The Use of Pairwise Comparisons for Decision Making May Lead to Grossly Inaccurate Results

Evangelos Triantaphyllou[1, 2, *] and Juri Yanase[3]

1: Division of Computer Science & Engin.
College of Engineering
Louisiana State University
Baton Rouge, LA 70803, USA
Email: etriantaphyllou@yahoo.com

2: Department of Medicine
Section of Hematology & Med Oncology
School of Medicine
Tulane University
New Orleans, LA 70112, USA

3: Complete Decisions, LLC
Baton Rouge, LA 70810, USA
http://www.completedecisions.com
Email: jurijuriy@aol.com

* : Corresponding Author

Date: November 7, 2024

Abstract:
Pairwise comparison matrices (also known as Saaty matrices) in conjunction with a finite discrete scale for their quantification, are considered by many as efficient and effective means for eliciting personal preferences from decision makers. This study demonstrates that under a highly optimistic assumption, called the ultra-accurate decision maker (UADM) assumption, the results obtained from such matrices may be grossly inaccurate. The true ranking of the compared entities may be altered and / or an abnormal rank grouping may occur when processing pairwise comparison matrices. The rates that such effects take place are exceptionally dramatic. Given that this occurs under the deliberately highly optimistic conditions of the UADM assumption, the implication is that under real-life circumstances these rates can be even more dramatic. The results are supported with theoretical and extensive computational investigations. The computational investigations are based on three novel and powerful tests introduced for the first time in this study. This is the first needed step to better understand the potential and limitations of pairwise comparisons before one attempts to remedy this problem.

Keywords:
Decision analysis; pairwise comparison matrix; ultra-accurate decision maker (UADM) assumption; ranking problem; Saaty matrices.

Please cite this paper as follows:

Triantaphyllou, E., and J. Yanase, 2024, **The Use of Pairwise Comparisons for Decision Making May Lead to Grossly Inaccurate Results**, *Computers & Industrial Engineering*, Vol. 198, December 2024, 110653.
https://doi.org/10.1016/j.cie.2024.110653

First available online on Oct. 22, 2024. This file is available on ResearchGate.net

1. Introduction

Since the introduction in the late 1970s of the analytic hierarchy process (AHP) (Saaty, 1977) for multi-criteria decision making (MCDM) / multi-criteria decision analysis (MCDA), the AHP has received enormous acceptance by researchers and practitioners. The AHP became more formally described in the late 1980s to early 2000s (Saaty, 1989; 2003). The AHP's wide and diverse acceptance is evident of the hundreds of thousands of citations of publications that Google's Scholar returns when one searches using keywords / key phrases such as "analytic hierarchy process," or "AHP." When one focuses the search within the Computers and Industrial Engineering journal, then more than six hundred publications appear under the previous search terms. This success of the AHP method in real-world managerial problems, is based on the intuitive appeal the method has to researchers and practitioners and also to its interesting mathematical properties. The development of dedicated software that is based on this method (such as Expert Choice (https://www.expertchoice.com/2021), or Decision Lens (https://www.decisionlens.com/)), has further enabled its wide adoption by these communities.

Table 1: The Saaty Scale for assigning numerical values to Pairwise Comparisons (adapted from (Saaty, 1977)).

Intensity of Importance on an absolute scale	**Definition**	**Interpretation**
1	Equal importance	Two entities contribute equally to the objective
3	Weak importance of one over the other	Experience and judgement slightly favor one entity over another
5	Essential or strong importance	Experience and judgement strongly favor one entity over another
7	Demonstrated importance	An entity is strongly favored, and its dominance is demonstrated in practice
9	Absolute importance	The evidence is favoring one entity 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 entity i has one of the above nonzero numbers assigned to it when compared with entity j, then j has the reciprocal value when compared with i	

A central concept in the AHP is the use of pairwise comparisons for eliciting intangible / qualitative / subjective preference data from decision makers. This part of the AHP process can be used in conjunction with any other method that needs to use such data for decision making. That is, the pairwise comparisons approach is not strictly limited to the AHP.

The concept of the pairwise comparisons is very simple but still powerful. It has long been understood that humans cannot assess relative weights of importance when they consider many entities (such as decision alternatives, evaluative criteria, and so on) simultaneously. The seminal publication by Miller (1956) is considered the foundation, from the psychological point of view, of the pairwise comparisons approach. Based on this, Saaty argued that humans can achieve the best results if they focus on a single pairwise comparison at-a-time and quantify them by means of a discrete ratio scale. For that, he proposed the scale depicted in Table 1. Next, all possible pairwise comparisons between all the pairs of the entities under consideration are considered simultaneously and the relative weights or human preferences of the entities are derived by employing certain mathematical procedures, such as the principal eigenvector approach.

For instance, if there are, say four entities denoted as A_1, A_2, A_3, and A_4, then the decision maker elicits pairwise comparisons by answering questions such as the following: "If entity A_i is compared to entity A_j in terms of a specific aspect that both share, then what is the result of this comparison (denoted as $a_{i,j}$) if the scale depicted in Table 1 is used?" If the answer, say *"weak importance of one over the other,"* is chosen as being the best answer from the scale in Table 1, then the pairwise comparison $a_{i,j}$ is assigned the numerical value 3 (see also Table 1). A similar interpretation applies for the rest of the entries in Table 1.

In this way, according to Saaty the decision maker needs to elicit $n(n-1)/2$ pairwise comparisons and form a pairwise comparison matrix, where n is the number of the entities that need to be considered. This matrix is a reciprocal one as Saaty recommends using the reciprocal value of $a_{i,j}$ for the entry $a_{j,i}$ once the value of the entry $a_{i,j}$ has been elicited. In terms of the previous hypothetical example, the entry $a_{i,j}$ will be assigned the value of 3, while the entry $a_{j,i}$ will be assigned the value of 1/3. The first diagonal entries will all be assigned the value of 1. That is, if there are n entities to be compared, such matrix is not a symmetric one, but it is rather a reciprocal one because by construction $a_{i,j} = 1/a_{j,i}$, and $a_{i,i} = 1$, for any $i, j = 1, 2, 3, \ldots, n$.

The validity of the pairwise comparisons approach has been challenged multiple times. Most of these studies focused on the inconsistency aspect of pairwise comparisons (see, for instance, (Linares, 2009), and (Kułakowski, 2018)). However, the present study uncovers some rather astonishing facts about the pairwise comparisons approach. That is, even when it is used in a setting that assumes highly consistent pairwise matrices, the pairwise comparisons approach may still yield strongly questionable results.

Other studies have focused on what is known as the rank reversal problem with the AHP. This was first reported in (Belton and Gear, 1983). This problem has received lots of attention and it became a controversial subject (see, for instance, (Triantaphyllou and Mann, 1989), (Barzilai and Golany, 1994), (Schenkerman, 1994), (Millet and Saaty, 2000), (Triantaphyllou, 2001), and (Faramondi, et al., 2022)). For a more recent literature review of this topic, one may check the following publication (Aires and Ferreira, 2018). This review considered 130 studies published between the years 1980 and 2015.

The previous studies consider the decision-making problem as a whole. That is, with a finite set of alternatives and a finite set of evaluative criteria. In contrast, <u>the present study focuses on ranking abnormalities that occur when a single and highly consistent pairwise comparison matrix is considered.</u>

The fact that the pairwise comparisons approach may result in strongly questionable results becomes evident if one considers, what is called in the present study, the **ultra-accurate decision maker (or UAMD) assumption.** That is, it will be assumed that the decision maker is ultimately accurate when he/she elicits individual pairwise comparisons. However, under this deliberately ultra-favorable assumption, it is demonstrated in this study by means of illustrative examples, theoretical analyses and extensive empirical / computational experimentations that the pairwise comparisons approach may still yield highly problematic results.

Therefore, **the main objective of this study** is to demonstrate how certain types of ranking abnormalities may occur under the UADM assumption and as an extension, in real-world applications of methods which are based on the use of pairwise comparisons.

This paper is organized as follows. The following section discusses some of the background developments that lead to the proposed approaches. Section 3 presents the key definitions and assumptions used in this study. Section 4 describes how a key condition, that of the reciprocity of the values when two entities are considered, may be violated under the UADM assumption. Section 5 examines this issue when a complete pairwise comparison matrix is considered. Certain types of ranking abnormalities are examined in Sections 6 and 7. A discussion and the conclusions are presented in Section 8, where the key contributions of this study are summarized.

2. Some Background Information

According to Lootsma (1990) there are two "schools" of thought in MCDA. He called them the French and the American school. Such distinction had started with (Schärlig, 1985) and (Colson and De Bruyn, 1989), but it became formal by Lootsma with his 1990 publication. The French school is based on the work by Roy when he introduced the ELECTRE family of methods (Roy, 1985; 1989; 1990), (Figueira, et al., 2010; and 2013), and continued with Brans, et al., with the PROMETHEE method (Brans, et al., 1986). The main idea is the use of the so-called "outranking relationships."

On the other hand, the American school is based on the work by Keeney and Raiffa (Keeney, et al., 1976; and 1993) and Saaty's AHP approach (Saaty, 1980). The main idea is first to determine the pertinent data in the form of a decision matrix and then use an aggregation approach to compute the final priority scores (weights) of the alternatives.

Some key concepts for the proposed approach were first highlighted more than 30 years ago in (Triantaphyllou and Mann, 1990), (Triantaphyllou, et al., 1994), and (Triantaphyllou and Mann, 1994). However, these concepts have not been explored adequately until the present study was undertaken. To see the main issue, consider a simple hypothetical example. Suppose that in reality when two entities, say A_1 and A_2, are compared, their true and thus unknown to the decision maker value of the pairwise comparison (denoted as $a_{1,2}$) is equal to, say 2.85. The decision maker has no means to 100% accurately determine this value. However, the decision maker can use as a preference elicitation tool Saaty's ratio scale (see also Table 1) and choose the verbal option that best approximates this (unknown) value of 2.85. Thus, it is assumed that the decision maker would choose the option *"weak importance of one over the other,"* which is assigned to the value of 3 in Table 1, because this option is numerically closest to the (real but unknow) value of 2.85. In other words, the decision maker will assign the numerical value of 3 to the pairwise comparison $a_{1,2}$ and at the same time, according to Saaty, the value of 1/3 to the symmetric pairwise comparison $a_{2,1}$.

In the previous early publications (e.g., (Triantaphyllou and Mann, 1990), (Triantaphyllou, et al., 1994), and (Triantaphyllou and Mann, 1994) two key concepts regarding pairwise matrices were introduced. This is the concept of the **Real Continuous Pairwise (or RCP)** matrix and the concept of the **Closest Discrete Pairwise (or CDP)** matrix. CDP matrices are highly consistent, with an average CI (consistency index) value less than 0.014 and a maximum CI value less than 0.028 in the experiments reported in (Triantaphyllou, et al., 1994).

To see their meaning, assume that the goal is to elicit the relative weights of three entities, say A_1, A_2, and A_3. Suppose that the real weights (denoted as $\omega(A_i)$, for i = 1, 2, 3) are as follows:

$$\omega(A_1) = 0.1449, \qquad \omega(A_2) = 0.5652, \text{ and} \qquad \omega(A_3) = 0.2899.$$

Given the above values, the RCP matrix is a pairwise matrix where the $a_{i,j}$ entry is the ratio of the corresponding true values (i.e., $a_{i,j} = \omega(A_i) / \omega(A_j)$). Thus, the RCP matrix that corresponds to the above (real) data is as follows:

$$\textbf{Matrix RCP} = \begin{bmatrix} 1 & 0.2564 & 0.5000 \\ 3.9002 & 1 & 1.9496 \\ 2 & 0.5129 & 1 \end{bmatrix}$$

For instance, the entry (1,2) is equal to 0.2564 because the ratio $\omega(A_1) / \omega(A_2)$ is equal to 0.1449 / 0.5652 (= 0.2564). The values of this matrix are unknown to the decision maker.

The CDP matrix is the matrix whose entries are the closest values when the Saaty scale (or any other discrete ratio scale for that matter) is used to approximate the corresponding entries of the RCP matrix. That is, the CDP matrix attempts to approximate the RCP matrix by using a ratio scale (like the one in Table 1). Given the previous RCP matrix, the corresponding CDP matrix when Saaty's scale is used is as follows:

$$\textbf{Matrix CDP} = \begin{bmatrix} 1 & 0.2500 & 0.5000 \\ 4 & 1 & 2 \\ 2 & 0.5000 & 1 \end{bmatrix}$$

For instance, the entry $a_{1,2}$ of the CDP matrix is equal to 0.2500 (or 1/4) because this is the closest value from the Saaty scale in Table 1 to the real value 0.2564.

In various publications Saaty (e.g., (Saaty, 1977; and 2008)) has accepted the notion of having "pre-existing preferences" and hence the foundation of the RCP and CDP matrices used in the present study. In those publications Saaty considered a total of 6 real-world case studies where the real preferences exist and can be computed objectively. Next, he employed his method of pairwise comparisons to accurately estimate these preference values.

