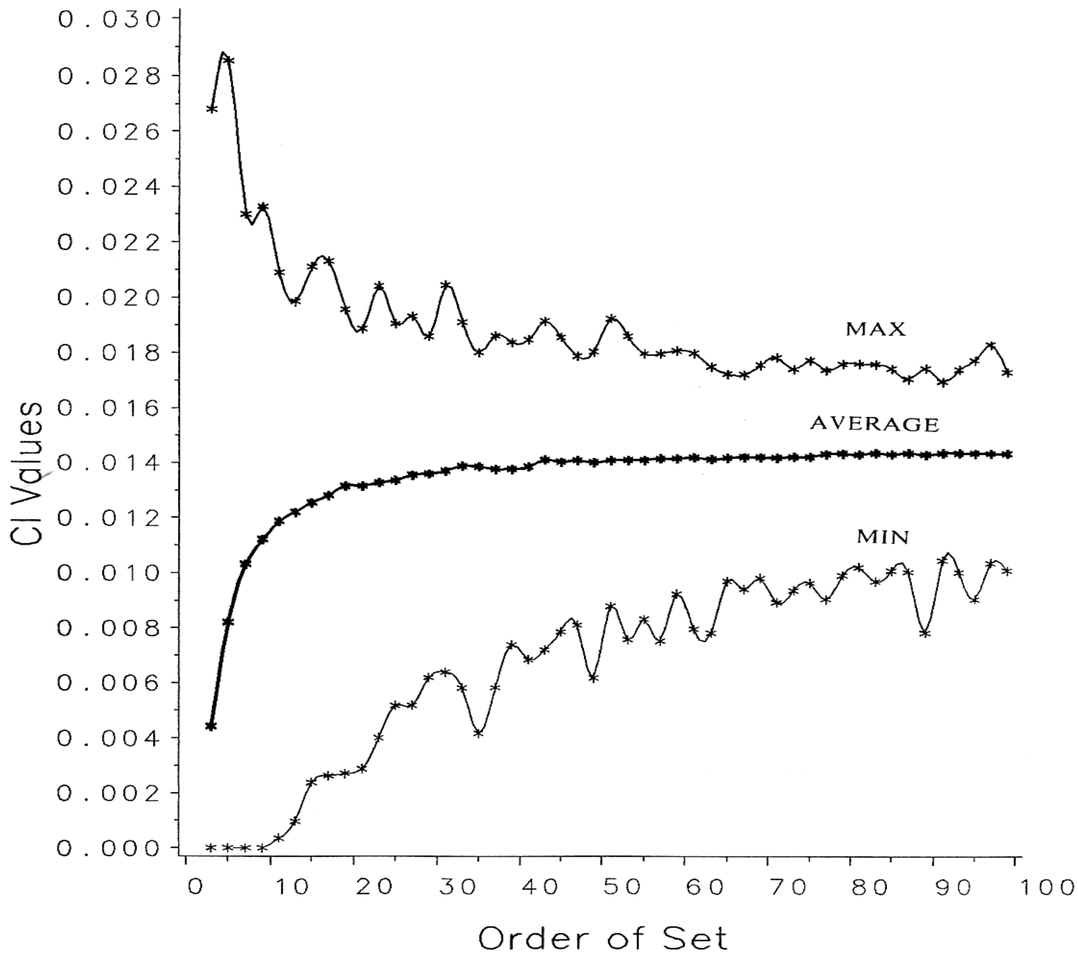


Figure 2. Maximum, Average, and Minimum CI Values of Random CDP Matrices When the Original Saaty Scale is Used.

3.3. Evaluative Criteria.

In Triantaphyllou and Mann (1990), the evaluation of the effectiveness of Saaty's eigenvalue method was based on a continuity assumption. Under this assumption the eigenvalue approach in some cases causes worse alternatives to **appear** better than alternatives that are **truly** better in reality.

Two kinds of ranking inconsistency were examined. The first kind is **"ranking reversal"**. For example, if the real ranking of a set of three members is (1, 3, 2) and a method yields (1, 2, 3) then a case of a ranking reversal occurs. The second kind is **"ranking indiscrimination"**. For example, if the real ranking of a set of three members is (1, 3, 2) and a method yields (1, 2, 2), that is, a tie between two or more members, then a case of ranking indiscrimination occurs. 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 RCP matrix. Therefore, the following two evaluative criteria can be introduced to investigate the effectiveness of any scale which attempts to quantify pairwise comparisons:

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 do 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 do 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**.

