Fundamenta Informaticae 144 (2016) 1–17
DOI 10.3233/FI-2016-1320
IOS Press

Important Facts and Observations about Pairwise Comparisons (the special issue edition)

Waldemar W. Koczkodaj*
Laurentian University, 935 Ramsey Lake Road
Sudbury, Ontario P3E 2C6, Canada
wkoczkodaj@cs.laurentian.ca

Grzegorz Redlarski†
Gdansk University of Technology
80-297 Gdansk, Narutowicza st. 11/12, Poland
grzegorz.redlarski@pg.gda.pl

Jacek Szybowski‡
AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Krakow, Poland
szybowsk@agh.edu.pl

Eliza Wajch
Siedlce University of Natural Sciences and Humanities
3 Maja 54, 08-110 Siedlice, Poland
eliza.wajch@wp.pl

Ludmil Mikhailov
The University of Manchester
M13 9PL, United Kingdom
ludi.mikhailov@mbs.ac.uk

Michael Soltys
California State University Channel Islands
One University Drive, Camarillo, CA 93012, USA,
michael.soltys@csuci.edu

Gaik Tamazian§
St. Petersburg State University
198207 Srendiy Prospekt, d. 41, St. Petersburg, Russia
g.tamazian@spbu.ru

Kevin Kam Fung Yuen
Xi'an Jiaotong-Liverpool University
Suzhou 215123, China
Kevin.Yuen@xjtlu.edu.cn

Abstract. This study has been inspired by numerous requests for clarification from researchers who often confuse Saaty's Analytic Hierarchy Process (AHP) with the pairwise comparisons (PC) method, taking AHP as the only representation of PC. This study should be regarded as an interpretation and clarification of past investigations of PC. In addition, this article is a reflection on general PC research at a higher level of abstraction: the philosophy of science. It delves into the foundations and implications of pairwise comparisons. Some results of this study are based on a recently published work by Koczkodaj and Szwarc. Finally, open problems have also been reported for future research.

*Address for correspondence: Laurentian University, 935 Ramsey Lake Road, Sudbury, Ontario P3E 2C6, Canada
†Supported in part by the European Union within the European Social Fund – Human Capital Operational Programme HC-OP
‡Research supported by the Polish Ministry of Science and Higher Education
§Supported, in part, by Russian Ministry of Science (Mega-grant no.11.G34.31.0068) and St.-Petersburg State University (grant no.1.37.165.2014)

Keywords: pairwise comparisons, inconsistency analysis, ratio scale, approximation.

1. Introduction

The first use of the method of Pairwise Comparisons (PC method) is often attributed to Ramon Llull, the 13th-century mystic and philosopher. His works created a basis for computation theory, considering his acknowledged influence on Gottfried Leibniz. Llull's contribution to PC, used for elections, was recently mentioned in [1]. Fechner (see [2]) is regarded as the precursor of using PC in Psychometrics. However, it was Waber who really used PC in such way. Similarly to Llull, Condorcet also used PC in [3], for voting but only for the binary choice: win or lose. In both cases, however, the PC method was not the main subject for the scientific investigation but rather its use. Thurstone (see [4]) proposed what is known as "The Law of Comparative Judgments," in 1927.

The seminal study [5] had a considerable impact on PC research, but led to AHP becoming a proprietary eponym for PC. However, we argue that AHP should not be equated with PC. The fixed ratio scale (the scale in this presentation) assumed in AHP makes it a subset of PC. PC is more general as it does not assume any particular scale, and allows us for using even non-numerical rankings. For instance, the non-numerical rankings of [6] are relations. The "scales" in [7] are arbitrary groups while the less general abelian linearly ordered groups (alo-groups) were used in [8, 9].

There is a naming problem with PC. Regretfully, PC has been wrongly used in its singular form. It is a rare form where an adjective is in the plural form. Similarly, it is unacceptable to replace the plural form in *"women's world cup"* by the singular form (woman). The plural form *"comparisons"* is needed in front of a method or a matrix. The term PC should not be confused with the term *"paired comparison"* that is commonly used for the binary entries.

Comparing two entities (also called stimuli or properties) in pairs is irreducible since one entity compared with itself trivially gives 1. Comparing only two entities gives us a PC matrix 2×2 and may be inaccurate but does not involve inconsistency.

Despite its long history, the PC method is still a very attractive subject for research. Ranking with a reference set of alternatives, as in [10, 11], is an example of such explorations. "To pair or not to pair" is not the question. "When and how to use PC" is the proper question.

2. Basic concepts of pairwise comparisons

Usually, we define an $n \times n$ *pairwise comparisons matrix* (abbreviated to *PC matrix*) simply as a square matrix $M = [m_{i,j}]$ such that $m_{i,j}$ are positive real numbers expressing ratios of entities E_i over E_j or properties for all $i, j = 1, \ldots, n$. A pairwise comparisons matrix $M = [m_{i,j}]$ is called:

(r) *reciprocal* if $m_{i,j} = \frac{1}{m_{j,i}}$ for all $i, j = 1, \ldots, n$ (then automatically $m_{i,i} = 1$ for each $i = 1, \ldots, n$);

(c) *consistent* (or *transitive*) if $m_{i,j} \cdot m_{j,k} = m_{i,k}$ for all $i, j, k = 1, 2, \ldots, n$.

Sometimes, it is convenient to transform the PC matrix $M = [m_{i,j}]$ by the logarithmic mapping to get the matrix $\ln(M) = [\ln(m_{i,j})]$ probably for the first time introduced in [12] and used in [13] for computing a consistent approximation of a generalized PC matrix.

For $a_{i,j} = \ln(m_{i,j})$, we have $m_{i,j} = \frac{1}{m_{j,i}}$ if and only if $a_{i,j} = -a_{j,i}$. Moreover, $m_{i,j} \cdot m_{j,k} = m_{i,k}$ if and only if $a_{i,j} + a_{j,k} = a_{i,k}$. Therefore, we may consider two kinds or reciprocities: multiplicative reciprocity $m_{i,j} * m_{j,i} = 1$ for M and additive reciprocity $a_{i,j} + a_{j,i} = 0$ for $\ln(M)$. Analyzing the multiplicative consistency for M and the additive consistency for $\ln(M)$ is of considerable importance. In [8], a unified framework for both multiplicative and additive reciprocity and consistency was introduced by a general notion of a reciprocal PC matrix over an abelian linearly ordered group. In [7], the notion of PC matrices over a group was introduced.