The whole idea in those Saaty publications was to demonstrate how accurately his pairwise comparison approach could estimate the true / objective values. In those case studies such values are the relative protein contents of some foods, the distances between some test physical objects, the distances between some world cities, the values of the GDP of some countries (at that time), the weights of some ordinary objects, and so on. Some additional examples of such real data approximated by means of pairwise comparisons were given again by Saaty in a power point presentation entitled "The Analytic Hierarchy Process (AHP) for Decision Making," (URL: http://www.cashflow88.com/decisiones/saaty1.pdf).

3. Key Definitions and Assumptions

This section begins by providing the formal definitions of the RCP and CDP matrices highlighted in the previous section.

Definition 1:

Given a vector $\Omega(A)$ with the weights $\omega(A_i)$ of n entities, where $\omega(A_i) > 0$ for $i = 1, 2, 3, \ldots, n$, the **Real and Continuous Pairwise (RCP) matrix** is the reciprocal matrix of dimension $n \times n$ with entries $a_{i,j}$ defined as follows:

$$a_{i,j} = \omega(A_i) / \omega(A_j), \text{ for } i, j = 1, 2, 3, \ldots, n.$$

Definition 2:

Given the Saaty scale, with numerical entries from the set $S = \{1/9, 1/8, 1/7, \ldots, 1/2, 1, 2, \ldots, 7, 8, 9\}$ and an RCP matrix with entries $a_{i,j}$, for $i, j = 1, 2, 3, \ldots, n$, the corresponding **Closest Discrete Pairwise (CDP) matrix** is the square matrix of dimension $n \times n$ with entries $b_{i,j}$ defined as follows:

$$b_{i,j} = b, \text{ where } b \in S \text{ and the difference } |\, b - a_{i,j} \,| \text{ is minimum.}$$

Definition 2 indicates that the entries of the CDP matrix are members of the numerical entries of the Saaty scale (i.e., members of the set S) that are closest to the corresponding entries of the RCP matrix. For instance, if the $a_{i,j}$ entry of the

RCP matrix is equal to, say the value 2.85, then the corresponding entry $b_{i,j}$ of the CDP matrix will be equal to 3. What is not very obvious at this point, is that while the RCP matrix is always a reciprocal one by construction (i.e., the relationship $a_{i,j} = 1/a_{j,i}$ is satisfied for all i, j values), this may not be true with the entries of the CDP matrix. That is, it is possible to have $b_{i,j} \neq 1/b_{j,i}$. This is investigated later as it has profound implications. A secondary issue is that in the simulation studies described later in this paper attention is given to have all the entries of an RCP matrix be from the continuous interval [1/9, 9] in order to be compatible with the numerical values indicated by Saaty's scale (see also Table 1).

The use of the proposed CDP testing approach does not depend on the traditional Saaty scale which is defined in the interval [1/9, 9]. This scale is used in this paper for illustrative purposes because it is the most well-known one. Any finite ratio scale can be used in conjunction with the CDP concept and not only Saaty's scale.

A key assumption in this study is that, unless it is otherwise stated, we will have all possible comparisons (i.e., $n(n-1)$) and not the usual $n(n-1)/2$ number. Furthermore, each comparison is elicited independently of the rest. This is stated formally as follows:

The Comparison Completeness and Independence Assumption:

Given n entities to be evaluated by means of the pairwise comparison approach, all possible $n(n-1)$ comparisons will be elicited. Furthermore, the comparisons will be elicited independently of each other.

Given the above assumption, the following is a key issue that dictates a fundamental condition that reciprocal comparisons should satisfy. It is termed the *reciprocity condition (RC)* and it is best stated as follows:

The Reciprocity Condition:

When pairwise comparisons $a_{i,j}$, between any two items A_i and A_j of a collection of n items are elicited independently of each other, then under the **reciprocity condition (RC)** the following constraint should hold:

$$a_{i,j} = 1/a_{j,i}, \text{ for any } i, j = 1, 2, 3, \ldots, n.$$

In the empirical studies described in this paper it is assumed that the decision maker who elicits pairwise comparisons is an ultra-accurate one. We will call this the ultra-accurate decision maker (or UADM) assumption. This concept is introduced in order to demonstrate in multiple ways that results obtained from pairwise comparisons may exhibit certain ranking abnormalities, even under the ideal (and hence utopic) condition employed by the UADM assumption. Formally, the UADM assumption is described as follows:

The UADM Assumption:

> The **ultra-accurate decision maker (UADM) assumption** states that when a decision maker elicits pairwise comparisons and forms a pairwise comparison matrix, then this matrix will be the CDP matrix of the RCP matrix derived from the vector of the unknown weights of the entities which are of interest to the decision maker.

The UADM assumption is very optimistic but in **favor** of pairwise comparisons approaches. Nevertheless, the error results reported in this paper are dramatic in a **negative sense**. That is, despite the ultimately favorable (i.e., overoptimistic) assumption of having an UADM case, the results are still dramatic as being excessively negative as they may lead to too many ranking abnormalities.

Therefore, in a real-world setting, results based on the use of pairwise comparisons may exhibit even worse abnormalities than the ones derived under the UADM assumption. Thus, abnormalities exhibited by processing pairwise comparison matrices have, on the average, as a lower bound results exhibited by CDP matrices when they are derived under the overoptimistic UADM assumption. This is stated formally as the lower bound results when using the UADM assumption as follows:

Lower Bounds When Using the UADM Assumption:

> The real-world results based on the use of pairwise comparisons may exhibit even worse ranking abnormalities than the ones derived under the UADM assumption. That is, results under the UADM assumption can function, on the average, as **lower bounds** for results derived under a real-world setting.

As it is demonstrated later, ranking abnormalities derived by matrices produced under the UADM assumption can be easily quantified, while abnormality results under real-world conditions may not be as easily quantifiable. In summary, the UADM assumption is straightforward and creates a powerful evaluative tool for some key issues pertinent to any decision-making process which is based on pairwise comparisons.

4. Violations of the Reciprocity Condition (RC) When a Single Pairwise Comparison is Considered Under the UADM Assumption

This is best demonstrated by means of a simple numerical example as follows:

4.1. <u>Example #1</u>: A Case When a Single Pairwise Comparison Violates the Reciprocity Condition Under the UADM Assumption

To see the motivation for the first computational study, consider the following hypothetical example. Suppose that two items, say A_1 and A_2, are compared under a common aspect and the true value of that comparison is equal to 3.4673. That is, it is assumed that the value of the ratio of the true values of A_1 and A_2 is equal to 3.4673. An ultra-accurate decision maker (UADM) cannot assign this value to this comparison as this value is assumed to be unknown to the decision maker. Instead, it is assumed that this ultra-accurate decision maker uses a ratio scale, such as the Saaty scale, and assigns the assessment that is <u>closest to the previous numerical value</u> of 3.4673. That is, it is assumed that this ultra-accurate decision maker assigns the value 3 (i.e., *"moderate importance of one over the other"* according to the Saaty scale as shown in Table 1) because this value is closer to 3.4673 than any other value from Saaty's scale (including the value 4).

However, if instead of the ratio $a_{1,2}$, the decision maker had considered the reciprocal pairwise comparison, that is, $a_{2,1}$, then the value of the ratio of the true weights for $a_{2,1}$ is equal to 0.2887 (= 1/3.4673). Observe that the value of 0.2887 is closer to the value 1/4 (= 0.2500) than it is to the value 1/3 (= 0.3333) from Saaty's scale. That is, this ultra-accurate decision maker would select the value 1/4 from the Saaty scale. This means that item A_1 is assessed to be between being of *"moderate importance of one over the other"* and of *"essential or strong importance"* when it is compared to A_2 according to the Saaty scale in Table 1. Obviously, there is a contradiction with the case when the ratio $a_{1,2}$ was considered earlier in this hypothetical example under the UADM assumption. Therefore, a violation of the reciprocal condition is *guaranteed to occur if we assume the decision maker is an ultra-accurate (UADM) one as defined above.*

On the other hand, if the true value was equal to a quantity such as 3.1200, then this ultra-accurate decision maker would have assigned the value of 3 from Saaty's scale to the comparison $a_{1,2}$, and the value of 1/3 to the comparison $a_{2,1}$. This is true because the reciprocal of 3.1200 is equal to 0.3205 and the closest value from Saaty's scale to 0.3205 is the value 1/3. In this second scenario no violation of the reciprocal condition will occur. **End of Example**

The previous hypothetical example demonstrates that the Reciprocity Condition will be violated when the true value is in a certain range and the decision maker is assumed to be an ultra-accurate (UADM) one. Therefore, the question is how can one identify such intervals of values where a violation of the reciprocal condition always occurs under the UADM assumption. This is explored theoretically next.

4.2. Identifying the Ranges of Values Where the Reciprocity Condition is Violated Under the UADM Assumption

Suppose that the true value of a pairwise comparison of two entities, say A_1 and A_2, is between the two consecutive integers n and $n+1$, and it is equal to $n + X$, where $n + 1 \geq n + X \geq n$ (or, equivalently, $1 \geq X \geq 0$). Let $\omega(A_1)$ and $\omega(A_2)$ denote the true values of the corresponding values of these two entities. Then, from the previous consideration it is implied that $\omega(A_1) / \omega(A_2) = n + X$. This analysis is **independent** of Saaty's scale.

When the Saaty scale is considered, the values of n in the previous scenario are equal to 1, 2, 3, …, 8. Note that the highest value in Saaty's scale is equal to 9, but since we have $n + 1$, the maximum value for n must be equal to 8. An ultra-accurate decision maker will never be able to assign the value $n + X$ to this pairwise comparison $\omega(A_1) / \omega(A_2)$ (unless $X = 0$ or 1). Instead, it is assumed that this ultra-accurate decision maker will assign the value that is closest to the true value from the numerical values implied by the Saaty scale (or any ratio scale with a finite number of choices). That is, either the value n or $n + 1$ will be assigned to this comparison, depending on which of these two values (i.e., n or $n + 1$) is closest to the true value of $n + X$. It turns out that the value of X that makes the difference is $X =$ 0.50. That is, it is assumed that if the true value is less than or equal to $n + 0.50$, then the value this ultra-accurate decision maker will chose is equal to n. Otherwise, this decision maker will choose the value $n + 1$.

Next, we consider the case of having true values taken from the interval $[1/(n + 1), 1/n]$. Suppose that the true value is equal to $1/(n + X)$. The question is, what is the value of X that can make a difference? As before, we assume that the decision maker is ultra-accurate and he/she would assign the value $1/(n + 1)$ or $1/n$, depending on which one is closest to $1/(n + X)$. As before, this ultra-accurate decision maker does not know the value of $1/(n + X)$, but he/she can approximate it by using either the value $1/(n + 1)$ or $1/n$, (for the Saaty scale we have: $n = 1, 2, 3, \ldots, 8$).

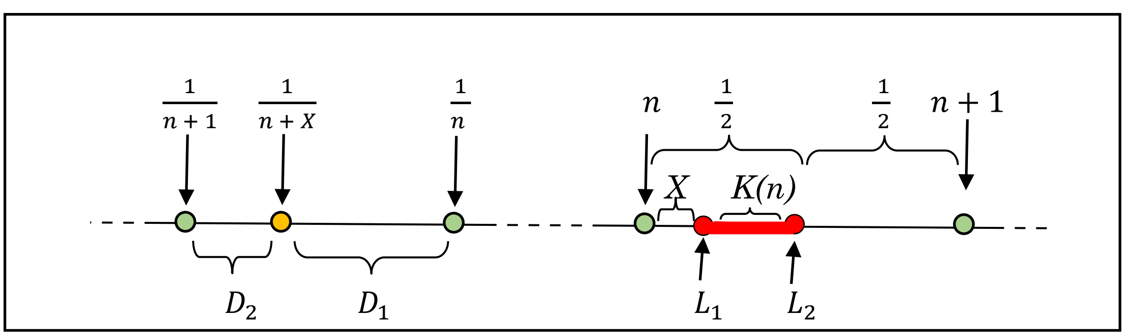

Figure 1: The main idea for computing the key differences D_1 and D_2. Note that $1 \geq X \geq 0$.