4. A Simulation Evaluation of Different Scales.

Different scales were evaluated by generating test problems and then recording the inversion and indiscrimination rates as described in criteria 1 and 2. Suppose that a scale defined on the interval [9, 1/9] (as described in section 2.1.) or an exponential scale (as described in section 2.2.) is defined on the interval [X, 1/X]. That is, the numerical value that is assigned to a pairwise comparison that was evaluated as: **"A is absolutely more important than B"** (i.e., the highest value) is equal to X. For instance, in the original Saaty scale (as well as in all the other scales in section 2.1.) X equals to 9.00. Under the assumption that a scale on the interval [X, 1/X] is used, the pairwise comparisons also take numerical values from the interval [X, 1/X]. In this case the entries of RCP matrices (as defined in section 3.1.) are **any numbers** from the interval [X, 1/X]. However, in CDP matrices the entries take values only from the discrete and finite set that is defined on the interval [X, 1/X]. We call it set Θ. For example, in the case of the original Saaty scale the entries of CDP matrices are members of the set Θ = {1/9, 1/8, 1/7, ..., 1/2, 1, 2, ..., 7, 8, 9}.

For the above reasons test problems for the case of the first and second evaluative criterion were generated as follows. First, N random membership values of N elements were randomly generated from the interval [0, 1]. These membership values were such that no ratio of any pair of them would be larger than X or less than 1/X. After the random membership values were generated, the corresponding RCP matrix was constructed. Next, from the RCP matrix and the discrete and finite set Θ the corresponding CDP matrix was determined. Then,the eigenvalue approach was applied on this CDP matrix and the new ranking of the N elements. The eigenvalue method was used because it is rather simple to apply and is the method used widely in the literature when only one decision maker is considered. The recommended ranking of the N elements is compared with the **actual** ranking which is determined from the real membership values that were generated in the beginning of this process. If a ranking inversion or ranking indiscrimination was observed, it was recorded so. This is exactly the testing procedure followed in the investigation of the original Saaty scale as it is reported in Triantaphyllou and Mann (1990).

A FORTRAN program was written which generated the N random membership values, the RCP and CDP matrices, and compared the two rankings as described above. Sets with $N = 3, 4, 5, \ldots, 30$ elements were considered. For each such set 21 scales defined on the interval [9, 1/9] (which correspond to the values $\alpha = 0, 5, 10, 15, \ldots, 90, 95, 100$) and 57 exponential scales which correspond to γ values equal to 0.02, 0.04, 0.06, ..., 1.10, 1.12, 1.14 were generated. The previous scales will also be indexed as scale 1, scale 2, scale 3, ..., scale 78.

In figures 3 and 4 the results of the evaluations of scales 1,2,3,..,21 (also called **class 1 scales**) in terms of the first and second criterion, respectively, are presented. Similarly, in figures 5 and 6 the results of the evaluations of scales 22, 23, 24,.., 78 (also called **class 2 scales**) in terms of the first and second criterion, respectively, are presented. It should be noted here that only 57 exponential scales were generated because in this way values of γ from zero to around to 1.00 can be considered. In the original Lootsma scales the value of γ was 0.50 and 1.00. In this investigation all the scales with $\gamma = 0.02, 0.04, 0.06, \ldots, 0.50, \ldots, 1.00, \ldots, 1.14$ are considered. For each case of a value of N and one of the 78 scales, 1,000 random test problems were generated and tested according to the procedure described in the previous paragraphs. The computational results of this investigation are depicted in figures 5 and 6.

At this point it should be emphasized that the present simulation results are contingent on how the random membership values were generated. Other possibilities, such as assigning membership values from a nonuniform distribution (such as the normal distribution), would probably favor other scales. However, the uniform distribution from the interval [0, 1] was chosen in this study (despite the inherited restrictions of this choice) because it is the simplest and most widely used in simulation investigations.