Let us recall that a group is an ordered pair $\langle X, \odot \rangle$, denoted briefly by X, where X is a set, while $\odot$ is a mapping from $X \times X$ to X such that the following conditions are satisfied:

(g1) if $a, b, c \in X$, then $(a \odot b) \odot c = a \odot (b \odot c)$ (associativity);

(g2) there exists exactly one element $1_X \in X$ (called the identity element of the group) such that, for each $a \in X$, the equality $a \odot 1_X = a$ holds;

(g3) for each $a \in X$, there exists $a^{-1} \in X$, called the inverse element of a, such that $a \odot a^{-1} = 1_X$.

In addition, if $a \odot b = b \odot a$ holds for all elements a, b of the group X, then the group is called abelian or commutative.

Now, let $X = \langle X, \odot \rangle$ be a group and let $M = [m_{i,j}]$ be an $n \times n$ matrix such that $m_{i,j} \in X$ for all $i, j \in \{1, \ldots, n\}$. Then we say that M is a *square matrix over the group* X. According to [7], the matrix M is called a *reciprocal* PC *matrix over the group* X if $m_{i,i} = 1_X$ and $m_{j,i} = m_{i,j}^{-1}$ for all $i, j \in \{1, \ldots, n\}$. The matrix M over X is called *consistent* if it satisfies the following consistency (equivalently, transitivity) condition with respect to $\odot$:

$$m_{i,j} \odot m_{j,k} = m_{i,k},$$

for all $i, j, k = 1, 2, \ldots, n$.

We denote the group of all positive real numbers equipped with their standard multiplication by $\mathbb{R}_+$. Most applications of PC matrices have been found for PC matrices over the group $\mathbb{R}_+$. For this reason, in the sequel, we assume that M is a reciprocal PC matrix over $\mathbb{R}_+$. Let M be of the form:

$$M = \begin{bmatrix} 1 & m_{1,2} & \cdots & m_{1,n} \\ \frac{1}{m_{1,2}} & 1 & \cdots & m_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{m_{1,n}} & \frac{1}{m_{2,n}} & \cdots & 1 \end{bmatrix}$$

where $m_{i,j}$ expresses a relative quantity, intensity, or preference of entity (or stimuli) E_i over E_j.

While every consistent matrix is reciprocal, the converse is false, in general. If the consistency condition does not hold, the matrix is *inconsistent* (or *intransitive*).

Consistent matrices correspond to the ideal situation where exact values exist for the ratios of entities: $E_1, \ldots, E_n$. For a consistent PC matrix, there is a vector $w = [w_1, \ldots w_n]$, unique up to a multiplicative constant, which generates this matrix, i.e., the condition $m_{ij} = w_i / w_j$ holds. We usually call w *the vector of weights*.

The main challenge for the PC method is the lack of consistency in the PC matrices, which in practice often takes place – in terms of realistic input data, most PC matrices are inconsistent since "to err is human". Only simple or academic examples are fully consistent and for such cases, a simplified PC matrix, proposed in [14], should be considered, as it requires only $n-1$ comparisons.

Given an $n \times n$ matrix M, which is not consistent, the theory attempts to provide a consistent $n \times n$ matrix M, which differs from matrix M "as little as possible". Such approximation is acceptable for applications when the inconsistency is at an acceptable level. Similarly to p-value in statistics, the inconsistency acceptable level has to be set arbitrarily for each application. However, the threshold 1/3 turns to be acceptable for most applications.

Given any strictly positive real values v_i, the PC matrix $[v_i/v_j]$ is always consistent. It is an important observation since it implies that a problem of approximation is really a problem of a norm selection and distance minimization. For the Euclidean norm, the vector of geometric means (equal to the principal eigenvector for the transitive matrix) is the vector of weights which generates it. One should note that it is also equal (when normalized) to an eigenvector corresponding to the principal eigenvalue. Needless to say, only optimization methods can approximate the given matrix for the assumed norm as addressed in Sec. 8. Such type of matrix is examined in [15] as an "error-free" PC matrix.

As it has been mentioned, the additive version of PC can be obtained by a logarithmic transformation. It has been recently analyzed by Yuen in [16]. Non-numerical PC have been introduced by Janicki in [17] and recently analyzed in [6] where the authors refer to the famous book [18] by the Nobelist K. Arrow (without commenting on Arrow's Impossibility Theorem) but the entire section devoted to compliance with the Arrow's consistency rules is in [19].

3. The ratio scale problem

A scale may be considered as a mapping from a set of qualitative assessments to the set of real numbers. The qualitative assessments arise from comparing pairs of objects that have a natural order, and are usually considered to form a constrained set. In the past, 7 ± 2 assessment gradations were considered due to psychological limitations of human thinking [20] but the lower "magical numbers", as low as 3, have been recently considered (see [21]).

Following [22], we formalize the notion of a scale in the following way. Let $\Lambda = \{\, 0, \pm 1, \ldots, \pm 8 \,\}$ be the set of numbers representing qualitative assessments (see Table 1). We assume that for each pair of objects (C_i, C_j) being compared with each other, a number $\lambda_{i,j} \in \Lambda$ is assigned in the following way:

> if the first object in the pair C_i is preferable to the second one C_j, then a positive number from Λ is assigned; otherwise the assigned number from Λ is negative if $i \neq j$, and it is 0 for $i = j$.

Since PC are reciprocal, it is reasonable to assume that if C_i is preferable to C_j to a certain extent expressed by $\lambda_{i,j} \in \Lambda$, then C_j is inferior compared to C_i with the same extent $\lambda_{j,i} \in \Lambda$ by the degree but of the opposite sign: $\lambda_{j,i} = -\lambda_{i,j}$.

A ratio scale is represented by a mapping f defined on Λ such that $f(-\lambda) = \frac{1}{f(\lambda)}$ for each $\lambda \in \Lambda$. This condition ensures that scale values $a_{i,j} = f(\lambda_{i,j})$ and $a_{j,i} = f(\lambda_{j,i})$ corresponding to reciprocal assessments are indeed reciprocal values:

$$a_{i,j} \cdot a_{j,i} = f(\lambda_{i,j}) \cdot f(\lambda_{j,i}) = f(\lambda_{i,j}) \cdot f(-\lambda_{i,j}) = f(\lambda_{i,j}) \cdot \frac{1}{f(\lambda_{i,j})} = 1.$$

As an example, we consider AHP ratio scale [5] which is determined by the values $1/9, \ldots, 1/2, 1, 2, \ldots, 9$. In the terms given above, AHP scale is represented by the following mapping:

$$f_S(\lambda) = (1 + |\lambda|)^{\mathrm{sgn}\lambda}. \tag{1}$$

Note that for $\lambda \geq 0$, the function (1) is a linear function of λ.

