



Article The Influence of Criteria Selection Method on Consistency of Pairwise Comparison

Vladimír Bureš *,†[®], Jiří Cabal [†], Pavel Čech [†][®], Karel Mls [†][®] and Daniela Ponce [†][®]

Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, 50003 Hradec Kralove, Czech Republic; jiri.cabal@uhk.cz (J.C.); pavel.cech@uhk.cz (P.Č.); karel.mls@uhk.cz (K.M.); daniela.ponce@uhk.cz (D.P.)

- * Correspondence: vladimir.bures@uhk.cz
- + These authors contributed equally to this work.

Received: 19 November 2020; Accepted: 8 December 2020; Published: 10 December 2020



Abstract: The more criteria a human decision involves, the more inconsistent the decision. This study experimentally examines the effect on the degree of pairwise comparison inconsistency by using the (im)possibility of selecting the criteria for the evaluation and the size of the decision-making problem. A total of 358 participants completed objective and subjective tasks. While the former was associated with one possible correct solution, there was no single correct solution for the latter. The design of the experiment enabled the acquisition of eight groups in which the degree of inconsistency was quantified using three inconsistency indices (the Consistency Index, the Consistency Ratio and the Euclidean distance) and these were analysed by the repeated measures ANOVA. The results show a significant dependence of the degree of inconsistency on the method of determining the criteria for pairwise evaluation. If participants are randomly given the criteria, then with more criteria, the overall inconsistency of the comparison decreases. If the participants can themselves choose the criteria for the comparison, then with more criteria, the overall inconsistency of the comparison decreases. For females, the dependence is the opposite, but it is not statistically significant.

Keywords: pairwise comparison; inconsistency measures; real data experiment; comparison criteria selection method; multi-criteria decision making

1. Introduction

In the realm of multi-criteria decision making, the process of selecting from options involves the ranking of a finite set of available alternatives. As a method for coping with this relatively simple task, pairwise comparisons have been the primary approach for several decades. Comparing alternatives has been a significant topic in fields of study such as cognitive science, decision sciences, psychology and computer science [1–3], and has enabled the establishment of modern multi-criteria decision-making methods such as multi-attribute value theory and the analytic hierarchy process [4]. The method involves two steps. First, pairwise comparisons of the alternatives are conducted. Second, the overall ranking is synthesised by using an appropriate algorithm [5].

The issue is that this procedure is connected with inconsistencies [6]. Consistency is "a cardinal transitivity condition of preferences on triplets of decision elements and represents the full decision-maker's coherence" [7]. Simply put, inconsistency as a concept associated with the pairwise comparisons technique is based on the idea of transitivity, i.e., establishment of two pairwise comparisons among three alternatives determines the last comparison between as yet unpaired alternatives [8]. If preferences are presented as ratios, then their consistency is based on the following idea: if an alternative a_i is preferred to an alternative $a_j x$ times and the alternative a_j is preferred to an

alternative $a_k y$ times, then the alternative a_i should be preferred to the alternative $a_k x \times y$ times [9]. When pairwise comparisons are executed, a priority vector of alternatives can be determined and utilised for the final ranking [5]. One of the first presentations of the inconsistency concept in pairwise comparisons was provided by Kendall [10] decades ago.

When working with *n* priorities, a decision-maker has to conduct a set of (n - 1) basic comparisons. Nonetheless, the issue is that this approach pushes an individual to make a direct choice of one object over another during the comparison rather than comparing all objects simultaneously [11]. A decision-maker is forced to make both a selection of one alternative over another without the possibility of comparing all alternatives at once, and set values to each pair of alternatives. Considering these assumptions, it is almost infeasible to reach an absolutely consistent priority matrix. Not surprisingly, the inconsistency in comparison of priorities also arises due to mistakes made by decision-makers [8]. There are many mistakes made by individuals who try to achieve good value judgements. They range from incomprehension of the decision context and doubts about their judgements, to the vagueness of the judgements and the omission of checking the comparisons of priority for their consistency [5,12].