To help formulate this problem, we calculate the distance D_1 between $1/(n + X)$ and $1/n$ as follows (see also Figure 1):

$$D_1 = \frac{1}{n} - \frac{1}{n + X} \quad (1)$$

Similarly, we calculate the distance D_2 between $1/(n + X)$ and $1/(n + 1)$ as follows:

$$D_2 = \frac{1}{n + X} - \frac{1}{n + 1} \quad (2)$$

The point, that the decision maker would use as a <u>threshold value</u> to assign this pairwise comparison to either $1/(n + 1)$ or $1/n$ value from Saaty's scale, is when these two distances are equal to each other. That is, when $D_1 = D_2$. From expressions (1) and (2) this requirement yields the following relationships:

$$D_1 = D_2, \text{ or: } \frac{1}{n} - \frac{1}{n + X} = \frac{1}{n + X} - \frac{1}{n + 1},$$

from which after some simple algebraic manipulations, we get:

$$X = \frac{n}{2n + 1} \quad (3)$$

From the above result it follows that the difference between the quantities 1/2 and X, denoted as K(n), is as follows (see also Figure 1):

$$K(n) = \frac{1}{2} - X = \frac{1}{2} - \frac{n}{2n + 1},$$

or after some simple algebraic manipulations:

$$K(n) = \frac{1}{2(2n+1)}. \quad (4)$$

The last result indicates that the value of $K(n)$ will never be negative nor equal to 0. Moreover, the value of $K(n)$ approaches the value of 0 from positive values as the value of n approaches positive infinity. In practical terms this means the following. As the number n of discrete scale integer values increases, the range of values of true weights that violate the reciprocity condition under the UADM assumption, becomes monotonically smaller as n increases, but it never disappears. The previous considerations lead to the proof of Theorem 1, which is independent of the Saaty scale.

<u>Theorem 1:</u>

Given two consecutive integer numbers n and $n + 1$ (where $n \in N$), the interval of values between these two integers that violate the Reciprocity Condition is as follows: $[L_1, L_2]$, where:

$$L_1 = \frac{2n(n+1)}{2n+1}, \text{ and } L_2 = \frac{2n+1}{2}.$$

Furthermore, the length of this interval is equal to $K(n)$, which approaches zero from positive values as the value of n approaches infinity.

Proof:

Given relation (3) for the value of X, we have the following relations regarding the left border point (i.e., L_1) and the right border point (i.e., L_2) of the interval with the values that exhibit Reciprocity Condition violations (see also Figure 1):

$$L_1 = n + X = n + \frac{n}{2n+1} = \frac{2n^2+n+n}{2n+1} = \frac{2n(n+1)}{2n+1}, \text{ and}$$
$$L_2 = n + \frac{1}{2} = \frac{2n+1}{2}.$$

Furthermore, it can be easily shown with some simple algebraic manipulations that $L_2 - L_1 = \frac{1}{2(2n+1)} = K(n)$. This concludes the proof of Theorem 1. **End of Proof**

Table 2 has the value of X when $n = 1, 2, 3, \ldots, 8$, and it is based on formula (3). These results were also confirmed experimentally via a simulation study that scanned the interval [1/9, 9] for possible values and testing for RC violations. This scanning was done with a step size equal to 0.001. The same table also depicts the size of the sub-intervals defined as $K(n) = 0.5000 - X$. Note that in general, as the value of n increases, the value of X approaches 0.5000 and the corresponding interval approaches 0. This is derived by taking limits in Formula (3) when n approaches infinity. The size of these intervals (i.e., the values of $K(n)$) in Table 1 is depicted in Figure 2.

Any value between the reciprocals of the L_1 and L_2 borders, violates the Reciprocity Condition. For instance, when $n = 5$, then the value $a = 0.1825$ is between the values $1/L_1$ (= 0.1833 = 1/5.4546) and $1/L_2$ (= 0.1818 = 1/5.5000, see also Table 2). This value a is best approximated by 1/6 (= 0.1667) when Saaty's scale is used. However, its reciprocal is equal to 5.4795 (= 1/0.1825) which is best approximated by the value 5 and not 6, when the Saaty scale is used. Hence, a reciprocity condition violation occurs.

Table 2: The intervals of values in the range [1, 9] where the Reciprocity Condition (RC) is violated.

n	Value of X: $X = \frac{n}{2n+1}$	Size of interval: $K(n) = \frac{1}{2(2n+1)}$	Left border: $L_1 = n + X$	Right border: $L_2 = n + 1/2$
1	0.3333	0.1667	1.3333	1.5000
2	0.4000	0.1000	2.4000	2.5000
3	0.4286	0.0714	3.4286	3.5000
4	0.4444	0.0556	4.4444	4.5000
5	0.4546	0.0454	5.4546	5.5000
6	0.4615	0.0385	6.4615	6.5000
7	0.4667	0.0333	7.4667	7.5000
8	0.4706	0.0294	8.4706	8.5000

These results indicate the ranges of true values a pairwise comparison should have under the UADM assumption to have a violation of the reciprocity condition. Such case was the one explored in Example #1. If the true value of a pairwise comparison is within any of the above intervals (as defined in the two right-most columns in Table 2), then a Reciprocity Condition violation is guaranteed to occur under the UADM assumption.

Furthermore, if one adds up all the intervals in the third left-most column in Table 2 (i.e., the one under the column with the $K(n)$ values), it follows that the sum is equal to 0.5403. In a similar manner the sum of all the sub-intervals in the range [1/9, 1] can be shown to be equal to 0.1121817. These results indicate that when one considers the range of values of the Saaty scale, that is the range [1/9, 9], then the sum of all sub-intervals where Reciprocity Condition (RC) violations occur is 7.34% of that range (i.e., of the range [1/9, 9]). A simulation program empirically confirmed this result.

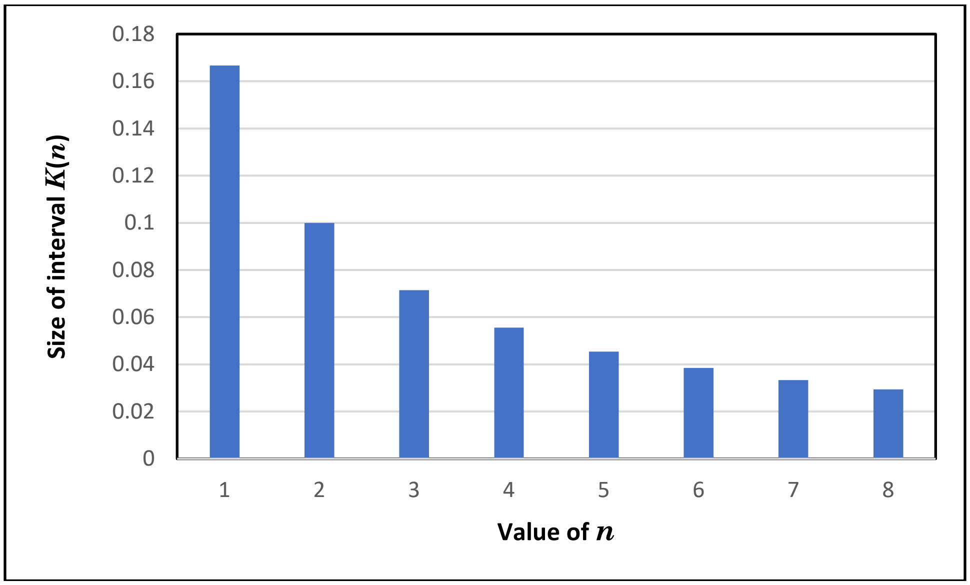

Figure 2: The sizes of the interval $K(n) = \frac{1}{2(2n+1)}$ where RC violations occur within the range [1, 9], or equivalently, when n = 1, 2, 3, …, 8.

5. Violations of the Reciprocity Condition When Pairwise Comparison Matrices are Elicited Under the UADM Assumption

This is best demonstrated by means of a simple numerical example as follows:

5.1. <u>Example #2:</u> A Case When a CDP Matrix Violates the Reciprocity Condition (RC) Under the UADM Assumption

The next question that is raised at this point is to examine the issue of Reciprocity Condition violation in the context of pairwise comparison matrices when the Saaty scale is used, and the decision maker is assumed to be an ultra-accurate one. To help fix ideas consider the following true values (or weights), normalized to add up to 1, of four entities A_1, A_2, A_3, and A_4:

$\omega(A_1) = 0.0939$

$\omega(A_2) = 0.4503$

$\omega(A_3) = 0.3252$
$\omega(A_4) = 0.1306$

Then, the corresponding RCP matrix is as follows:

$$\textbf{Matrix RCP} = \begin{bmatrix} 1 & 0.2086 & 0.2889 & 0.7194 \\ 4.7931 & 1 & 1.3847 & 3.4483 \\ 3.4616 & 0.7222 & 1 & 2.4903 \\ 1.3900 & 0.2900 & 0.4016 & 1 \end{bmatrix}$$

Note that any pairwise comparison derived by using these values is within the interval [1/9, 9], thus the Saaty scale is applicable. For instance, the comparison $a_{2,3}$ in position (2, 3) is equal to 1.3847 (= 0.4503 / 0.3252). Similarly for the rest of the entries of the previous RCP matrix. An ultra-accurate decision maker at the very best will assign the closest values from Saaty's scale to the ones in this RCP matrix. Recall that the Saaty values are equal to {1/9, 1/8, 1/7, …, 1, 2, 3, …, 7, 8, 9} (see also Table 1). When such ultra-accurate decision maker considers this case, it is assumed that the following CDP matrix will be derived:

$$\textbf{Matrix CDP} = \begin{bmatrix} 1 & \mathbf{0.2000} & 0.2500 & 0.5000 \\ \mathbf{5} & 1 & 1 & 3 \\ 3 & 0.5000 & 1 & 2 \\ 1 & 0.2500 & 0.3333 & 1 \end{bmatrix}$$

For instance, for the comparison $a_{2,3}$ in location (2, 3) the value 1 will be used by this ultra-accurate decision maker, because this is the one which is closest to the true (and hence unknown) value of 1.3847, as shown in the previous RCP matrix. Similarly for the rest of the entries in this CDP matrix. However, this matrix with values consistent with Saaty's scale exhibits 5 violations of the Reciprocity Condition. For example, note that the entry (2,3) in matrix CDP is not the reciprocal of the value of entry (3,2). Specifically, the following pairs of 5 symmetric entries exhibit violations of the Reciprocity Condition:

Entries (1, 3) = 0.2500	and (3, 1) = 3
Entries (1, 4) = 0.5000	and (4, 1) = 1
Entries (2, 3) = 1	and (3, 2) = 0.5000
Entries (2, 4) = 3	and (4, 2) = 0.2500
Entries (3, 4) = 2	and (4, 3) = 0.3333

The only case without such violation is the pair of symmetric entries (1, 2) and (2, 1), shown in boldface in the previous CDP matrix. **End of Example**

5.2. The Frequency the Reciprocity Condition (RC) is Violated in Random CDP Matrices of Various Sizes Under the UADM Assumption

The previous example motivates the creation of the following test, denoted as Test 1 in this paper.

Table 3.a: Part 1 of 2 of the summary of the results of the first simulation study. The number of random samples is 1,000,000 per dimension of the CDP matrices.

Dimension of CDP matrix	K1	K2	K3	K4	K1 + K4	Average 1	Ratio 1 (%)
2	94,039	280,813	592,466	32,682	126,721	0.13	74.21
3	205,130	347,921	318,323	128,626	333,756	0.39	61.46
4	370,779	294,345	153,610	181,266	552,045	0.78	67.17
5	552,492	213,758	56,284	177,466	729,958	1.30	75.69
6	708,879	131,199	17,169	142,753	851,632	1.97	83.24
7	821,950	70,392	4,500	103,158	925,108	2.76	88.85
8	894,263	34,973	1,054	69,710	963,973	3.68	92.77
9	939,183	15,856	179	44,782	983,965	4.75	95.45
10	964,737	6,991	38	28,234	992,971	5.95	97.16
11	979,740	3,023	4	17,233	996,973	7.27	98.27
12	988,342	1,268	0	10,390	998,732	8.73	98.96
13	993,622	518	1	5,859	999,481	10.33	99.41
14	996,368	199	0	3,433	999,801	12.06	99.66
15	998,041	81	0	1,878	999,919	13.92	99.81
16	998,969	28	0	1,003	999,972	15.91	99.90
17	999,405	5	0	590	999,995	18.04	99.94
18	999,717	2	0	281	999,998	20.30	99.97
19	999,857	2	0	141	999,998	22.70	99.99
20	999,925	1	0	74	999,999	25.24	99.99

Counter K1: Cases with RC violations and with ranking abnormalities.
Counter K2: Cases with NO RC violations and with ranking abnormalities.
Counter K3: Cases with NO RC violations and NO ranking abnormalities.
Counter K4: Cases with RC violations and NO ranking abnormalities.
Counter K1 + K4: All cases that exhibited at least one RC violation.

Average 1: The average number of RC violations per CDP matrix.

Ratio 1 (%): The percent of cases with ranking abnormalities of all cases with RC violations.
Ratio 1 = (K1 x 100) / (K1 + K4).

<u>Test 1:</u> (The Reciprocity Condition Test)

> If a pairwise comparison approach is used to estimate the relative weights and ranks of some entities under the UADM assumption, then the reciprocity condition (RC) should not be violated by any comparison pair of these entities.