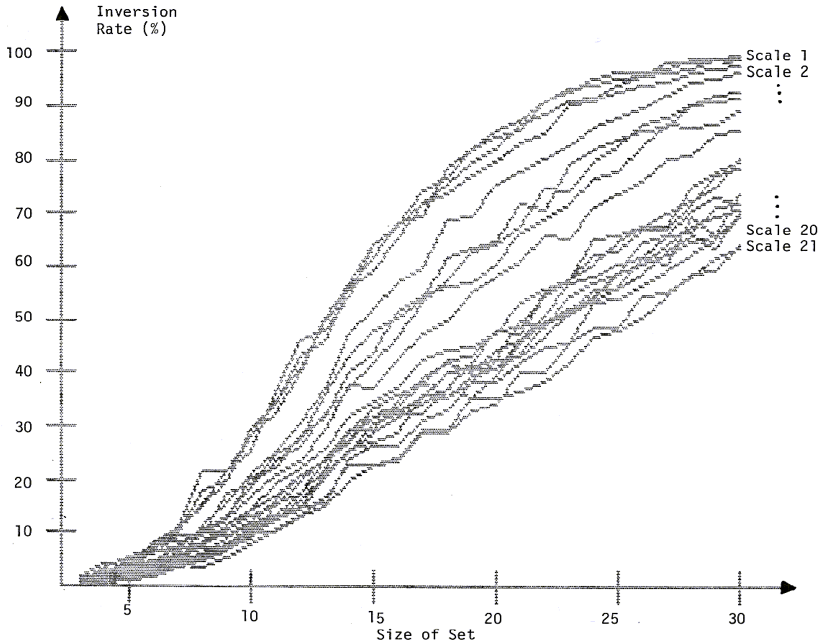


Figure 3.
Inversion Rates for Different Scales and Size of Fuzzy Set (Class 1 Scales).

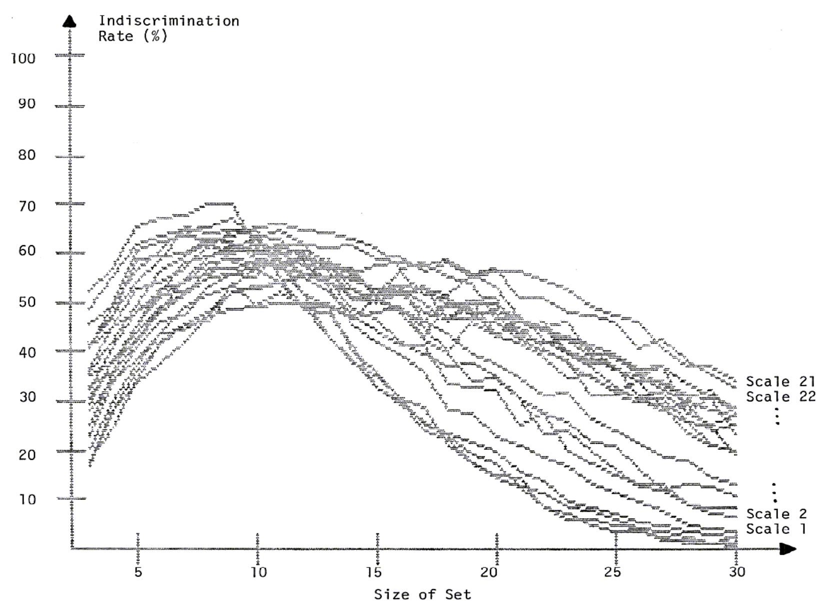


Figure 4.
Indiscrimination Rates for Different Scales and Size of Fuzzy Set (Class 1 Scales).

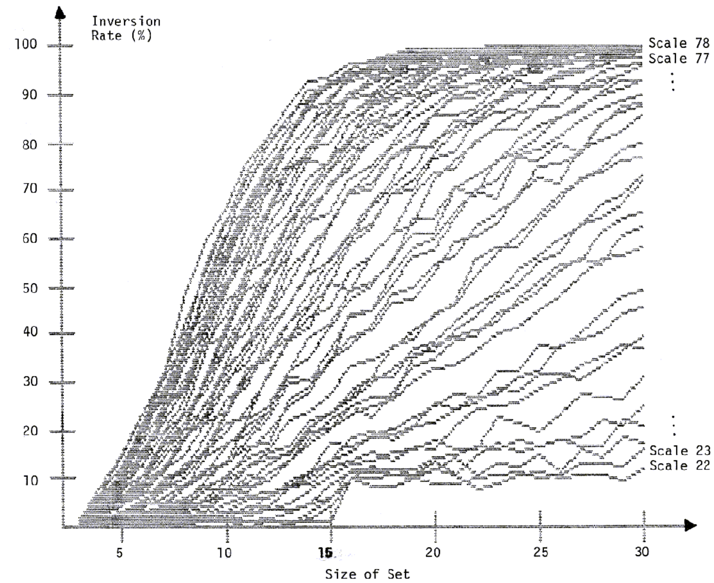


Figure 5.
Inversion Rates for Different Scales and Size of Fuzzy Set (Class 2 Scales).