In this section, we will review alternative scales for pairwise comparisons proposed in [23, 24, 25, 26] and address the problem of choosing the most appropriate scale for a given scenario. The scales are given as mappings defined on Λ.

The Ma-Zheng scale [23] is defined in the following way:

$$f_{MZ}(\lambda) = \left(\frac{9}{9 - |\lambda|}\right)^{\mathrm{sgn}\lambda}. \tag{2}$$

The motivation for (2) was to propose a scale that would be linear for $\lambda \leq 0$, in the same way as AHP scale is linear for $\lambda \geq 0$.

The Lootsma scale [24] was proposed in the context of the multiplicative AHP and is based on psychological insights. It is defined as follows:

$$f_L(\lambda) = c^{\lambda}, \tag{3}$$

where $c > 1$ is a scale parameter.

The Donegan-Dodd scale [25] was suggested to handle the most extreme assessments ($\lambda = \pm 8$ in our notation). Its mapping is based on the inverse hyperbolic tangent function:

$$f_{DD}(\lambda) = \exp\left[\tanh^{-1}\left(\frac{\sqrt{3}|\lambda|}{14}\right)\right]^{\mathrm{sgn}\lambda}. \tag{4}$$

The Salo-Hämäläinen scale [26] is related to so-called *balanced scales* and was designed to give more uniform ('balanced') distribution of scale values compared to AHP scale. It is defined in the following way:

$$f_{SH} = \left(\frac{0.5 + |\lambda|s}{0.5 - |\lambda|s}\right)^{\mathrm{sgn}\lambda}, \tag{5}$$

where s is a scale parameter and is usually equal to 0.05 or $1/17$.

A number of studies were carried out to compare the scales to each other and determine the most suitable for practical use. According to [27], the Lootsma-like scale is the most appropriate based on the scale transitivity criterion. The authors also describe a way to derive the proper value of the scale parameter c. In [28], scales of Saaty, Ma-Zheng, Lootsma and Donegan-Dodd were compared to each other using a Monte-Carlo simulation study. The criterion for comparison was the symmetry of the priority value distribution; according to this criterion, the Ma-Zheng scale was optimal. The readers are also encouraged to reference [29, 30, 22] for further investigation of the scale problem.

A mathematical proof that a small ratio scale (1 to 3) has the most desired mathematical properties (e.g., convexity) was provided in [31]. However, it is not the only reason postulated in this study for using the ratio scale in PC. The main reason arises from the natural language, by the use of comparative adjectives (e.g., big $\rightarrow$ bigger $\rightarrow$ biggest).

Table 1. Representation of qualitative assessments as integer numbers.

$\lambda \in \Lambda$	Qualitative assessment
0	equivalence
2	weak superiority
4	strong superiority
6	very strong superiority
8	absolute superiority
1,3,5,7	intermediate assessments

The Fülöp's constant was introduced in [32]. Subsequently, it was employed in [31] for the derivation of the main result (the small scale). As proved in [31], there exists $a_0 > 0$ such that for any $a > 0$, the univariate function f_a defined as:

$$f_a(x) = (e^x - a)^2 + \left(e^{-x} - 1/a\right)^2 \tag{6}$$

is strictly convex if and only if $1/a_0 \leq a \leq a_0$.

When the condition $1/a_0 \leq a \leq a_0$ is fulfilled for all i,j, then f_a can be transformed into the convex programming problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n-1} f_{a_{in}}(x_i) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} f_{a_{ij}}(x_{ij}) \\ \text{s.t.} \quad & x_i - x_j - x_{ij} = 0,\ i = 1, \ldots, n-2,\ j = i+1, \ldots, n-1. \end{aligned} \tag{7}$$

with a strictly convex objective function to be minimized (see [32], Proposition 2). It implies that the programming problem (7) and the equivalent problem (6) have a unique solution which can be found using standard local search methods. The mentioned constant equals to $a_0 = ((123 + 55\sqrt{5})/2)^{1/4} = \sqrt{\frac{1}{2}\left(11 + 5\sqrt{5}\right)} \approx 3.330191$, what is a reasonable bound for real-life problems [32, 31].

However, an unbounded ratio scale is indispensable for measurable entities such as distances, areas, or temperatures if the precise measurements are available or could be obtained. There is an ongoing dispute about ratio scales for entities lacking the "unit" (e.g., emotions or safety). Usually, the ratio scale is assumed as arbitrary and the Likert scale is usually a ratio scale with small values that are "close enough" to 3, on each branch (negative and positive) as Fig. 1 demonstrates.

Figure 1. Comparing two scale items.

For the following PC matrix of the size 3×3:

$$M3 = \begin{bmatrix} 1 & 9 & x \\ 1/9 & 1 & 9 \\ 1/x & 1/9 & 1 \end{bmatrix}$$

there is no such x in the AHP scale that the matrix $M3$ is consistent. The sum of sets $A = \{1, \ldots, 9\}$ and $B = \{\frac{1}{n} : \ n \in A\}$ of AHP matrices' elements is not closed under the multiplication (e.g., 9×9 does not belong to the sum of sets). It means that the set of values in $A \cup B$ fails to create a group or even the weakest algebraic structure, which is a magma (or groupoid). Evidently, it is not a good news from the mathematical point of view for developing any theoretical foundations.

Unfortunately, the same reasoning does apply to any other fixed value of the upper bound for a rating scale. Evidently, the scale $[1, \infty)$ (and its inverse $(0, 1]$) is a scale which does not suffer from the above lack of the mathematical closure problem. It appears that such a scale is consistent with the *zero, one, and infinity* rule often used in computer science having the interpretation of: the only reasonable numbers are zero, one, and infinity [33].

The 0-1-∞ rule implies that we should not impose arbitrary limits on the number of instances of a particular entity in the design of algorithms or heuristics. If more than a single instance of it is allowed, then the size should not have a fixed limit. Certainly, practicality may impose limits but this should be done by design and necessity rather than chance. Binary alternatives are the most commonly used in the decision making process (e.g., go left or right, buy or sell, pass or fail, etc.). There are well-defined cardinalities other than zero, one, and the arbitrary cardinality (represented by ∞ in the above rule). Most Indo-European languages have developed a concept of pairs or couples. It is also demonstrated by: "This" and "Not this" or "This" and "The other," or "One way" and "The other way". Obviously, all computer technology is based on the binary system.