Test 1 provides the impetus to consider a computational study to explore how often Test 1 is violated in random test problems, like the one described in Example #2. Thus, a simulation / empirical study was designed to explore how often Reciprocity Condition (RC) violations occur in random CDP matrices of various sizes. The goal was to explore how often closest discrete pairwise (CDP) comparison matrices exhibit at least one reciprocity violation. For this purpose, a computer program was written using the C# programming language (by using Microsoft's Visual Studio 2022). The program run on an XPS Dell desktop computer with a core i7 processor and the Windows 10 operating system.

First, vectors with random values in the interval [0, 1] of sizes 2 to 20 were generated. Only vectors that were compatible with Saaty's values were considered. That is, the maximum value of any pairwise comparison had to be less than or at most equal to 9, while the minimum value of any such pairwise comparison had to be greater than or at least equal to 1/9. Next, RCP matrixes with true values were considered and the corresponding CDP matrices were derived as in the previous numerical Example #2. After that, the CDP matrices were checked for violations of the Reciprocity Condition. Even if a single pairwise comparison was identified to violate the Reciprocity Condition (RC), this was recorded as a CDP matrix that violated this condition. The number of random replications (samples) was equal to 1,000,000 because with these many random replications the results were clearly statistically significant.

The results of this part of the first simulation study are tabulated in Table 3.a. These results are recorded in the columns with labels <K1>, <K4>, <K1 + K4>, and <Average 1>. The concept of "ranking abnormalities" is explained in the next section. The results under column <K1 + K4> represent all cases when a CDP matrix exhibited at least one violation of the RC condition. The size of the CDP matrices was equal to 2, 3, 4, …, 20.

It is the results under column <K1 + K4> that were used in Figure 3. The average number of RC violations per CDP matrix is recorded in Table 3.a under the column with the label <Average 1>.

The results in Figure 3 are dramatic. For CDP matrices of size 2 the frequency of having RC violations is greater than 10%, while for CDP matrices of size 3 the frequency jumps to more than 30%, for size 4 it jumps to more than 50% and for size 7 and above all the way to size 20 it is more than 90% while it is almost equal to 100% for sizes greater than 9. If all $n(n-1)$ possible pairwise comparisons are elicited (and not only $n(n-1)/2$ of them) **in real-world applications of a pairwise comparison-based approach (such as the AHP), then the number of Reciprocity Condition (RC) violations may be even more dramatic (on**

the average). This is stated because even under the highly favorable UADM assumption the number of such RC violations is dramatic.

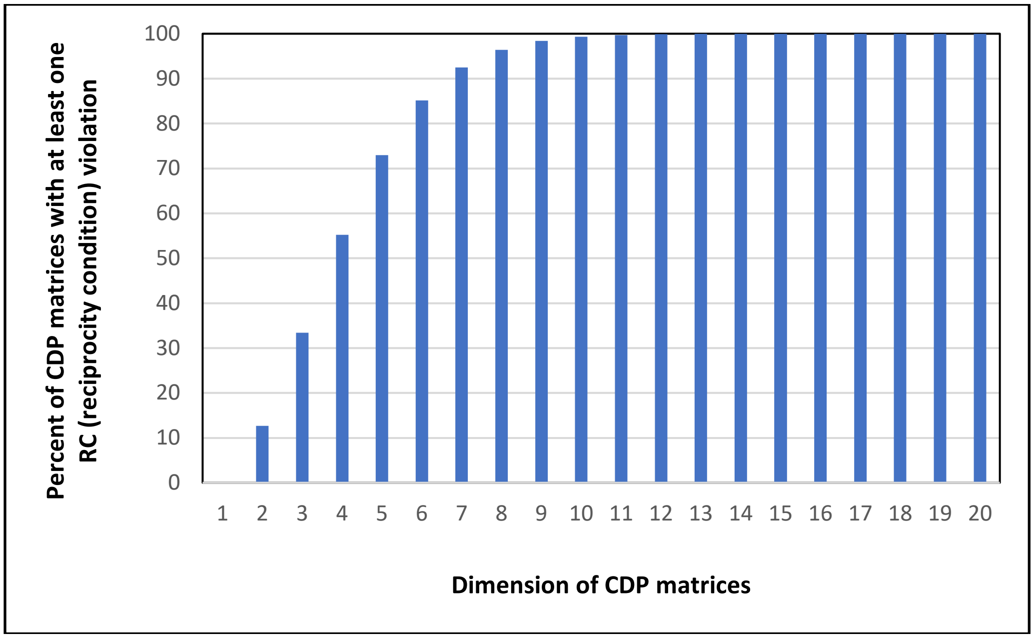

Figure 3: Percent of CDP matrices with at least one Reciprocity Condition (RC) violation. This figure is based on the sum of counters K1 and K4 (see also Table 3.a). The number of random samples per size of CDP matrix is equal to 1,000,000.

6. Some Ranking Abnormalities and Reciprocity Condition Violations When Random Pairwise Comparison Matrices are Elicited Under the UADM Assumption

This is best demonstrated by means of a simple numerical example as follows:

6.1. <u>Example #3:</u> A Case where a CDP Matrix with Reciprocal Condition (RC) Violations Leads to a Ranking Abnormality

Suppose that in some application the decision maker wishes to estimate the relative weights of four entities (say, four decision alternatives A_1, A_2, A_3, and A_4 of some decision problem). These entities are assumed to have the following relative true weights (normalized to add up to 1):

$\omega(A_1) = 0.1315$
$\omega(A_2) = 0.3078$
$\omega(A_3) = 0.2247$
$\omega(A_4) = 0.3361$

When these values are considered, the following RCP matrix with all the actual (true) pairwise comparisons is formed:

$$\textbf{Matrix RCP} = \begin{bmatrix} 1 & 0.4273 & 0.5854 & 0.3913 \\ 2.3405 & 1 & 1.3700 & 0.9159 \\ 1.7084 & 0.7299 & 1 & 0.6685 \\ 2.5555 & 1.0919 & 1.4959 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.1315 \\ 0.3078 \\ 0.2247 \\ 0.3361 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 4 \\ 2 \\ 3 \\ 1 \end{bmatrix} \end{array}$$

For instance, the entry (2, 3) of the above matrix is equal to 1.3700 because the ratio for entity A_2 over A_3 (i.e., $\omega(A_2) / \omega(A_3)$ is equal to 0.3078 / 0.2247 (= 1.3700). Similar is the case with the rest of the entries of the above matrix. In the previous description of the RCP matrix the vector with the real weights and corresponding ranks is shown above as well for easy reference.

As stated before, the above four values of the true weights and the entries of the previous pairwise comparison matrix are assumed to be unknown to the decision maker. Thus, the decision maker is using pairwise comparisons and the Saaty scale to best assess these relative weights. As before, under the assumption of having an ultra-accurate decision maker the following pairwise matrix (CDP matrix) with entries from Saaty's scale (i.e., the set of values {1/9, 1/8, 1/7, …, 1, 2, 3, …, 7, 8, 9}) is derived:

$$\textbf{Matrix CDP} = \begin{bmatrix} 1 & 0.5000 & 0.5000 & 0.3333 \\ 2 & 1 & 1 & 1 \\ 2 & 0.5000 & 1 & 0.5000 \\ 3 & 1 & 1 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.1384 \\ 0.3062 \\ 0.2165 \\ 0.3389 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 4 \\ 2 \\ 3 \\ 1 \end{bmatrix} \end{array}$$

For instance, the entry (2, 3) is equal to 1, because this is the closest value from Saaty's scale for the true entry 1.3700. Similarly with the rest of the entries of the above CDP matrix.

The above CDP matrix violates the reciprocal condition twice. This becomes evident if one considers the following two pairs of symmetric entries:

Entries (2, 3) = 1 and (3, 2) = 0.5000
Entries (3, 4) = 0.5000 and (4, 3) = 1

The standard Saaty approach dictates that the decision maker elicits only $n(n - 1)/2$ pairwise comparisons and not all the $n(n - 1)$ comparisons shown in the previous CDP matrix (where $n = 4$). For this example, this means the decision maker under the traditional approach would make $4(4 - 1)/2 = 6$ pairwise comparisons. Therefore, at this point two scenarios are considered.

Under the first scenario we consider the first diagonal submatrix of the CDP matrix. That is, the elements (i, j) where $j \geq i$, for $i = 1, 2, 3$, and 4. The rest of the comparisons are simply the reciprocals of the ones in the first diagonal submatrix. In this way, the resulted matrix <u>cannot</u> have any Reciprocity Condition violations. In a similar manner, under the second scenario, we consider the second diagonal submatrix. That is, the elements (i, j) where $j \leq i$, for i = 1, 2, 3, and 4. The rest of the comparisons are simply the reciprocals of the ones in the second diagonal submatrix. As before, there cannot be any Reciprocity Condition violations.

The first scenario results in matrix A, while the second scenario results in matrix B shown next. In this study these two matrices will be called "Saaty matrices" because they are produced by using only $n(n-1)/2$ pairwise comparisons and then filling the rest of the entries by using the reciprocals of the previous $n(n-1)/2$ comparisons.

$$\textbf{Matrix A} = \begin{bmatrix} 1 & 0.5000 & 0.5000 & 0.3333 \\ 2 & 1 & 1 & 1 \\ 2 & 1 & 1 & 0.5000 \\ 3 & 1 & 2 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.1252 \\ 0.2771 \\ 0.2330 \\ 0.3647 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 4 \\ 2 \\ 3 \\ 1 \end{bmatrix} \end{array}$$

$$\textbf{Matrix B} = \begin{bmatrix} 1 & 0.5000 & 0.5000 & 0.3333 \\ 2 & 1 & 2 & 1 \\ 2 & 0.5000 & 1 & 1 \\ 3 & 1 & 1 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.1259 \\ 0.3314 \\ 0.2343 \\ 0.3084 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 4 \\ 1 \\ 3 \\ 2 \end{bmatrix} \end{array}$$

Besides these two new matrices, the corresponding weights implied by them are depicted along with the corresponding rankings. The vectors with the relative weights were derived by employing a standard dominant eigenvector approach according to which the geometric means of each row are computed and then normalized so the weights would add up to 1 (Aczél and Saaty, 1983), (Forman and Peniwati, 1998).

An examination of these results reveals that when matrix B is considered, the ranking is different than the one under matrix A, which happens to be the same as the one under the RCP matrix (i.e., the real ranking which is unknown to the decision maker). As matter of fact, under the second scenario entity A_4 is ranked as having rank 2, while entity A_2 as having rank 1 (i.e., the top rank). This occurs while in reality entity A_4 has a real weight equal to 0.3361 and the real weight of A_2 is equal to 0.3078. That is, although the true weight of entity A_4 **is 9.2% higher** than the one for entity A_2, entity A_4 looks as being inferior to A_2 under matrix B. This is clearly a ranking abnormality (rank inversion).

Interestingly, if one considers the vectors of relative weights and compares them in terms of the real weights (under the RCP matrix) then the weights under matrix A have a sum of squared differences equal to 0.00186903, while when the weights under matrix B are considered then the sum of their squared differences is equal to 0.00144777 (i.e., it is smaller). But it is the comparison of the ranks that reveals the ranking abnormality described above. **End of Example**

6.2. A Simulation Study to Explore How Often Random Pairwise Comparison Matrices Exhibit Reciprocity Condition Violations and Ranking Abnormalities

Example #3 is the incentive to establish Test 2. This test first considers an RCP matrix and then the corresponding CDP matrix. Next, from the CDP matrix it considers the upper and lower triangle submatrices. Given a triangle submatrix, it forms a complete Saaty matrix by building the rest of the matrix by using the reciprocal relationship $a_{i,j} = 1/a_{j,i}$, for any i, j = 1,2,3,…, n. If the CDP matrix has exhibited at least one RC violation, then these two Saaty matrices will be non-identical. Otherwise, they will be identical.

If the two Saaty matrices are different, then they may or may not yield the same ranking of all the alternatives. If they are different and the two rankings derived from them are different too, then at least one of them has a ranking that is different than the hidden (and thus "true" in the sense of Example #3) ranking. In such case, the two rankings may differ in terms of the top alternative (Case A-1 in Figure 4), or they may agree in terms of the top alternative but disagree with the ranking of at least another alternative(s) (Case A-2). In this study such cases are called **ranking abnormalities**. If the rankings of the two Saaty matrices agree with each other, their common ranking is compared to the hidden (i.e., the "true") ranking of the initial data. As before, two cases of ranking abnormalities can be defined. The common ranking and the "true" ranking disagree in terms of the top alternative (Case B-1), or they agree in terms of the top alternative, but they disagree in terms of the ranking of at least another alternative(s) (Case B-2). The previous four cases of ranking abnormalities are presented conceptually in Figure 4.