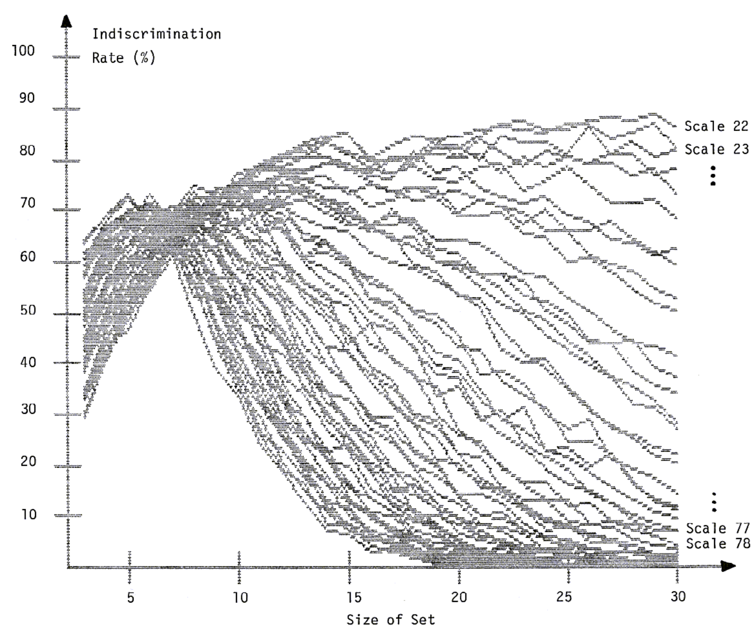


Figure 6.
Indiscrimination Rates for Different Scales and Size of Fuzzy Set (Class 2 Scales).

5. Evaluation of the Computational Results.

Figures 3, 4, 5, and 6 depict how the previous 78 different scales perform in terms of the two evaluative criteria. Figures 3 and 4 depict the inversion and indiscrimination rates (as derived after applying the two evaluative criteria) for class 1 scales. That is, for the scales defined in the interval [9, 1/9]. Similarly, figures 5 and 7 depict the inversion and indiscrimination rates for the exponential scales (or class 2 scales). It is also interesting also to observe here that when both classes of scales are evaluated in terms of the second criterion (indiscrimination rates in figures 4 and 6), then they perform worse when the size of the set is in the region 8 to 12.

Clearly, there is no single scale which outperforms all the other scales for any size of set. Therefore, there is no scale or a group of scales which is better than the rest of the scales in terms of **both evaluative criteria**. However, the main problem is to determine which scale or scales are more efficient.

Since there are 78 different scales for which there are relative performance data in terms of two evaluative criteria, it can be concluded that this is 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. For instance, if ranking indiscrimination of the elements is not of main concern to the decision maker, then the weight of the ranking reversals should assume its maximum value (i.e., becomes equal to 1.00). 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 $1.00 \geq W_1 \geq 0.00$). In this way, a total of 100 different combinations of weights were examined.

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 (1983)). In Triantaphyllou and Mann (1989) 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 7. Similarly, the results regarding the worst scales are depicted in figure 8. 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 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. It is also interesting to observe from figure 7 that sometimes under similar weights of the two evaluative criteria, the same scale might be classified as the best. The same is also true for the worst scales depicted in figure 8. This phenomenon suggests that sometimes an **approximated assessment** of the relative weights is adequate to successfully determine either the best or worst scale.