4. The eigenvalue problem

For the Euclidean norm, the vector of geometric means is the best approximation to a consistent matrix. However, the "best" approximation of an inconsistent PC matrix is not generated by the vector of geometric means or by the eigenvector, although both are very accurate approximations for small inconsistencies. The problem is that no one actually knows "how small is small," although it should not be a surprise since a similar problem can be found in statistics with the p-value and α-value.

It is worth noticing that the normalized vector of the geometric means and the normalized eigenvector (corresponding to the principal eigenvalue) are identical for a consistent PC matrix. The Monte Carlo study was conducted to verify whether the eigenvector approximation is better than the geometric means as proposed in [34], and it disproved such a claim on the basis of 1,000,000 matrices. However, its proponent still insists on the superiority of an eigenvector solution although Example 1, represented by Table 1 in [35], is better approximated by the geometric means than by an eigenvector. The same reasoning is applicable to most examples in Table 2.

The eigenvector approximation was criticized in the 1980s when the geometric mean method was proposed in [12]. More recently, better argumentation was used in [36]. In [37], it was established that the geometric mean is the only method for deriving weights from multiplicative PC that satisfy

fundamental consistency requirements. According to [37], it is immune to the scale-inversion as well as to rank reversal.

The compelling argument against the eigenvector "superiority" is provided by John Fichtner in his 1984 report (see [38]). His eigenvector metric was misused by many researchers who have cited it. Fichtner never claimed that his eigenvector distance should be used instead of the Euclidean distance.

> Nevertheless here two methods for getting priority vectors from reciprocal matrices (geometric row mean and right eigenvector) and two possible scales (1;2;3;... and $1;\sqrt{2};2;2\cdot\sqrt{2};...$) will be compared with the help of 4 qualitative comparison matrices given by Saaty (1980 ; relative visual brightness p. 38 f. and 58, wealth of nations, p. 40 f. and 59 and estimating distances p. 42 f.). The result can be seen in table 1. From these few examples it seems impossible to decide, which method or scale is better.

Figure 2. Excerpt from Fichtner's 1984 report.

In fact, Fichtner's 1984 report findings have been misused as Fig. 2 so well demonstrates. The Fichtner's distance for eigenvectors between PC matrices A and B of the same degree $n > 2$ was defined on pages 37-38 of [38] as follows:

$$\begin{aligned}\delta(A,B) &= \sqrt{(w_1(A)-w_1(B))^2+...+(w_n(A)-w_n(B))^2} \\ &+ \frac{1}{2(n-1)}\left|\lambda_{\max}^{(A)}-\lambda_{\max}^{(B)}\right| \\ &+ \frac{1}{2(n-1)}\left|\lambda_{\max}^{(A)}+\lambda_{\max}^{(B)}-2n\right|\cdot\begin{cases}0 & \text{if} \quad A=B\\ 1 & \text{if} \quad A\neq B\end{cases}\end{aligned}$$

where:

$w_i(A)$ and $w_i(B)$ are, respectively, the i-th components of the normalized principal eigenvectors of PC matrix A and of PC matrix B;

$\lambda_{max}^{(A)}$ and $\lambda_{max}^{(B)}$ are the corresponding principal eigenvalues;

Note: the above excerpt was taken from an old typewritten manuscript and transcribed as accurately as possible.

We leave it to the readers to assess whether it is "simpler and more accurate" than the Euclidean distance.

5. The inconsistency concept

Erroneously, [5] is often credited for providing the first consistency definition of a PC matrix A. In fact, the condition "if and only if $a_{i,j} \cdot a_{j,k} = a_{i,k}$ for $i, j, k = 1, 2, ..., n$." was defined and examined before 1977 by at least these four studies published between 1939 and 1961: [39, 34, 40, 41].

According to [42], a single triad (x, y, z) and its inconsistency indicator ii are expected to have the following properties:

(A.1) $ii(x, y, z) = 0$ if and only if $y = x \cdot z$,

(A.2) $ii(x, y, z) \in [0, 1)$ - the "complete inconsistency" (represented by 1) cannot be achieved,

(A.3) for a consistent triad $ii(x, y, z) = 0$ with $xz = y$, increasing or decreasing x, y, z results in increasing $ii(x, y, z)$.

The third axiom can be summarized by "the further we go away from $y = x \cdot z$, the bigger is the inconsistency value". However, the formal specification of this axiom is not as easy as it seems to be and an independent study on it is under review. In [43], the inconsistency indicator is introduced for a triad (x, y, z) as:

$$ii(x, y, z) = 1 - \min\left\{\frac{y}{xz}, \frac{xz}{y}\right\}.$$

It is equivalent to:

$$ii(x, y, z) = 1 - e^{-\left|\ln\left(\frac{y}{xz}\right)\right|}.$$

ii is extended to a PC matrix by $\max$ for all triads in it and refereed to as Kii (*Koczkodaj's inconsistency indicator*). It is important to note here that the definition of Kii allows us to localize the inconsistency in a PC matrix.

Let us look at the following two examples from [42]:

- $ii(1.5, 2, 2.5)$ will increase if 1.5 or 2.5 are increased, since 1.5*2.5 is already greater than 2. On the other hand, decreasing 2 should also increase the inconsistency.
- $ii(1.5, 2.5, 1.2)$ will increase if 2.5 is increased, since it is greater than 1.5*1.2=1.8, but decreasing 1.5 or 1.2 should also increase inconsistency for the same reason.

In [5], the *consistency index* (CI) was proposed as a scaled deviation of the difference between PC matrix size and the principal eigenvalue:

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

with the highly disputable acceptance threshold of 10% (later on decreased for a higher size of n).