One may define more cases of ranking abnormalities. For instance, another possibility is to disagree on the top alternative and one more alternative, or to disagree on all the alternatives or even to be the opposite ranking of each other. Ranking difference may also be defined as the sum of the absolute differences between pairs of ranks.

A more comprehensive study on how to evaluate differences in rankings is provided in (Ray and Triantaphyllou, 1998; and 1999). Given the above considerations, Test 2 is defined as follows:

<u>Test 2:</u> (Ranking Abnormalities when the UADM assumption is considered)

Let a pairwise comparison approach be used on some "hidden weights" via an RCP / CDP transformation process under the UADM assumption. Next, the derived CDP matrix is used to derive two traditional Saaty matrices by considering the upper and lower triangular submatrices of it. The rankings derived by these two Saaty matrices should not exhibit any ranking abnormalities.

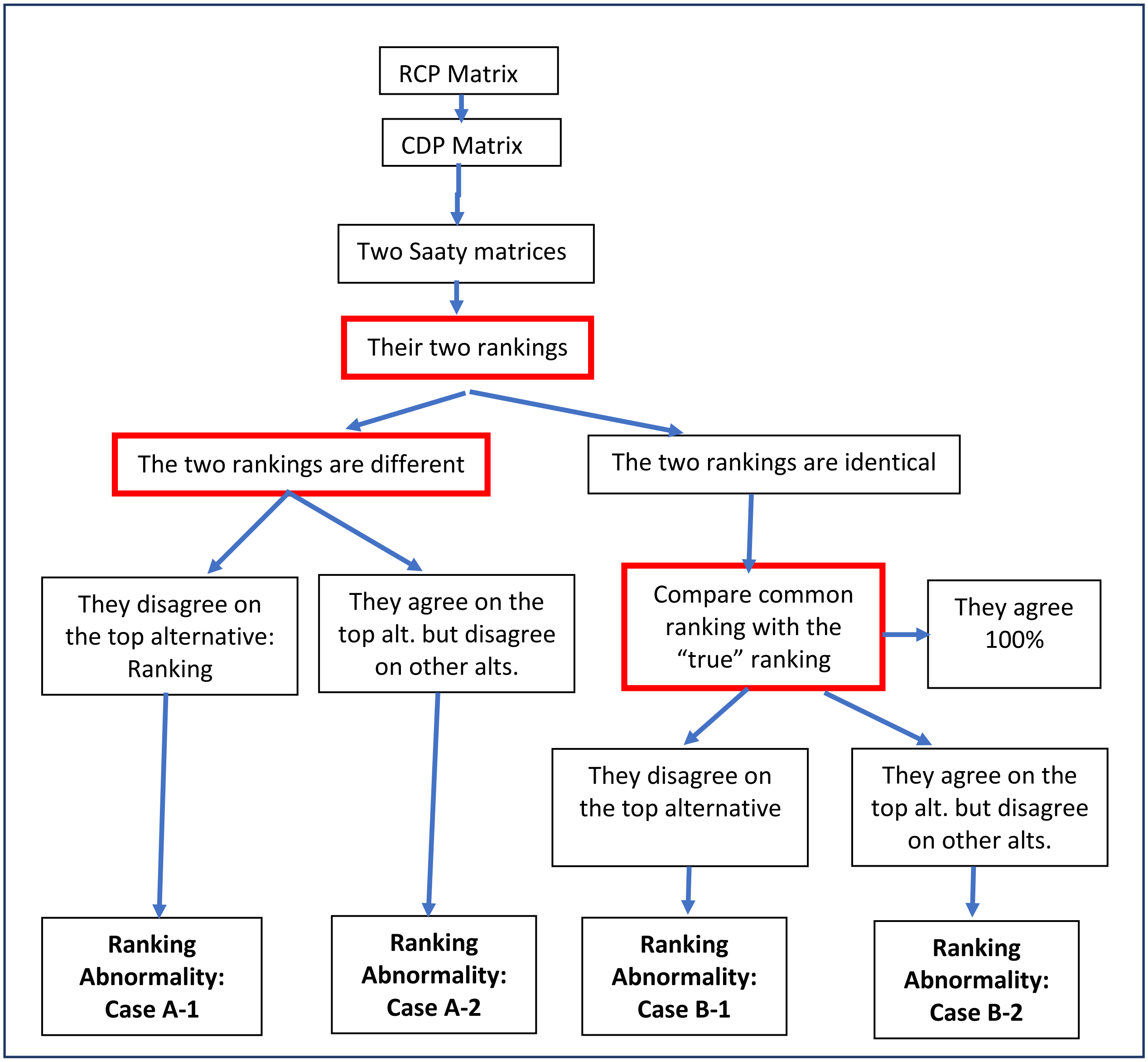

Figure 4: The four cases of ranking abnormalities considered in the first simulation study.

The previous simulation study was extended in order to explore how often random CDP matrices may have Reciprocity Condition violations and ranking abnormalities as described above.

As before a computer program was written in C# and run on an XPS Dell desktop computer with a core i7 processor and the Windows 10 operating system. The computational results are summarized in Tables 3.a and 3.b, and are depicted graphically in Figures 5, 6, 7, and 8.

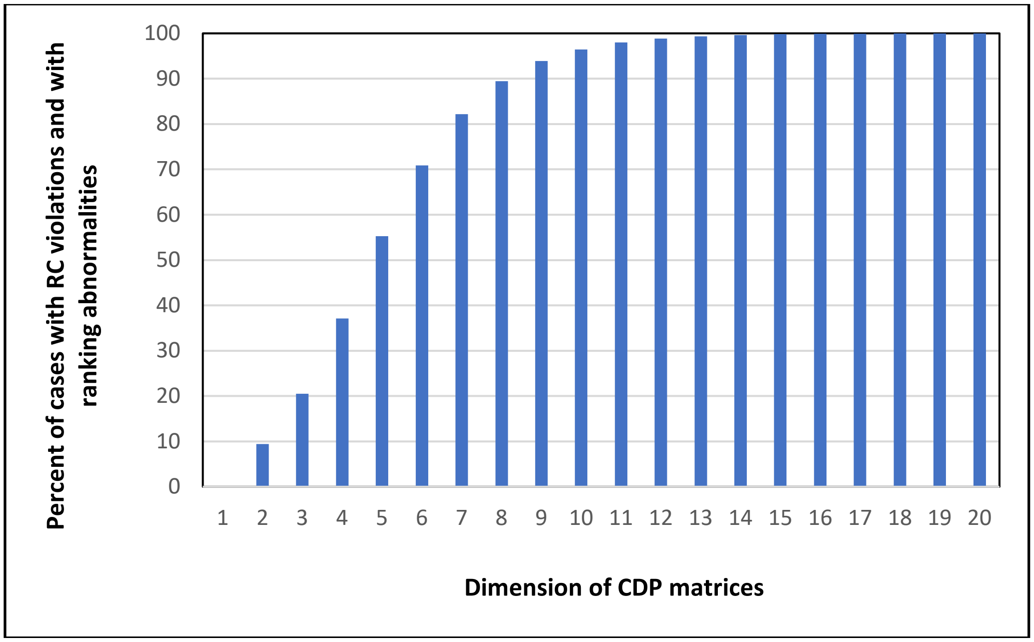

Figure 5: Percent of CDP matrices **with** Reciprocity Condition (RC) violations and **with** ranking abnormalities for CDP matrices of various sizes. This figure is based on counter K1 (see also Table 3.a). The number of random samples is equal to 1,000,000.

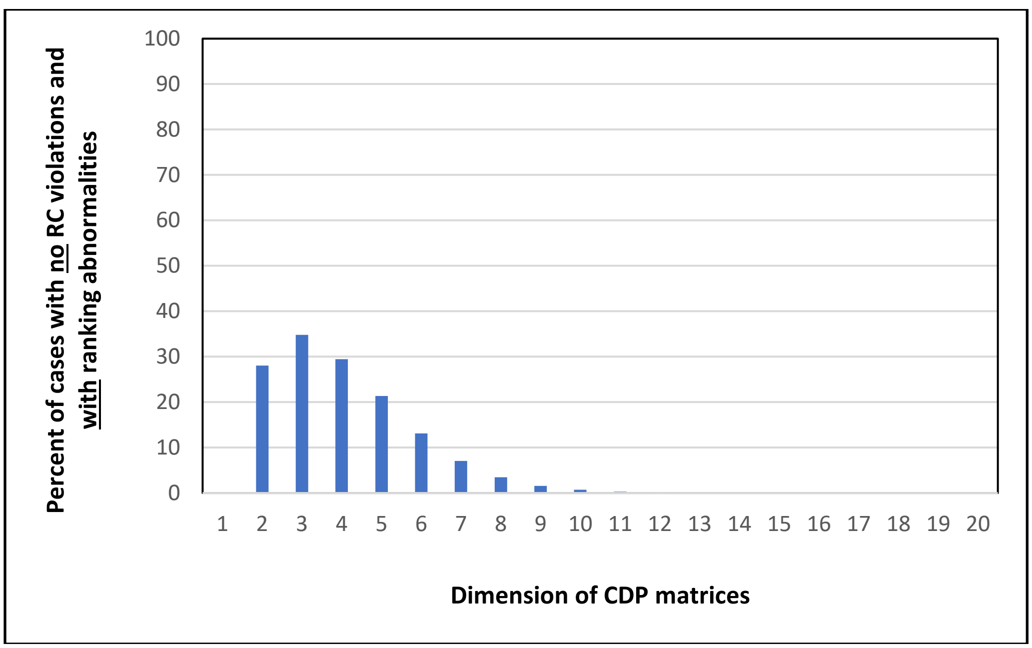

Figure 6: Percent of CDP matrices with **no** Reciprocity Condition (RC) violations and **with** ranking abnormalities for CDP matrices of various sizes. This figure is based on counter K2 (see also Table 3.a). The number of random samples is equal to 1,000,000.

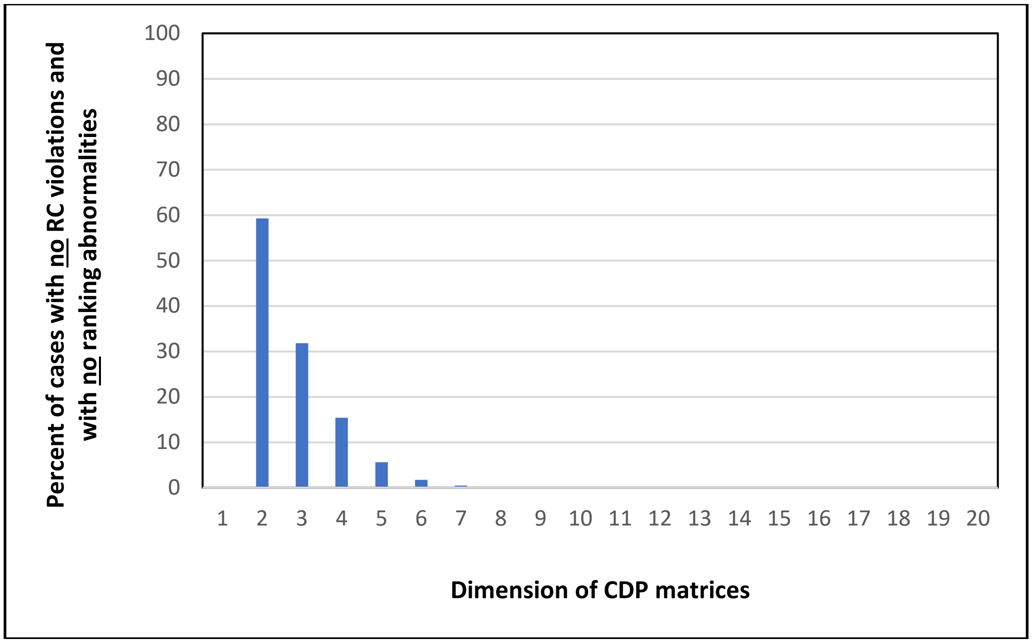

Figure 7: Percent of CDP matrices with **<u>no</u>** Reciprocity Condition (RC) violations and with **<u>no</u>** ranking abnormalities for CDP matrices of various sizes. This figure is based on counter K3 (see also Table 3.a). The number of random samples is equal to 1,000,000.