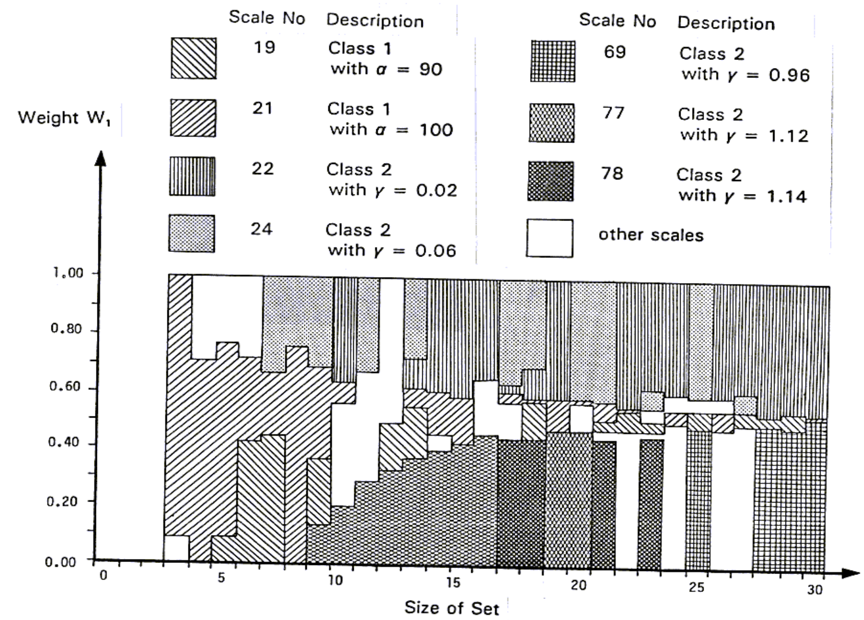


Figure 7.
The Best Scales

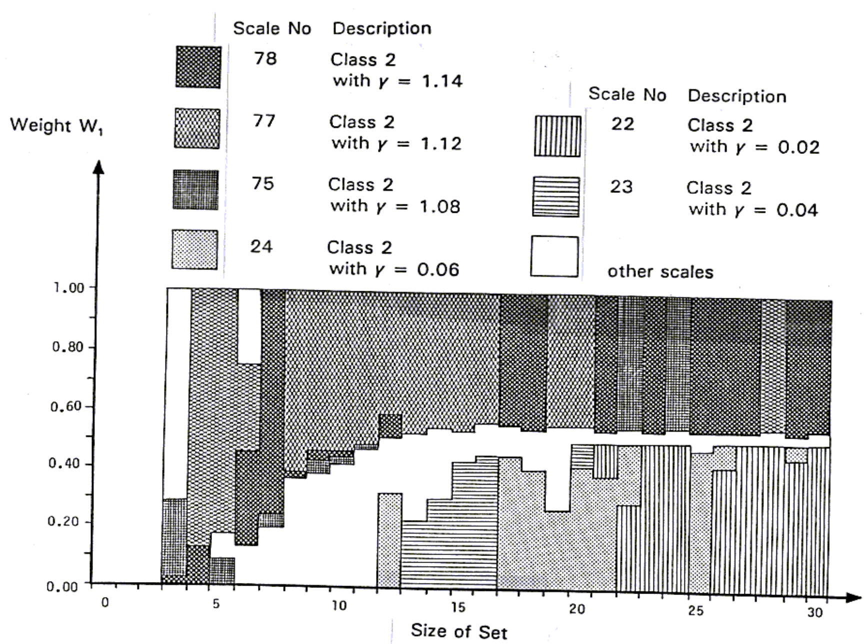


Figure 8.
The Worst Scales

6. Concluding Remarks.

This paper revealed that the scale issue is a complex problem. The results demonstrated that there is no single scale which can always be classified as the best scale or as the worst scale for all cases. The present investigation is based on the assumption that there exists a real-valued rating of the comparison between two entities, that ideally represents the individual preference. However, the decision-maker cannot express it, hence he has to use a scale with finite and discrete options.

In order to study the effectiveness of various scales, we furthermore assumed the scenario in which the decision maker is able to express his judgments as accurate as possible. Under this scenario, it is assumed that the decision maker is able to construct CDP matrices with pairwise comparisons instead the unknown RCP matrices. Based on this setting, a number of computational experiments was performed to study how the ranking derived by using CDP matrices differs from the real (and hence unknown) ranking implied by the RCP matrices. The computational results reveal that there is no single scale which is best in all cases. It should be emphasized here that given an RCP matrix (and a scale with numerical values), then there is **one and only one** CDP matrix which best approximates it. Moreover, this CDP matrix may or may not yield a different ranking than the ranking implied by the RCP matrix.