The corner PC matrix CPC is a counter-example for the incorrectness of the eigenvalue-based inconsistency with the mathematical reasoning in [42]. Having $x > 1$, we set $CPC[1, n] := x$, $CPC[n, 1] := 1/x$ and 1's in as all remaining CPC elements. It may seem that there is only one triad but in fact, there are $n-1$ of them. It is easy to see that $Kii = 1-1/x$ and is invariant of n while CI changes with n and its upper bound is estimated in [42] to x/n^2 hence no matter how large x is assumed, there is such n that CI is acceptable despite the presence of $n-1$ triads $(1, x, 1)$. In particular, there may be triads $(1, 10^6, 1)$ where

the consistency condition requires $1 \cdot 1$ to be 1 and it is equal to 10^6 or more and CI is still below any assumed threshold. CPC Matrix is not just a curiosity. Another counter example, FPC matrix is demonstrated in Section 6 [42] with $x > 1$ above the main diagonal and $1/x$ below the main diagonal. Considering x to be a real number, the cardinality of all possible triads in a CPC matrix $n \times n$ is $O(n)$ ($O(n^3)$ for FPC matrix) but the set of all values for x, hence all possible triads $(1, x, 1)$, is of cardinality $2^{\aleph_0}$ in both cases.

It seems that all inconsistency indicators, based on *"central tendency"*, suffer from the similar problem because the consistency condition has (by the definition) the requirement "for all triads" and with the growing PC matrix size ($n \longrightarrow \infty$), the value of inconsistency indicator vanishes but the local estimation error does not. If we accept the inconsistency "vanishing problem" as normal, by the same logic, we should not worry about one "nuke" left behind by the collapsed Great Empire since it is only one lost weapon for approximately 7,000,000,000 inhabitants of this planet. The nuke must be located and possibly destroyed. The same goes for an inconsistency. Once located and found as unacceptable, the inconsistency must be reduced before any further computations take place. Ignoring the unacceptable inconsistency and computing the approximation is as incorrect as the following "proof". Let us assume $x = 0$ giving us $x = 2{*}x$ since $0 = 2{*}0$. If so, we can simplify it by dividing both side of the equation $x = 2{*}x$ by x, getting $0 = 1$, which we know is not exactly so. Our error comes from knowing that $x = 0$ and dividing over it. It should never take place. Similarly, we should not use unacceptably large inconsistent assessment to estimate our solution but AHP (and other inconsistency based on 'central tendency") ignores such principle. As a consequence, an error of the arbitrarily large value is tolerated as demonstrated above.

The opponents of the above examples may say: "...but dividing by 0 is an exceptional phenomenon". Apparently, $\frac{1}{10^{-n}}$ is not just one case. It generates arbitrary large values for growing n, although the divisor is not exactly 0 and there are ∞ of them. Drawing mechanical conclusions, based on the "central tendency," may be risky, as the following example demonstrates. Let us look at the average weight of 80 kg (200 lbs) in a given population of 100 men. It does not imply that there is even a single man with the weight of 80 kg nor near to that average. For example, it may happen that 50 men may each weigh 70 kg and 50 weigh 90 kg, which results in average of 80 kg. In the case of inconsistency indicators, the unacceptably high inconsistency in one triad may not contribute much to "the average" but it is unacceptable. Indeed, most of the local approximation error aberrations occur and the argument that "we only lost one nuclear weapon" (which may be "accidentally found" by terrorists) is unacceptable to most of us. However, it does take place for the eigenvalue-based inconsistency as demonstrated by [42] for the CPC counter-example.

Common sense dictates a simple rule: "locate the worse inconsistency and inspect it; if unacceptable, reduce it". Certainly, if we are unable to reduce the inconsistency or there is an important reason (e.g., time) to continue with the inconsistency, that may do so. It is also important to notice that there are no inconsistencies in the simplified version of PC, recently introduced in [14]. However, the zero inconsistency indicator does not guarantee that assessments are accurate. The consistent (or doctored data) ignorance can give result in 0 inconsistency. In fact, playing safe and giving all 1s for all entries of the entire PC matrix gives 0 inconsistency, but may be a sign of the total ignorance, where everything is of equal importance. Certainly, everything may be of equal (or unknown) importance and in such case, all PC matrix entries should be equal to one.

6. The inconsistency reduction process

Localizing the inconsistency was proposed in [43]. The consistency-driven process was postulated since it comes from the GIGO (garbage-in, garbage-out) principle and common sense. However, it is still not certain whether we should begin the reduction process, starting with the worse triad (as the common sense may dictate) or to improve the triad with the smallest inconsistency. In the second case, smaller errors may propagate through the PC matrix. A Monte Carlo study has been launched and the preliminary findings support the foreknowledge in 1993.

It is safe to assume that processing (e.g., approximating) random numbers gives random results. In the case of PC, the "randomness" of the input is associated with the inconsistency. It is evidenced in Fig. 3.

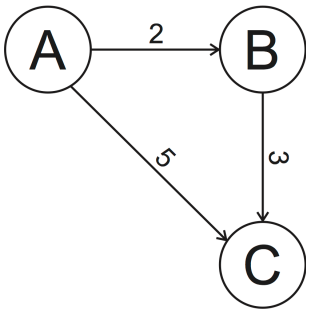

Figure 3. An inconsistent triad (2,5,3)

If we assume that $A/B = 2$ and $B/C = 3$, evidently A/C cannot be 5 but rather 6. The problem is that these three comparisons might have been conducted with the help of the Internet, on three different continents, based on the "undisputed taste" (as earlier noted). No one can assert which of these three comparisons is (or is not) accurate.

The illustration of the inconsistency is even more drastic by the PC matrix $M3$, which is generated by the triad $(2, 1/2, 2)$. It gives unnormalized weights $v = [1, 1, 1]$ ($[1/3, 1/3, 1/3]$ in its normalized form). Vector v generates a trivial PC matrix with all entries equal to 1 representing ignorance (lack of knowledge). For this reason, we have to improve all individual triads. Figuring out which strategy should be employed, is the key issue for future research.

So far, many PC method studies fail to identify the real problem with verification of the approximation accuracy. When we try to verify results on real examples related to subjective data (say, public satisfaction), how can we be sure what is and what is not "the right solution"? After all, it is said that the taste is indisputable and few academic examples prove nothing.

The reduction process in [43, 44] was based on the common sense approach. It called for finding the most inconsistent triad and reducing the inconsistency in this triad. The reduction of the triad with the smallest inconsistency is another strategy. Why? It is suspected that the error propagation may be smaller than the triad with the worst inconsistency. It is probably impossible to provide an analytical reasoning why an inconsistency is undesirable (other than the GIGO adage but is it not a proof). It is not even easy to conduct a Monte Carlo experimentation to support it although such preliminary study was conducted (see [45]) with the randomly generated images for its area estimation.
"Summary of findings" in [46] includes:

> "The practical implications of identity (6) is that the numerical estimates produced by Saaty's eigenvalue-based method and the geometric mean procedure of de Jong (1984) and others should be close to each other when a respondent's judgments are not too far from being consistent."