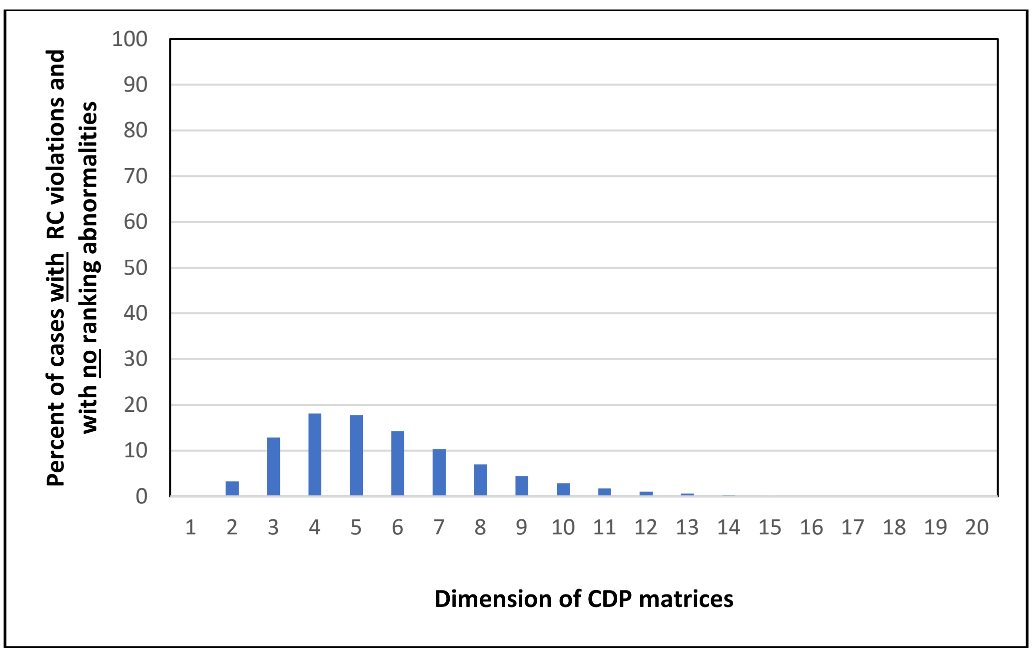

Figure 8: Percent of CDP matrices **<u>with</u>** Reciprocity Condition (RC) violations and with **<u>no</u>** ranking abnormalities for CDP matrices of various sizes. This figure is based on counter K4 (see also Table 3.a). The number of random samples is equal to 1,000,000.

Table 3.b: Part 2 of 2 of the summary of the results of the first simulation study. The number of random samples is 1,000,000 per dimension of the CDP matrices.

Dimension of CDP matrix	C1	C2	C3	C4	C5
2	374,852	94,039	0	94,039	280,813
3	553,051	204,206	4,239	199,967	348,845
4	665,124	335,749	12,134	323,615	329,375
5	766,250	469,947	22,890	447,057	296,303
6	840,078	589,825	36,490	553,335	250,253
7	892,342	690,781	51,224	639,557	201,561
8	929,236	772,391	68,769	703,622	156,845
9	955,039	837,818	86,837	750,981	117,221
10	971,728	886,734	105,730	781,004	84,994
11	982,763	922,625	125,560	797,065	60,138
12	989,610	948,165	145,639	802,526	41,445
13	994,140	966,587	166,097	800,490	27,553
14	996,567	978,931	186,310	792,621	17,636
15	998,122	987,144	206,596	780,548	10,978
16	998,997	992,184	226,623	765,561	6,813
17	999,410	995,372	246,765	748,607	4,038
18	999,719	997,269	266,532	730,737	2,450
19	999,859	998,527	285,791	712,736	1,332
20	999,926	999,104	304,869	694,235	822

Counter C1: All ranking abnormalities.
Counter C2: The number of cases the two Saaty matrices derived contradicting rankings.
Counter C3: The number of cases the rankings of the two Saaty matrices disagreed on the top alternative (i.e., Case A-1 in Figure 4).
Counter C4: The number of cases the rankings of the two Saaty matrices agreed on the top alternative but disagreed on the ranking of other alternative(s) (i.e., Case A-2 in Figure 4).
Counter C5: The number of cases the rankings of the two Saaty matrices agreed 100% with each other but their common ranking disagreed with the way some non-top alternatives were ranked when the original (true) data are considered (i.e., Case B-2 in Figure 4).
Note that the case the two Saaty matrices yielded identical rankings and the top ranking is different than the true top ranking is not shown because this case (i.e., Case B-1 in Figure 4) never happened.

Note that the following relationships follow from Tables 3.a and 3.b:
C1 = K1 + K2; This is true by definition.
C1 = C2 + C5; This is true by definition.
C2 = C3 + C4; This is true by definition.
C2 ≤ K1 + K4; This is true because RC violations may exist but the two Saaty matrices may still yield identical rankings.

Figures 5 to 8 indicate that when reciprocity condition (RC) violations occur, then it is very likely to have ranking abnormalities as well. This is also indicated by the high values of Ratio 1 in Table 3.a (in the right-most column). However, if no reciprocity violations occur, then the likelihood to have a ranking abnormality is (relatively speaking) much smaller. Furthermore, the likelihood to have both RC violations and ranking abnormalities at the same time is very high (see also Figure 5). The results become even more dramatic if one considers that these findings were derived under the highly optimistic UADM (ultra-accurate decision maker) assumption.

7. Abnormal Rank Groupings that May Occur When Consecutive Pairwise Comparison Matrices are Considered

The motivation for this type of exploration is simple. One first considers a vector with some random weights. Next the RCP and CDP matrices are formed as before. It is assumed that this CDP matrix can capture the data needed (in terms of pairwise comparisons) to accurately estimate the relative weights and eventually the ranking of the entities of interest. The weights derived from the first CDP matrix are used to derive an RCP matrix. From the second RCP matrix a new CDP matrix is derived and so on. If all the CDP matrices were accurate, then the consecutive CDP matrices and derived weights (and rankings) must be identical and not change. However, as it is demonstrated next, this may not be the case. The above is the essence of Test 3.

<u>Test 3:</u> (The Ranking Preservation Test)

> Let a pairwise comparison approach be used and a set of weights and corresponding ranking be inferred. Next, these weights are used as the "hidden data" to infer a new set of weights and ranking via the RCP / CDP transformation process under the UADM assumption. Suppose that this process iterates until two consecutive iterations reach identical results in terms of the inferred weights. If the pairwise comparison approach is effective, then the original ranking and the rankings inferred at the last iteration should be identical.

Note that Test 3, and Tests 1 and 2, **are necessary but not sufficient**. For Test 3 this means that if a pairwise comparison process is reliable, then the ranking preservation condition must be satisfied. However, if the condition of Test 3 is satisfied, this does not guarantee that the decision-making process is reliable.

7.1 <u>Example #4:</u> An Abnormal Rank Grouping When a Sequence of CDP Matrices is Considered

The data for this example are taken from a case study first reported by Saaty when he had introduced the AHP method for the first time (Saaty, 1977). This is the example when using a pairwise comparison matrix to estimate the weights of five simple daily objects. According to this case study we have five objects with some weights (which are unknown to the decision maker). Following are their normalized true weights (as given in (Saaty, 1977)):

Item Name	Relative Weight	Rank
Object 1	0.10	4
Object 2	0.39	1 (= top)
Object 3	0.20	3
Object 4	0.27	2
Object 5	0.04	5
Total =	**1.00**	

Iteration 1:

When the previous five weights are used, the following RCP matrix, termed matrix RCP1, is derived. This matrix has the true values and thus it is assumed to be hidden from the decision maker.

$$\textbf{Matrix RCP1} = \begin{bmatrix} 1 & 0.2564 & 0.5000 & 0.3704 & 2.5000 \\ 3.9000 & 1 & 1.9500 & 1.4444 & 9.7500 \\ 2 & 0.5128 & 1 & 0.7407 & 5 \\ 2.7000 & 0.6923 & 1.3500 & 1 & 6.7500 \\ 0.4000 & 0.1026 & 0.2000 & 0.1482 & 1 \end{bmatrix}$$

The corresponding CDP Matrix (termed matrix CDP1) is derived as in previous examples, and it is as follows:

$$\textbf{Matrix CDP1} = \begin{bmatrix} 1 & 0.2500 & 0.5000 & 0.3333 & 2 \\ 4 & 1 & 2 & \mathbf{1} & 9 \\ 2 & 0.5000 & 1 & 0.5000 & 5 \\ 3 & \mathbf{0.5000} & 1 & 1 & 7 \\ 0.5000 & 0.1111 & 0.2000 & 0.1429 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.1011 \\ 0.3910 \\ 0.1997 \\ 0.2660 \\ 0.0422 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 4 \\ 1 \\ 3 \\ 2 \\ 5 \end{bmatrix} \end{array}$$

It can be observed that a Reciprocity Condition violation occurs when one considers the following pair of entries (shown in boldface, above):

Entry (2, 4) = 1, and entry (4, 2) = 0.5000

This CDP matrix yields the relative weights and corresponding ranking shown above. The average squared difference of the above weight vector when it is compared to the true (and thus unknown to the decision maker) vector is equal to:

$$[(0.10 - 0.1011)^2 + (0.39 - 0.3910)^2 + (0.20 - 0.1997)^2 + (0.27 - 0.2660)^2 + \\ + (0.04 - 0.0422)^2] / 5 = 0.00006846.$$

Iteration 2:

The weights derived from matrix CDP1, are used to build a new RCP matrix (termed RCP2) which is as follows:

$$\textbf{Matrix RCP2} = \begin{bmatrix} 1 & 0.2585 & 0.5063 & 0.3801 & 2.3957 \\ 3.8675 & 1 & 1.9579 & 1.4699 & 9.2654 \\ 1.9753 & 0.5107 & 1 & 0.7508 & 4.7322 \\ 2.6311 & 0.6803 & 1.3320 & 1 & 6.3033 \\ 0.4174 & 0.1079 & 0.2113 & 0.1587 & 1 \end{bmatrix}$$

The corresponding CDP Matrix (termed CDP2) is as follows:

$$\textbf{Matrix CDP2} = \begin{bmatrix} 1 & 0.2500 & 0.5000 & 0.3333 & 2 \\ 4 & 1 & 2 & \mathbf{1} & 9 \\ 2 & 0.5000 & 1 & 1 & 5 \\ 3 & \mathbf{0.5000} & 1 & 1 & 6 \\ 0.5000 & 0.1111 & 0.2000 & 0.1667 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.0985 \\ 0.3808 \\ 0.2234 \\ 0.2513 \\ 0.0460 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 4 \\ 1 \\ 3 \\ 2 \\ 5 \end{bmatrix} \end{array}$$

This matrix too exhibits a case of Reciprocity Condition violation when the same pair of entries is considered as before. This matrix yields the relative weights and corresponding ranking shown above. The average squared difference of the above weight vector when it is compared to the true (and thus unknown to the decision maker) vector is equal to:

$[(0.10 – 0.0985)^2 + (0.39 – 0.3808)^2 + (0.20 – 0.2234)^2 + (0.27 – 0.2513)^2 +$
$+ (0.04 – 0.0460)^2] / 5 = 0.000204028.$

Iteration 3:

The weights derived from matrix CDP2 are used to build a new RCP matrix (termed RCP3; not shown next for simplicity of the presentation) which yields the following CDP matrix:

$$\textbf{Matrix CDP3} = \begin{bmatrix} 1 & 0.2500 & 0.5000 & 0.3333 & 2 \\ 4 & 1 & 2 & 2 & 8 \\ 2 & 0.5000 & 1 & 1 & 5 \\ 3 & 0.5000 & 1 & 1 & \mathbf{5} \\ 0.5000 & 0.1250 & 0.2000 & \mathbf{0.1667} & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.0949 \\ 0.4114 \\ 0.2151 \\ 0.2333 \\ 0.0454 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 4 \\ 1 \\ 3 \\ 2 \\ 5 \end{bmatrix} \end{array}$$

For example, entry (2, 5) of matrix RCP3 (not shown) is equal to 0.3808 / 0.0460 = 8.2783, which results entry (2, 5) in matrix CDP3 to be equal to 8. Similarly for entry (4, 5) = 0.2513 / 0.0460 = 5.4630 → 5 for entry (4, 5) in matrix CDP3 and so on.

Note that entry (5, 4) = 0.0460 / 0.2513 = 0.1831 is marginally closer to 0.1667 which is entry (5, 4) in matrix CDP3. This causes a violation of the Reciprocity Condition as the entry (4, 5) = 5 $\neq$ 6 = 1/0.1667. This matrix yields the relative weights and corresponding ranking shown above.