An alternative assumption to the current one, which accepts that there exists a real-valued rating of the comparison between entities, is to consider the premise that maybe the real entity is the CDP matrix as given by the decision maker. In this case the RCP matrix is maybe just an illusion. In the later case the preference reversal leads to a very different conclusion: if the CDP is the only "real" thing, then it means that the individual should point at the interval $[1/V_i, 1/V_{i-1}]$ or $[V_{i-1}, V_i]$ rather than to the values V_i. That is, the preference reversal effects indicate that two objects will be indifferent (since their ranking changes in the interval).

To determine the appropriate scale in a given situation certain factors have to be analyzed. First the number N, of the items to be compared, has to be known. Secondly, the relative importance of the two evaluative criteria has to be assessed. These evaluative criteria deal with possible ranking inversions and ranking indiscriminations that may result when a scale is used. When these factors have been assessed figure 7 depicts the best scale for each case. Similarly, figure 8 depicts the worst scale for each case.

For instance, suppose that one has to evaluate the membership values of a set with 15 members. Furthermore, suppose that ranking reversal is considered, in a particular application, far more severe than ranking indiscrimination. In other words, the weight of the first evaluative criterion is considered to be higher than the weight of the second criterion. Using this information, we can see that figure 7 suggests to use scale 22 from class 2 (i.e., an exponential scale with parameter $\gamma = 0.02$). Moreover, figure 8 suggests that the **worst** scale for this case is scale 77 from class 2 (i.e., an exponential scale with parameter $\gamma = 1.12$).

The same figures also indicate that the choice of the best or worst scale is not clear under certain conditions. For instance, when the number of members is greater than 15 and the two evaluative criteria are of almost equal importance. In cases like this, it is recommended to experiment with different scales in order to increase the insight into the problem, before deciding on what is the best scale for a given application.

The computational experiments in this paper indicate (as shown in figure 7) that exponential scales are more efficient than the original Saaty scale (i.e., Scale 1). Only two Saaty-based scales (i.e., scales 19 and 21) are present in figure 7. In matter of fact, for sets with up to 10 elements Scale 21 was best over a wide range of weights. It is also worth noting that all the worst scales in figure 8 came from the exponential class.

However, as the various examples in section 2.3 suggest, human beings seem to use exponential scales in many diverse situations. Therefore, exponential scales appear to be the most reasonable way for quantifying pairwise comparisons. The computational results in this paper provide a guide for selecting the most appropriate exponential scale for quantifying a given set of pairwise comparisons.

Finally, it needs to be emphasized here that the scale problem is a very crucial issue when membership values of the members of a fuzzy set are determined by using pairwise comparisons. These membership values can provide the data for many real life decision-making problems. An alternative point of view of this study would be to perform in the future a similar investigation with methods which **do not** use pairwise comparisons and thus are counterparts of the pairwise comparison methodologies. However, since pairwise comparisons provide a flexible and also realistic way for estimating these type of data, it follows that an in depth understanding of all the aspects of the scale problem is required for a successful solution of a decision-making problem.

Acknowledgements

The authors would like to thank the referees for their thoughtful comments which significantly improved the quality of this paper.

REFERENCES

Belton, V., and Gear, T. (1983). "On a Short-coming of Saaty's Method of Analytic Hierarchies", Omega, 11, 228-230.

Chen, S.-J. and Hwang, C.-L. (1992). "Fuzzy Multiple Attribute Decision Making: Methods and Applications", Lecture Notes in Economics and Mathematical Systems, No 375, Sringer-Verlag, Berlin.

Chu, A.T., Kalaba, R.E., and Spingarn, K. (1979). "A Comparison of Two Methods for Determining the Weights of Belonging to Fuzzy Sets", Journal of Optimization Theory and Applications, 27/4, 531-538.

Donegan, H.A., and Dodd, F.J. (1991). "A Note on Saaty's Random Indexes", Mathematical and Computer Modeling, 15/10, 135-137.

Gupta, M.M., Ragade, R.K. and Yager, R.Y., editors. (1979). "Fuzzy Set Theory and Applications", North-Holland, New York.

Federov, V.V., Kuz'min, V.B., and Vereskov, A.I. (1982). "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.