Identity (6) is $\hat{w} = \hat{\hat{w}} + O(\sigma^2)$ where w_i are parameters (eigenvector or geometric means) and $\sigma^2 > 0$ is their common variance. However, the old adage that "one bad apple spoils the barrel" seems to be more applicable here: even one massively inconsistent triad may contribute to significant errors. An approximation of a PC matrix is meaningful, only if the initial inconsistency is acceptable (that is, located, brought under control and/or reduced to a certain predefined minimum; in our analogy, always remove an overripe fruit promptly if it is possible to find it).

In practical applications, a high value of the inconsistency indicator is a "red flag," or a sign of potential problems. A distance-based inconsistency reduction algorithm focuses, at each step, on an inconsistent triad and "corrects" it by replacing it with a consistent (or, more generally, a less inconsistent) triad. It resembles "Whac-a-Mole," a popular arcade game. One difference is that instead of a mole, we have three array elements, as explained above. After hitting the mole (which generally results in some other "moles" appearing), the next triad is selected according to some rule (which may be for example the greedy algorithm) and the process is repeated. Numerous practical implementations (e.g., a hazard ratio system for abandoned mines in Northern Ontario) have shown that the inconsistency converges relatively quickly. However, the need for rigorous proof of the convergence (that is, showing that the whacked moles always have the tendency of coming out less and less eagerly) was evident. An approximation of a PC matrix is meaningful only if the inconsistency in it is acceptable. This means that we need to localize the inconsistency and reduce it to a certain predefined minimum, if it is too high.

7. Not-so-inconsistent matrices and Monte Carlo experiments

Using completely random PC matrices for testing has very little scientific merits, since they are simply random numbers and defy all principles of learning (machine or natural). Common sense dictates to use matrices somehow inconsistent but not just a proverbial "bunch of random numbers". We will call such a PC matrix "not-so-inconsistent" (NSI) PC matrix. The NSI matrix was defined in [47] as follows. We obtain an NSI PC matrix M from a random vector v with positive coordinates by: $M = [v_i/v_j]$ where $i, j = 1, 2, \ldots, n$. We deviate M by random multipliers $m_{ij} := m_{ij} * rand()$.

NSI PC matrices have been recently used in [48] to verify how quick is the convergence to the consistency by the reduction process. The results were very positive. The number of iterations to reduce the inconsistency below the accepted level (assumed as $1/3$ for the distance-based inconsistency) were 10 at the most.

8. Deriving priorities from PC matrices

In PC we compare two elements, and assign a value, which represents an assessment of the relative preference of E_i over E_j . If E_i is preferred to E_j then $a_{ij} > 1$, otherwise $0 < a_{ij} < 1$. A full set of assessments for n elements requires $n(n-1)/2$ comparisons. The eigenvector method (EV), proposed in [5], is based on the Perron-Frobenius theorem. It provided a proof that the right principal

eigenvector of A can be used as a priority vector, so the EV solves the equation: $Aw = \lambda_{max} w$, where λ_{max} is the principal eigenvalue of A. However, the EV method has numerous drawbacks, discussed in the previous sections. Numerous alternative prioritization methods have been proposed. Many are based on the optimization approach. These methods need an objective function, which measures the accuracy of the approximation of the initial comparison assessments. Thus, the problem of priority derivation is formulated as an optimization task of minimizing the objective function, subject to normalization and possible additional constraints.

The Direct Least Squares (DLS) method (probably introduced in [12]) is based on the assumption that the errors between the initial assessments and solution ratios should be minimized, and therefore uses the Euclidean distance metric (or its square) as an objective function. The prioritization task is formulated as a constrained optimization problem:

$$\text{MinED} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - w_i/w_j)^2 \right)^{1/2}$$

subject to: $\sum_{i=1}^{n} w_i = 1, w_i > 0$.

On page 312 in [49], authors state:

> *"Remark: The most frequently used metric to measure closeness and accuracy in n-dimensional spaces is the Euclidean metric."*

Common sense dictates that there is a reason why the Euclidean metric is the most frequently used. It has been used for hundreds of years since it is simple, elegant and accurate. The above quotation was not only true in 1984 but it is still true at the time of this study (after 31 years). It is evident why authors followed it with an "explanation":

> *"However, the Euclidean metric does not address the question of inconsistency."*

Common sense dictates that no approximation can really give a satisfactory solution for heavily polluted data, hence the inconsistency analysis should precede any approximation, as discussed in former sections. With the proper inconsistency analysis, MinED is the most practical solution since it is simple and accurate. After all, the statistical results (based on 1,000,000 randomly generated case in [47] and numerous smaller Monte Carlo experimentation) support it.

9. Notable applications

The number of applications of PC is ever growing. It is currently one of a few valid methods for processing subjective data. In the past, the PC method was used for decision making on the national level, related to nuclear weapons or energy [50]. Certainly, making such decisions must be guided by common sense "rules" and GIGO is one of them. In [51] the authors propose a research program for developing a formal framework for ranking procedures based on the PC method. They also provide a case study based on [52].

It is also important to point out that one of the sizable countries in the European Union has already passed a law to use PCs for evaluations at the national level, as documented in [53], where scientific

entities have been evaluated. Credits for introducing PC to the ministerial regulations in Poland should be given to Prof. R. Slowinskii, a Member of the Polish Academy of Science.

10. Conclusions

There is still much to be done in PC. The biggest challenge is not just the technology: hardware and software, but rather the theory itself. However, complicating the existing theory is not a solution. The "best practices," as they are often used in the Software Engineering approach and more Monte Carlo experimentations are needed.