The average squared difference of the above weight vector when it is compared to the true (and thus unknown to the decision maker) vector is equal to:

$[(0.10 – 0.0949)^2 + (0.39 – 0.4114)^2 + (0.20 – 0.2151)^2 + (0.27 – 0.2333)^2 +$
$+ (0.04 – 0.0454)^2] / 5 = 0.000417606.$

Iteration 4:

As before, the last weights are used to build a new RCP matrix (termed RCP4; not shown for simplicity) which yields the following CDP matrix (termed matrix CDP4):

$$\textbf{Matrix CDP4} = \begin{bmatrix} 1 & 0.2500 & 0.5000 & 0.3333 & 2 \\ 4 & 1 & 2 & 2 & 9 \\ 2 & 0.5000 & 1 & 1 & 5 \\ 2 & 0.5000 & 1 & 1 & 5 \\ 0.5000 & 0.1111 & 0.2000 & 0.2000 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.0956 \\ 0.4245 \\ 0.2168 \\ 0.2168 \\ 0.0463 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 3 \\ 1 \\ 2 \\ 2 \\ 4 \end{bmatrix} \end{array}$$

For instance, entry (2, 5) = 0.4114 / 0.0454 = 9.0617 → 9. This matrix yields the relative weights and the corresponding ranking shown above. The average squared difference of the above weight vector when it is compared to the true (and thus unknown to the decision maker) vector is equal to:

$$[(0.10 - 0.0956)^2 + (0.39 - 0.4245)^2 + (0.20 - 0.2168)^2 + (0.27 - 0.2168)^2 + (0.04 - 0.0463)^2] / 5 = 0.000872356.$$

It is also noticeable that now two objects that in reality are distinct (i.e., objects 3 and 4) appear to be of the same size (i.e., equal to 0.2168).

Iteration 5:

As before, the last weights are used to build a new RCP matrix (termed RCP5; not shown for simplicity) which yields the following CDP matrix (termed matrix CDP5):

$$\textbf{Matrix CDP5} = \begin{bmatrix} 1 & 0.2500 & 0.5000 & 0.5000 & 2 \\ 4 & 1 & 2 & 2 & 9 \\ 2 & 0.5000 & 1 & 1 & 5 \\ 2 & 0.5000 & 1 & 1 & 5 \\ 0.5000 & 0.1111 & 0.2000 & 0.2000 & 1 \end{bmatrix} \quad \begin{array}{c} \underline{\textbf{Weights}} \\ \begin{bmatrix} 0.1028 \\ 0.4211 \\ 0.2151 \\ 0.2151 \\ 0.0459 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{\textbf{Rank}} \\ \begin{bmatrix} 3 \\ 1 \\ 2 \\ 2 \\ 4 \end{bmatrix} \end{array}$$

The average squared difference of the above weight vector when it is compared to the true (and thus unknown to the decision maker) vector is equal to:

$$[(0.10 - 0.1028)^2 + (0.39 - 0.4211)^2 + (0.20 - 0.2151)^2 + (0.27 - 0.2151)^2 + (0.04 - 0.0459)^2] / 5 = 0.000850376.$$

As before, objects 3 and 4 appear to be of the same size (i.e., but now they are equal to 0.2151).

When a new RCP and the corresponding CDP matrix are formed, they are identical to the ones derived in iteration 5. That is, the results of iteration 6 and beyond (not shown) are identical to those of iteration 5. Thus, this testing process terminates here.

The ranking at the last iteration is different than the initial ranking. The difference between these two rankings can be defined as the sum of the absolute differences of the individual ranks. For the current example this is equal to 3 because |4−3| + |1−1| + |3−2| + |2−2| + |5−4| = 3. This score can be normalized by dividing it by the number of items (i.e., by 5). Thus, the normalized difference of the two rankings is equal to 0.60 (= 3/5). It is also observed that the two rankings agree in terms of the top alternative (marked by the rank equal to 1).

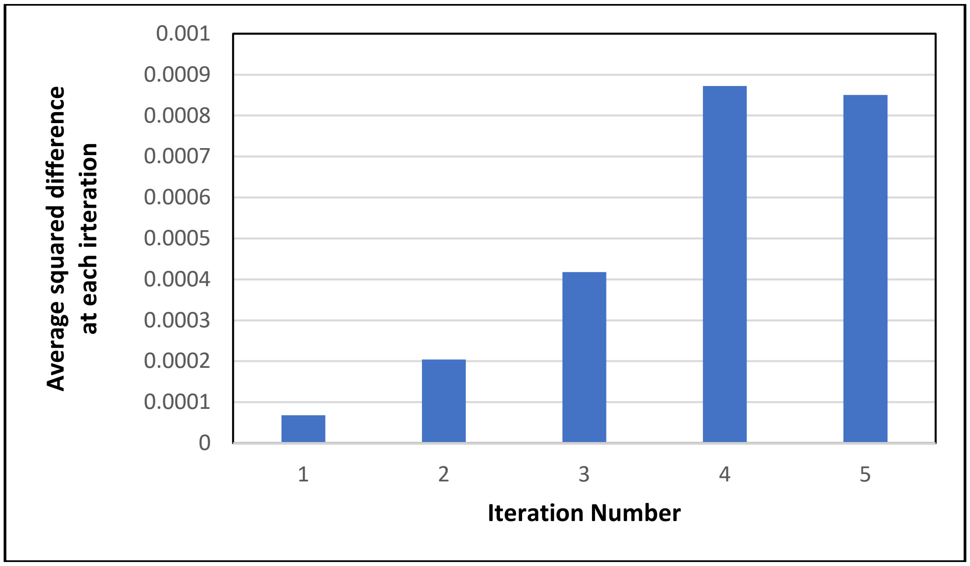

Figure 9: The average of the squared differences at each iteration of the process described in Example #4.

It is noticeable that when this testing process converges, entities (objects) 3 and 4 are of the same rank (i.e., 2; see the ranks at iteration 4 or 5). However, when the initial weights are considered, then it turns out that in reality object 4 has a weight that is 35% higher than that of object 3. Nevertheless, at the end of iteration 4 or 5 both objects are shown to be of the same weight despite the 35% difference of their true weights. That is, this major difference between these two objects has been lost during the testing process.

Figure 9 depicts the averages (i.e., normalized by dividing by 5) of the squared differences between the initial (and thus true) set of weights and the weights derived at each iteration. The errors at each iteration are not cumulative as inconsistencies can work either way and sometimes they may cancel out, even partially. This is indicated in Figure 9 where the average squared difference is (slightly) lower at iteration #5 than the value at iteration #4. However, the main trend is that the average squared differences, on the average, keep increasing.

It is clear, that each time a CDP matrix is formed, the difference grows dramatically (i.e., up to 1,264 %). In conclusion, with the derivation of a new CDP matrix, something is being lost and this loss can accumulate very rapidly. This is to be expected as a CDP matrix is in essence a discretized approximation of an RCP matrix. What is not very obvious, however, is the dramatic rate of the loss.

The next step of this investigation was to repeat this exploration but use some other datasets (vectors with the initial weights) as the starting point. Such data came from some real-life case studies reported in papers published by Saaty (i.e., (Saaty, 2008; and 1977)). The results of these cases are summarized in Table 4. These results show that out of 6 case studies, 4 (i.e., the majority) had this type of ranking problems. This compelled us to perform a more extensive simulation-based study as it is explained in the next section.

Table 4: Summary of the Abnormal Rank Groupings that Occur When Consecutive Pairwise Comparison Matrices are Considered with Data from some Case Studies published in the Literature by Saaty.

Case Study No	Size N	Publication	Domain Information	C1	C2	C3	C4
#1	7	(Saaty, 2008)	Comparing protein contents of 7 foods (a food item with 0 protein content was replaced by another one with a very small protein content)	2	5	1	**Yes**
#2	4	(Saaty, 1977)	Estimating distances from 4 objects based on visual input	1	2	0	**No**
#3	6	(Saaty, 1977)	Relative distances between the city of Philadelphia, PA, and 6 other cities in the world	1	4	1	**Yes**
#4	7	(Saaty, 1977)	Relative GDP of 7 countries in year 1977	3	7	2	**Yes**
#5	5	(Saaty, 1977)	Relative weights of 5 common daily life objects **(same as Example #4)**	3	5	1	**Yes**
#6	5	(Saaty, 2008)	Relative areas (surfaces) of 5 simple geometric figures	1	2	0	**No**

Notes:

C1 = Number of RC (Reciprocity Constraint) violations.
C2 = Number of iterations until convergence of the weights vector was achieved.
C3 = Number of times two consecutive rankings disagreed with each other.
C4 = The final ranking was different than the original (true / hidden) ranking.

7.2. The Frequency Abnormal Rank Groupings Occur When Consecutive Pairwise Comparison Matrices are Considered Under the UADM Assumption

A simulation / empirical study was conducted to explore how frequently this type of ranking abnormalities may occur with random data. The previous suite of C# computer programs was expanded to serve the needs of this study. Computational tests on simulated problems similar to Example #4, run for CDP matrices of size $n = 2, 3, 4, \ldots, 20$.

The results of this study are depicted in Table 5. These results reveal that on the average a rank grouping occurs after 1-7 iterations. More specifically, this rank grouping occurred on the average after (approximately, see also Table 5) 6.74 iterations when $n = 20$, 5.01 iterations when $n = 10$, and 2.32 iterations when $n = 4$.

Table 5: Summary of the simulation runs regarding Test 3. A random CDP matrix was generated and then a sequence of CDP matrices was derived from it and tested consecutively according to Test 3. The number of random samples is 1,000,000 per dimension of the CDP matrices.

Dimension of CDP matrix	**Average no. of iterations before convergence**	**Percentage of times the final ranking was different than the original (true / hidden) ranking**	**Average difference* between the final and the original (true / hidden) rankings**
2	1.0000	28.2011	0.1410
3	1.4547	55.5195	0.3642
4	2.3240	85.7079	0.7611
5	2.5056	92.2367	1.0773
6	3.4883	97.6851	1.5583
7	3.7700	99.2203	1.9480
8	4.4019	99.6943	2.4684
9	4.4958	99.9177	2.8861
10	5.0072	99.9742	3.4128
11	5.1566	99.9925	3.8572
12	5.5383	99.9980	4.3837
13	5.6331	99.9993	4.8475
14	5.9567	99.9994	5.3703
15	6.1379	100	5.8691
16	6.2407	100	6.3558
17	6.3972	100	6.8544
18	6.5162	100	7.3432
19	6.6613	100	7.8453
20	6.7370	100	8.3328

*** Note:**
The average difference of the rankings was computed as the sum of the absolute values of the individual rank differences divided by the number of items (i.e., the dimension of the CDP matrix). As the dimension increases, such differences increase as the range of values naturally increases from 1 to *n* (i.e., the size of the CDP matrices).

As was the case before, this type of ranking abnormality happens dramatically often even for CDP matrices of small sizes. It reaches almost a frequency of more than 90% with sizes of 5 and higher (see also the third column from the left of Table 5), and more than 99% for sizes 7 and higher. When one realizes that such dramatic results were obtained under the ultra-accurate decision maker (UADM) and thus very optimistic assumption, then the conclusion is that in real-world applications of the pairwise comparisons approach the situation, on the average, may be even more dramatic (see also the assumptions stated in Section 3).

8. Discussion and Conclusions

If one assumes the possibility of using a finer granularity in the available decision choices than what the Saaty scale suggests, then the error rates reported in this paper will definitely improve (on the average). The higher the granularity of the choices is, the better the computational results would be when the proposed RCP/CDP-based testing approaches are used. However, there is a crucial challenge: Can decision makers choose among decision choices of higher and higher granularity? The higher the granularity is, the better it is computationally in terms of the tests run in this study. But how about being **practical in the real-world**? Where is the limit on granularity?

To gain a perspective of how granularity affects violations of the reciprocity condition, note that under the original Saaty scale we have that in the interval [1, 9] 6.75% of its size is in ranges that guarantee a violation of the reciprocal property will occur. The value of 6.75% is calculated by observing that the sum of the segments where such violations occur is equal to 0.5403 (see also Section 4.2), or 6.75% of the size of the interval [1, 9]. If this idea is expanded with scales that assume the largest integer value to be equal, say to 10, 11, 12, and 13, then the corresponding percentages are equal to 6.30%, 5.90%, 5.57%, and 5.27%, respectively. If the interval of values of the scales keeps expanding, then the previous percentage will become smaller and smaller approaching the value of zero.

At the extreme hypothetical scenario of having infinite granularity, discretization (rounding) is not needed as in such scenario one can always assign the precise and true (actual) numerical value to any pairwise comparison, under the UADM assumption. However, this would be totally utopic in the real-world. This is one of the first key aspects that Saaty had considered while developing the AHP.

In his 1980 book on the AHP (Saaty, 1980) Saaty referred to Weber's 1846 law on perceivable stimulus of measurable magnitude. This issue is also mentioned in the (Triantaphyllou, Lootsma, et al., 1994) publication on the evaluation of 78 different scales. As recently as in 2001, Saaty had mentioned Weber's 1846 law again along with Fechner's 1860 law on sequences of just noticeable increasing stimuli (Saaty, 2001, page 21).

As result of these findings from psychology, Saaty had asserted that his scale of <u>discrete</u> choices is the best <u>balance</u> between the need for high granularity and the capability of people to distinguish between adjacent domination relationships. If a decision maker feels strongly that he/she can make a distinction between numerical values such as 1.2 and 1.5 in a particular application, then this is fine, and it could / should be incorporated in the set of possible choices. However, as it is described above, there is an upper limit on the granularity of the decision choices one is expected to make.