Hihn, J.M., and Johnson, C.R. (1988). "Evaluation Techniques for Paired Ratio-Comparison Matrices in a Hierarchical Decision Model", Measurement in Economics, Psysics-Verlag, 269-288.

Khurgin, J.I., and Polyakov, V.V. (1986). "Fuzzy Analysis of the Group Concordance of Expert Preferences, defined by Saaty Matrices", Fuzzy Sets Applications, Methodological Approaches and Results, Akademie-Verlag Berlin, 111-115.

Lootsma, F.A. (1988). "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, 57-88.

Lootsma, F.A., Mensch, T.C.A. and Vos, F.A. (1990). "Multi-Criteria Analysis And Budget Reallocation In Long-Term Research Planning", European Journal of Operational Research, 47, 293-305.

Lootsma, F.A., (1990). "The French and the American School in Multi- Criteria Decision Analysis", Recherche Operationnele / Operations Research, 24/3, 263-285.

Lootsma, F.A. (1991). "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.

Ma, D. and Zheng, X. (1991). "9/9-9/1 Scale Method of the AHP", Proceedings of the 2nd International Symposium on the AHP, Vol. 1, Pittsburgh, PA, 197-202.

Marks, L.E. (1974). "Sensory Processes, The New Psychophysics", Academic Press, New York.

Michon, J.A., Eijkman, E.G.J., and de Klerk, L.F.W. (eds.), (1976). "Handboek der Psychonomie", (in Dutch), Van Loghum Slaterus, Deventer.

Miller, G.A. (1956). "The Magical Number Seven Plus or Minus Two: Some Limits on our Capacity for Processing Information", Psychological Review, 63, March, 81-97.

Roberts, F.S. (1979). "Measurement Theory", Addison-Wesley, Reading, Mass.

Saaty, T.L. (1974). "Measuring the Fuzziness of Sets", Cybernetics, Vol. 4, No. 4, 53-61.

Saaty, T.L., (1977). "A Scaling Method for Priorities in Hierarchical Structures", Journal of Mathematical Psychology, 15/3, 234-281.

Saaty, T.L. (1978). "Exploring the Interfaces Between Hierarchies, Multiple Objectives and Fuzzy Sets", Fuzzy Sets and Systems, Vol. 1, No. 1, 57-68.

Saaty, T.L. (1980). "The Analytic Hierarchy Process", McGraw Hill International, 1980.

Stevens, S.S. and Hallowell Davis, M.D. (1983). "Hearing, its Psychology and Physiology". American Institute of Physics, New York.

Triantaphyllou, E. and Mann, S.H. (1989). "An Examination of the Effectiveness of Multi-Dimensional Decision-Making Methods: A Decision-Making Paradox", International Journal of Decision Support Systems, 5, 303-312.

Triantaphyllou, E., Pardalos, P.M., and Mann, S.H. (1990a). "A Minimization Approach to Membership Evaluation in Fuzzy Sets and Error Analysis", Journal of Optimization Theory and Applications, 66/2, 275-287.

Triantaphyllou, E., Pardalos, P.M., and Mann, S.H. (1990b). "The Problem of Determining Membership Values in Fuzzy Sets 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.

Triantaphyllou, E., and Mann, S.H. (1990). "An Evaluation of the Eigenvalue Approach for Determining the Membership Values in Fuzzy Sets", Fuzzy Sets and Systems, 35/3, 295-301.

Triantaphyllou, E., and Mann, S.H. (1993). "An Evaluation of the AHP and the Revised AHP when the Eigenvalue Method is Used under a Continuity Assumption", Computers and Industrial Engineering, under review.

UPDATED REFERENCE:

Triantaphyllou, E., and Mann, S.H. (1994). "An Evaluation of the AHP and the Revised AHP when the Eigenvalue Method is Used under a Continuity Assumption", Computers and Industrial Engineering, 26/3, 609-618.

Triantaphyllou, E. (1993). "A Quadratic Programming Approach In Estimating Similarity Relations", IEEE Transactions on Fuzzy Systems, 1/2, 138-145.

Vargas, L.G. (1982). "Reciprocal Matrices with Random Coefficients", Mathematical Modeling, 3, 69-81.

Zwicker, E. (1982). "Psychoakustik", Springer, Berlin.