Much has been done in this direction but more research is definitely still needed. Firstly, the input values should be extended to indicate the certainty of the assessment. It seems that fuzzy logic is the best approach to express uncertainty. Although the fuzzy logic is widely accepted for modeling uncertainly and approximate reasoning, Saaty vigorously opposed its use in two of his publications by using "On the invalidity of fuzzifying numerical judgments" in [54] and "There Is No Mathematical Validity for Using Fuzzy Number Crunching" in [55], despite hundreds of well-documented successful applications (e.g., [56] with 695 citations as of 2015-03-27). The hierarchical classifier has not been given enough attention. In 1960s and 1970s, the hierarchical approach was popularized by many researchers in systems analysis. The top-down design (based on a hierarchical structure) was promoted by IBM researchers Harlan Mills and Niklaus Wirth, in the late 1960s and the early 1970s. The top-down strategy results in a hierarchical structure. Dijakstra may be the first researcher who examined the influence of hierarchy to control complexity in [57] in 1969. The HIPO model (abbreviation for Hierarchical Input Process Output model) was introduced by IBM in early 1970s in [58], with a follow-up book [59]. However, the hierarchical classifier has been assumed rather than computed in most applications. The hierarchical clustering seems to offer the best solution for a non-supervised learning.

Acknowledgments

The title of this study was inspired by the title of a report published at ResearchGate (`https://www.researchgate.net`) downloaded by Prof. K. Arrow, Nobel Laureate.

References

[1] Faliszewski P, Hemaspaandra E, Hemaspaandra LA. Using complexity to protect elections. Communications of the ACM. 2010;53(11):74–82. doi:10.1145/1839676.1839696.

[2] Fechner G. Elements of Psychophysics. Vol. I. Henry Holt editions in psychology. New York : Holt, Rinehart & Winston; 1966.

[3] de Condorcet N. Essay on the Application of Analysis to the Probability of Majority Decisions. Paris: l'Imprimerie Royale; 1785.

[4] Thurstone LL. A law of comparative judgment. Psychological Review. 1927;34(4):273.

[5] Saaty TL. A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology. 1977;15(3):234–281. doi:10.1016/0022-2496(77)90033-5.

[6] Janicki R, Zhai Y. On a pairwise comparison-based consistent non-numerical ranking. Logic Journal of IGPL. 2012;20(4):667–676. doi:10.1093/jigpal/jzr018.

[7] Koczkodaj WW, Szybowski J, Wajch E. Inconsistency indicator maps on groups for pairwise comparisons. International Journal of Approximate Reasoning. 2016;69(2):81–90. Available from: http://arxiv.org/abs/1502.06160. doi:http://www.sciencedirect.com/science/article/pii/S0888613X15001747.

[8] Cavallo B, D'Apuzzo L. A general unified framework for pairwise comparison matrices in multicriterial methods. International Journal of Intelligent Systems. 2009;24(4):377–398. doi:10.1002/int.20329.

[9] Ramík J. Pairwise comparison matrix with fuzzy elements on alo-group. Information Sciences. 2015;297:236–253. doi:10.1016/j.ins.2014.11.010.

[10] Kułakowski K. A heuristic rating estimation algorithm for the pairwise comparisons method. Central European Journal of Operations Research. 2013;23(1):187–203. doi:10.1007/s10100-013-0311-x.

[11] Kułakowski K. Heuristic Rating Estimation Approach to The Pairwise Comparisons Method. Fundamenta Informaticae. 2014;133(4):367–386. doi:10.3233/FI-2014-108.

[12] Jensen RE. An alternative scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology. 1984;28(3):317–332.

[13] Koczkodaj WW, Orłowski M. Computing a consistent approximation to a generalized pairwise comparisons matrix. Computers & Mathematics with Applications. 1999;37(3):79–85.

[14] Koczkodaj WW, Szybowski J. Pairwise comparisons simplified. Applied Mathematics and Computation. 2015;253:387–394.

[15] Temesi J. Pairwise comparison matrices and the error-free property of the decision maker. Central European Journal of Operations Research. 2011;19(2):239–249. doi:10.1007/s10100-010-0145-8.

[16] Yuen K, Kam F. Pairwise opposite matrix and its cognitive prioritization operators: comparisons with pairwise reciprocal matrix and analytic prioritization operators. Journal of the Operational Research Society. 2012;63(3):322–338. Available from: http://dx.doi.org/10.1057/jors.2011.33.

[17] Janicki R. Pairwise Comparisons Based Non-Numerical Ranking. Fundamenta Informaticae. 2009;94(2):197–217. doi:10.3233/FI-2009-126.

[18] Maskin E, Amartya Sen with contributions from Kenneth J Arrow PKP Partha Dasgupta, Stiglitz JE. The Arrow Impossibility Theorem. Kenneth J. Arrow Lecture Series. Columbia University Press; 2014. Available from: http://www.jstor.org/stable/10.7312/mask15328.

[19] Janicki R, Koczkodaj WW. A Weak Order Approach to Group Ranking. Computers & Mathematics With Applications. 1996;32(2):51–59. doi:10.1016/0898-1221(96)00102-2.

[20] Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review. 1956;63(2):81. Available from: http://dx.doi.org/10.1037/h0043158.

[21] Clermont KM. Procedures Magical Number-3 - Psychological Bases for Standards of Decision. Cornell Law Review. 1987;72(6):1115–1156. Available from: http://scholarship.law.cornell.edu/facpub/266.

[22] Kusherbaeva V, Sushkov Y, Tamazian G. Scales and methods for deriving weights in the analytic hierarchy process. Vestnik St Petersburg University: Mathematics. 2011;44(4):282–291. doi:10.3103/S106345411104008X.

[23] Ma D, Zheng X. 9/9–9/1 scale method of AHP. In: Proceedings of the 2nd International Symposium on the AHP. vol. 1; 1991. p. 197–202.

[24] Lootsma FA. Scale sensitivity in the multiplicative AHP and SMART. Journal of Multi-Criteria Decision Analysis. 1993;2(2):87–110. doi:10.1002/mcda.4020020205.

[25] Donegan HA, Dodd FJ, McMaster TBM. A new approach to AHP decision-making. The Statistician. 1992;41(3):295–302. Available from: http://www.jstor.org/stable/2348551. doi:10.2307/2348551.

[26] Salo AA, Hämäläinen RP. On the measurement of preferences in the analytic hierarchy process. Journal of Multi-Criteria Decision Analysis. 1997;6(6):309–319. doi:10.1002/(SICI)1099-1360(199711)6:6< 309::AID-MCDA163>3.0.CO;2-2.

[27] Ji P, Jiang R. Scale transitivity in the AHP. Journal of the Operational Research Society. 2003;54(8):896–905. Available from: http://www.jstor.org/stable/4101660.

[28] Beynon M. An analysis of distributions of priority values from alternative comparison scales within AHP. European Journal of Operational Research. 2002;140(1):104–117. doi:10.1016/S0377-2217(01)00221-1.