Using pairwise comparisons is a widely accepted approach for eliciting the people's preferences needed to solve important real-life decision-making problems. This can be easily documented by searching databases of publications under keywords / key phrases such as "pairwise comparisons," "Saaty matrices," "analytic hierarchy process," and so on.

The rank reversal / rank abnormalities problem in the context of the AHP is a well-known subject in MCDM / MCDA for more than 40 years (e.g., since the first publication by Belton and Gear (1983). Since that time many others have examined this issue. However, such publications have focused on the entire MCDM problem (which may involve multiple alternatives and criteria and thus multiple pairwise comparison matrices).

The present study focused on a <u>single pairwise comparison matrix</u> (the RCP matrix and then the corresponding CDP matrix). As CDP matrices are highly consistent by construction (but not necessarily perfectly consistent), there is no significant role played by inconsistency in the judgements. It is the first time in the literature that the issue of **reciprocity violations** is studied. The present study shows that reciprocity violations can play a detrimental role in the quality of the final outcome (ranking of the evaluated entities). The results are dramatic, despite the fact that a deliberately very optimistic assumption (i.e., the UADM assumption) is used in the testing process. This is a first in the literature. This is what makes the proposed UADM-based testing process to be innovative and also crucial to better understand the potential of the pairwise comparisons approach.

It is the different scales in the RCP and corresponding CDP matrices (the first is based on a continuous scale while the second on a discrete scale) and the discretization that occurs when we transform the RCP to the CDP matrix that cause violations of the reciprocity condition (RC). This is guaranteed to occur regardless of which discrete scale is used (see also Theorem 1).

As the world renown management expert, Dr. H. James Harrington has said: "Measurement is the first step that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it" (Good Reads, 2022). Understanding the true capability pairwise comparisons have for solving real-life problems is of ultimate importance.

This study has demonstrated that quantifying pairwise comparisons may lead to major failures. This happens under a specially designed, rather utopic, condition termed in this study as the ultra-accurate decision maker (UADM) assumption. As a result, failures under real-world conditions may, on the average, be even more dramatic.

This study introduced and investigated three fundamental conditions (described as Tests 1, 2, and 3) that must be satisfied by any approach based on pairwise comparisons. These tests are necessary but not sufficient to prove that a pairwise comparison approach (PCA) is effective. In other words, even if a PCA passes all the ranking tests described in this paper, it may still not be perfect and it may fail other ranking tests in the future.

In Section 4 it was proven that the Reciprocity Condition when comparing under the UADM assumption, say entity A_1 with entity A_2, and then independently entity A_2 with entity A_1, may be violated. This happens for certain ranges of values under the UADM assumption. This study demonstrated in Section 5 that such Reciprocity Condition violations may occur with pairwise comparison matrices under the UADM assumption. An extensive computationally study revealed that such violations may occur at dramatic rates even for small size of such matrices. This is the domain of Test 1.

Section 6 demonstrated in terms of an example and a computational study that pairwise comparison matrices may result in rank reversals when one compares the ranks derived after the use of pairwise comparisons and compare them with the true (and hence hidden to the system) rankings of the entities of interest. This is the domain of Test 2.

Section 7 described that other types of ranking abnormalities may occur as well. Such abnormalities are defined as a rank grouping of entities which in reality are distinct. This is the domain of Test 3.

The cause of the ranking problems is twofold: (1) Even the very small level of inconsistency, as described at the end of Section 2, of the CDP matrices (which is well below what Saaty had suggested as "acceptable") can cause ranking abnormalities, and (2) the violations of the reciprocity constraint. The violations of the reciprocity constraint is the most serious cause for such ranking abnormalities. Pertinent are the results in Tables 3.a, 3.b, 4, and 5, and also in the related figures (i.e., Figures 5 to 8).

The previous abnormality rates are quite dramatic, even under the highly optimistic UADM assumption. These results provide a strong impetus to seek ways to improve the performance of methods which are based on pairwise comparison matrices. This study also provides a view of some of the weaknesses that exist with the way such matrices are processed.

This study has demonstrated that when one uses the $n(n-1)/2$ comparisons required by the traditional approach, then there may be too many ranking problems. However, if one uses all $n(n-1)$ pairwise comparisons (i.e., twice as many than the traditional approach of $n(n-1)/2$ comparisons), even when the reciprocity constraint is violated, it still becomes a better approach than the traditional one. This is because more useful information can be incorporated into the relative weight calculation process.

For instance, if the rankings from both Saaty matrices agree with each other, then it is very likely that the common best ranked item is the same as the true (and thus hidden) best ranked item (see also the note in Table 3.b). Such results can only be derived if all the pairwise comparisons are elicited. Other important results may also be deduced if no reciprocity violations are detected in a pairwise comparison matrix. That is, the presence of RC violations may function as a warning flag when using pairwise comparisons. Therefore, one first needs to understand the impact of having all $n(n-1)$ comparisons, and then, in a future study, determine a remedy procedure of what to do when reciprocity violations occur.

In future work we will explore the possibility of identifying more types of ranking abnormalities and corresponding tests when pairwise comparisons are used for decision making. A key part in such investigations will again be the UADM assumption. After this phase of the research is completed, the next logical phase would be to seek ways for mitigating the problems created due to reciprocity violations. The following tabulation provides the state-of-the-art before of this study and the contributions made as result of this study.

Summary Points:

Part 1 of 2: What was already known on the topic?

1. The concepts of the RCP and CDP matrices existed and were used in some limited studies to evaluate the AHP, but not to evaluate pairwise comparisons per se.

2. Ranking problems with the AHP have been studied, but not at the level of pairwise comparison matrices when they are highly consistent (as the CDP matrices are).

Part 2 of 2: What did the study add to our knowledge?

1. The concept of the ultra-accurate decision maker (UADM) is formally introduced as a powerful evaluative tool of pairwise comparison approaches (Section 3).

2. Three fundamental tests are introduced and computationally explored regarding key properties of matrices defined using pairwise comparisons.

3. It is possible for some pairwise comparisons to fail the Reciprocity Condition.

4. Groups of values within the numerical range of the Saaty scale (i.e., in the interval [1/9, 9]) are identified that always violate the Reciprocity Condition (Section 4.2).

5. If all $n(n-1)$ possible pairwise comparisons are elicited (and not only $n(n-1)/2$ of them) **in real-world applications, then the number of Reciprocity Condition (RC) violations may be extreme.** This is true because even under the highly favorable UADM assumption the number of such RC violations is very dramatic (Section 5.2).

6. The presence of RC violations causes pairwise comparison methods to have rank inversions in dramatic frequencies (Section 6.2). As before, this may be even more extreme in real-world applications.

7. The presence of RC violations causes pairwise comparison methods to have abnormal rank groupings in dramatic frequencies (Sections 7.1 and 7.2). As before, this may be even more dramatic in real-world applications.

8. The RCP / CDP approach under the UADM assumption is an effective way for assessing effectiveness of a pairwise comparisons approach when a finite discrete ratio scale is used.

9. There is a strong need to better understand and ultimately deal with the Reciprocity Condition violation problem.

References

Aczél, J. and Saaty, T.L. (1983). Procedures for synthesizing ratio judgements. *Journal of Mathematical Psychology*, 27(1), 93-102.

Aires, R.F.D.F. and Ferreira, L. (2018). The rank reversal problem in multi-criteria decision making: a literature review. *Pesquisa Operacional*, 38, 331-362.

Barzilai, J. and Golany, B. (1994). AHP rank reversal, normalization and aggregation rules. *INFOR: Information Systems and Operational Research*, 32(2), 57-64.

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

Brans, J.P., Vincke, P. and Mareschal, B. (1986). How to select and how to rank projects: The PROMETHEE method. *European Journal of Operational Research*, 24(2), 228-238.

Colson, G. and De Bruyn, C. (1989). Models and methods in multiple objectives decision making. In *Models and Methods in Multiple Criteria Decision Making* (pp. 1201-1211). Pergamon.

Faramondi, L., Oliva, G., Setola, R., and Bozoki, S. (2023). Robustness to Rank Reversal in Pairwise Comparison Matrices Based on Uncertainty Bounds. *European Journal of Operational Research*, 304(2), 676-688.

Figueira, J.R., Greco, S., Roy, B. and Słowiński, R. (2010). ELECTRE methods: Main features and recent developments. In *Handbook of Multicriteria Analysis* (pp.51-89).

Figueira, J.R., Greco, S., Roy, B. and Słowiński, R. (2013). An overview of ELECTRE methods and their recent extensions. *Journal of Multi‐Criteria Decision Analysis*, 20(1-2), 61-85.

Forman, E. and Peniwati, K. (1998). Aggregating individual judgments and priorities with the analytic hierarchy process. *European Journal of Operational Research*, 108(1), 165-169.

Good Reads. (2022). Quote by Dr. H. James Harrington, Available from http://www.goodreads.com/quotes/632992-measurement-is-the-first-step-that-leads-tocontrol-and , accessed in March 2022.

Keeney, R, and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, New York, NY, USA.

Keeney, R.L., Raiffa, H. and Meyer, R.F. (1993). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press. (This is an updated version of the original 1977 publication by Keeney and Raiffa).

Kułakowski, K., 2018. Inconsistency in the ordinal pairwise comparisons method with and without ties. European Journal of Operational Research, 270(1), pp.314-327.

Linares, P., 2009. Are inconsistent decisions better? An experiment with pairwise comparisons. European Journal of Operational Research (EJOR), 193(2), pp.492-498.

Lootsma, F.A., 1990. The French and the American school in multi-criteria decision analysis. RAIRO-Operations Research, 24(3), pp.263-285.

Miller, G.A., 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological review, 63(2), p.81.

Millet, I. and Saaty, T.L., 2000. On the relativity of relative measures–accommodating both rank preservation and rank reversals in the AHP. European Journal of Operational Research (EJOR), 121(1), pp.205-212.

Ray, T.G. and Triantaphyllou, E., 1998. Evaluation of rankings with regard to the possible number of agreements and conflicts. *European Journal of Operational Research*, 106(1), 129-136.

Ray, T. and Triantaphyllou, E. (1999). Procedures for the evaluation of conflicts in rankings of alternatives. *Computers & Industrial Engineering*, 36(1), 35-44.

Roy, B. (1985). Méthodologie multicritère d'aide à la décision, *Collection Gestion, Economica*, Paris, France.

Roy, B. (1989). The Outranking Approach and the Foundations of ELECTRE Methods, *Document* 53, LAMSADE, Université de Paris-Dauphine, France.

Roy, B. (1990). The outranking approach and the foundations of ELECTRE methods. In *Readings in Multiple Criteria Decision Aid* (pp. 155-183). Springer, Berlin, Heidelberg, Germany.

Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3), 234-281.

Saaty, T.L. (1980). *The Analytic Hierarchy Process*. Book published by McGraw-Hill International. New York, NY, USA.

Saaty, T.L. (1989). Group decision making and the AHP. In *The Analytic Hierarchy Process* (pp. 59-67). Springer, Berlin, Heidelberg, Germany.

Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. European Journal of Operational Research (EJOR), 48(1), 9-26. http://dx.doi.org/10.1016/0377-2217(90)90057-I.

Saaty, T.L., 2001. Fundamentals of the Analytic Hierarchy Process. Chapter 2 in the book "The Analytic Hierarchy Process in Natural Resource and Environmental Decision Making," (pp. 15-35). Schmoldt, D.L., Kangas, J., Mendoza, G.A. and Pesonen, M. (Editors). Springer Science & Business Media. Dordrecht, South Holland, The Netherlands.

Saaty, T.L., 2003. Decision-making with the AHP: Why is the principal eigenvector necessary. European Journal of Operational Research (EJOR), 145(1), pp.85-91.

Saaty, T.L. (2008). Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. RACSAM-Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. *Matematicas*, 102(2), 251-318.

Schärlig, A. (1985). Décider sur plusieurs critères: panorama de l'aide à la décision multicritère (Vol. 1). PPUR Presses Polytechniques.

Schenkerman, S. (1994). Avoiding rank reversal in AHP decision-support models. *European Journal of Operational Research*, 74(3), 407-419.

Triantaphyllou, E. and Mann, S.H. (1989). An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox. *Decision Support Systems*, 5(3), 303-312.

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., Lootsma, F.A., Pardalos, P.M. and Mann, S.H. (1994). On the evaluation and application of different scales for quantifying pairwise comparisons in fuzzy sets. *Journal of Multi‐Criteria Decision Analysis*, 3(3), 133-155.

Triantaphyllou, E. and Mann, S.H. (1994). A computational evaluation of the original and revised analytic hierarchy process. *Computers & Industrial Engineering*, 26(3), 609-618.

Triantaphyllou, E. (2001). Two new cases of rank reversals when the AHP and some of its additive variants are used that do not occur with the multiplicative AHP. *Journal of Multi - Criteria Decision Analysis*, 10(1), 11-25.