In the realm of inconsistency research, there is an acknowledged anticipation that the growing number of required comparisons is related to increasing inconsistency [13–15]. The main motivation for this study is associated with the endeavour to develop more realistic assumptions. Factors influencing inconsistency are neither well nor completely understood. When inconsistency values are generated (e.g., for simulation purposes), the aforementioned anticipation presupposes that they are only linearly dependent on the problem size. Our research investigates factors that can contradict this presumption. The identified research gap reflects situations in two separate research domains in which experts cope with inconsistency. Both types of studies, theoretical and empirical, exploring properties of the inconsistency quantification methods and comparison matrices have already been published in the fields of mathematics and computer science [16–19]. Empirical comparison matrices are seldom analysed. In practice, randomly generated (simulated) comparison matrices are used almost exclusively. As far as our knowledge extends, there are only a few studies dealing with alternative comparison matrices based on empirically collected data. For a few exceptions from this practice, see for instance [20–26]. These studies are grounded either in a demonstrative experiment [27], or a regular experimental study [21]. In addition to this perspective, there are also research works based on empirical studies from the fields of cognitive science or psychology [28–32]. Nevertheless, published manuscripts are neither directed at individuals and their multi-criteria decision making peculiarities, nor do they utilise available numerical measures enabling if not cardinal, then at least ordinal comparison.

The aim of the present paper is to analyse the effects of subjective and objective problem types on the inconsistency of the decisions, as measured by selected inconsistency coefficients. The results of an experimental investigation clarify whether the size of the problem, the choice of the alternatives, or gender, moderate different problem types with regard to the level of inconsistency.

2. Inconsistency of Pairwise Comparison Matrices

The analytic hierarchy process (AHP) developed by Saaty [7] is a well-known method for pairwise comparison. It has been modified or extended in various ways from the very beginning in order to avoid its weak points [33,34]. In specific situations, such as group decision making, it is recommended to substitute AHP for more appropriate methods such as step-wise weight assessment ratio analysis [35] or Eckenrode's rating technique [36]. In order to enable the practical usability of pairwise comparison, a threshold inconsistency was determined. Comparison matrices with inconsistency values below the threshold are considered as usable for further analysis and the decision model developed by a decision-maker is considered as acceptable. Inconsistency values above the threshold indicate the need to reconstruct the model. Liang et al. [37] state that without any threshold, the decision-maker is left with significant issues of deciding whether judgements need to be revised or can be accepted.

Furthermore, the number of criteria and the scale of evaluation have to be considered, which makes the situation even more tangled. The generally accepted 10% rule of thumb associated with AHP [38] has long been criticised [14,39]. Hence, several amendments have been developed, such as values of 5 and 8% for three and four criteria, respectively [40]. Thresholds were also determined based on various statistical studies [41–44]. Although some other methods have been proposed to determine consistency thresholds [45,46], the majority of them are associated with complete pairwise comparison matrices. This feature prevents them from being used directly for incomplete pairwise comparison matrices.

The purpose of the pairwise comparison matrix is to capture partial information about all pairs of alternatives that the decision-maker compares with each other. In each comparison, the decision-maker assigns weights to alternatives expressing his/her preferences for the alternatives. However, the weights are not given directly. Instead, the decision-maker enters (estimation of) weight ratios corresponding to the alternatives being compared (if a multiplicative scale was used). Applying mathematical formalism, a matrix of pairwise comparisons is a mathematical structure in the form of a square matrix $A = (a_{ij})_{n \times n}$, where $a_{ij} > 0$ is an estimation of the weight ratio w_i to w_j ; w_i, w_j being weights for alternatives i, j, respectively. The matrix A is said to be consistent if, and only if, $a_{ik} = a_{ij}a_{jk}$ for all i, j, k. It can immediately be seen that for $a_{ij} = \frac{w_i}{w_i}$, A is consistent.