[29] Dong Y, Xu Y, Li H, Dai M. A comparative study of the numerical scales and the prioritization methods in AHP. European Journal of Operational Research. 2008;186(1):229–242. doi:10.1016/j.ejor.2007.01.044.

[30] Elliott MA. Selecting numerical scales for pairwise comparisons. Reliability Engineering & System Safety. 2010;95(7):750–763. doi:10.1016/j.ress.2010.02.013.

[31] Fülöp J, Koczkodaj WW, Szarek SJ. A different perspective on a scale for pairwise comparisons. In: Transactions on computational collective intelligence I. vol. 6220 of Lecture Notes in Computer Science. Springer; 2010. p. 71–84. doi:10.1007/978-3-642-15034-0_5.

[32] Fülöp J. A method for approximating pairwise comparison matrices by consistent matrices. Journal of Global Optimization. 2008;42(3):423–442. doi:10.1007/s10898-008-9303-0.

[33] MacLennan BJ. Principles of Programming Languages: Design, Evaluation, and Implementation. Oxford University Press, Inc.; 1999. ISBN: 0-19-511306-3.

[34] Hill RJ. A note on inconsistency in paired comparison judgments. American Sociological Review. 1953;18(5):564–566. doi:10.2307/2087442.

[35] Saaty TL. On the measurement of intengibles. A principal Eigenvector approach to relative measurement derived from paired comparisons. Notices of the American Mathematical Society. 2013;60(2):192–208. doi:10.1090/noti944.

[36] e Costa CAB, Vansnick JC. A critical analysis of the eigenvalue method used to derive priorities in AHP. European Journal of Operational Research. 2008;187(3):1422–1428. doi:10.1016/j.ejor.2006.09.022.

[37] Barzilai J. Deriving weights from pairwise comparison matrices. Journal of the Operational Research Society. 1997;48(12):1226–1232. Available from: http://www.jstor.org/stable/3010752. doi:10.2307/3010752.

[38] Fichtner J. Some thoughts about the mathematics of the analytic hierarchy process. Munich, Germany: Institut für Angewandte Systemforschung u. Operations-Research; 1984. 8403.

[39] Kendall MG, Smith BB. On the method of paired comparisons. Biometrika. 1940;31(3/4):324–345.

[40] Gerard HB, Shapiro HN. Determining the degree of inconsistency in a set of paired comparisons. Psychometrika. 1958;23(1):33–46. doi:10.1007/BF02288977.

[41] Slater P. Inconsistencies in a schedule of paired comparisons. Biometrika. 1961;48(3/4):303–312. Available from: http://www.jstor.org/stable/2332752.

[42] Koczkodaj WW, Szwarc R. On Axiomatization of Inconsistency Indicators for Pairwise Comparisons. Fundamenta Informaticae. 2014;132:485–500.

[43] Koczkodaj WW. A new definition of consistency of pairwise comparisons. Mathematical and Computer Modelling. 1993;18(7):79–84.

[44] Koczkodaj WW, Szarek SJ. On distance-based inconsistency reduction algorithms for pairwise comparisons. Logic Journal of IGPL. 2010;18(6):859–869.

[45] Koczkodaj WW, Almowanes A, Kakiashvili T, Duncan G. Monte Carlo Study of the Random Image Area Estimation by Pairwise Comparisons. Interdisciplinary Topics in Applied Mathematics, Modeling and Computational Science. 2015;117:271.

[46] Genest C, Rivest LP. A statistical look at Saaty's method of estimating pairwise preferences expressed on a ratio scale. Journal of Mathematical Psychology. 1994;38(4):477–496. doi:10.1006/jmps.1994.1034.

[47] Herman MW, Koczkodaj WW. A Monte Carlo study of pairwise comparisons. Information Processing Letters. 1996;57(1):25–29. doi:10.1016/0020-0190(95)00185-9.

[48] Koczkodaj WW, Kosiek M, Szybowski J, Xu D. Fast convergence of distance-based inconsistency in pairwise comparisons. Fundamenta Informaticae (in print). 2015;137(3):355–367.

[49] Saaty TL, Vargas LG. Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios. Mathematical Modelling. 1984;5(5):309–324. doi:10.1016/0270-0255(84)90008-3.

[50] Kutbi II. A pragmatic pairwise group-decision method for selection of sites for nuclear power plants. Nuclear Engineering and Design. 1987;100(1):49 – 63. ISSN: 0029-5493. Available from: `http://www.sciencedirect.com/science/article/pii/0029549387900719`. doi:http://dx.doi.org/10.1016/0029-5493(87)90071-9.

[51] Barbara S, Soltys M. Complex ranking procedures. Fundamenta Informaticae Special Issue on Pairwise Comparisons. 2015;.

[52] Soltys M. Fair Ranking in Competitive Bidding Procurement: A Case Analysis. Procedia Computer Science. 2014;35(0):1138–1144. doi:10.1016/j.procs.2014.08.207.

[53] Koczkodaj WW, Kułakowski K, Ligęza A. On the quality evaluation of scientific entities in Poland supported by consistency-driven pairwise comparisons method. Scientometrics. 2014;99(3):911–926.

[54] Saaty TL, Tran LT. On the invalidity of fuzzifying numerical judgments in the Analytic Hierarchy Process. Mathematical and Computer Modelling. 2007;46(7):962–975. doi:10.1016/j.mcm.2007.03.022.

[55] Saaty TL. There is no mathematical validity for using fuzzy number crunching in the analytic hierarchy process. Journal of Systems Science and Systems Engineering. 2006;15(4):457–464. doi:10.1007/s11518-006-5021-7.

[56] Chang DY. Applications of the extent analysis method on fuzzy AHP. European Journal of Operational Research. 1996;95(3):649–655. doi:10.1016/0377-2217(95)00300-2.

[57] Dijkstra EW. Complexity controlled by hierarchical ordering of function and variability. In: Naur P, Randel P, editors. Software Engineering. NATO Science Committee; 1969. p. 181–185. Available from: `http://homepages.cs.ncl.ac.uk/brian.randell/NATO/nato1968.PDF`.

[58] HIPO — A Design Aid and Documentation Technique; 1974. Manual No. GC20-1851-0.

[59] Katzan H. Systems Design and Documentation: An Introduction to the HIPO Method. Computer Science Series. Van Nostrand Reinhold New York; 1976. ISBN: 10:0442242670, 13:978-0442242671.