Since in general the elements of *A* are estimations of weight ratios, it is easy for this matrix to be inconsistent. There are several methods for quantifying the inconsistency contained in a pairwise comparison matrix (see e.g., [9,38]).

Quantification of inconsistency can be associated with either ordinal or cardinal comparisons. Some measures determine parameters with the help of a large set of comparison matrices generated on a random basis [15,47,48]. Methods and techniques applied to the quantification of inconsistency vary from the perspective of ease of calculation, degree of similarity to other indices, behaviour [9], or their focus on either means or extreme values [49]. Basic informally defined characteristics of inconsistency indices are provided by Brunelli [38] and are focused on (1) the most inconsistent part of the matrix (e.g., generalized K index [50]); (2) index formula as a reference to or function of the w vector (e.g., Consistency Index [7]); and (3) the existence of analytic solution (e.g., Euclidean distance). There are studies using different sets of matrices with values determined by selected criteria in order to compare and evaluate inconsistency quantification techniques [14,51,52]. A complete illustration and explanation of these indices goes beyond the focus of this study. Readers are invited to find relevant details in the original sources. It is important to note that some indices were originally tied to the consistency concept while others were associated with the measurement of inconsistency. Despite this terminological issue, the main rationale of all such indices is the same: a greater inconsistency is connected with a greater value of an index.

3. Materials and Methods

3.1. Participants and Materials

A call for participation was issued at the authors' university in order to acquire subjects for the data gathering process. Potential applicants were motivated by a reward of CZK 3000 (equivalent to EUR 120) for five randomly selected participants. Only participants who provided an email address from the university domain at the end of the trial could participate in the draw. The email addresses were stored separately without the possibility of connecting them with experimental outcomes retrospectively. Altogether, 358 subjects enrolled in the experiment, consisting of students of various study programs ranging from soft disciplines such as tourism management or financial management to hard disciplines such as applied computer science.

The acquired set of subjects represented a heterogeneous group of individuals. Therefore, the domain used for inconsistency measurement had to be thoroughly selected, as the testing domain and related task had to be comprehensible to all subjects. Eventually, a simulated decision-making

task of selecting a mobile phone (subjective problem type) and an area-based ordering of geometrical shapes (objective problem type) were presented to the subjects.

For the former, a set of 15 criteria associated with the properties of mobile phones was prepared; namely, the manufacturer, display size, resolution of the front camera, resolution of the back camera, battery capacity, the possibility of changing the battery, memory size, type of external memory card, operating system, weight, processor type, cordless battery charging, availability of original accessories, dual SIM and the resistance of the mobile phone to environmental forces.

For the other task, 7 shapes with known unequal areas were drawn: a circle, ellipse, rectangle, square, trapezoid, triangle and rhombus.

3.2. Procedure

The experiment was conducted in a dedicated computer lab located in the Faculty of Informatics and Management, University of Hradec Králové. A proprietary web-based application was developed based on the Hypertext Markup Language, Cascading Stylesheet, PHP: Hypertext Preprocessor, MySQL and JavaScript. This application assisted with the gathering, checking and saving of the data. The core of this application was focused on acquiring the input and formatting the output. The third-party public library, a product of Codeproject, was used for calculations of the eigenvalues. The measurement of the time was another functionality used for identifying unreliable data segments.

At the beginning of the experiment, the subjects were informed about the main objective and purpose of the project and an introductory explanation of the pairwise comparison was presented. The evaluation process was not associated with any time limitation. All subjects were allowed to participate in the study only once. A task assigned to a participant was based on a random selection of a combination of testing modes.

There were three tested properties: the number of evaluated criteria (a matrix size of either five or seven), objective or subjective problem type (i.e., unique correct alternative ordering exists or does not), and the possibility of the subjects selecting the criteria for the comparison from a predefined set (free criteria choice method), or a random assignment of the criteria.

Each participant evaluated two comparison matrices only: one for the objective problem type and another one for the subjective problem. Because the two problems are completely different, all findings came from a "between-subject" experiment (see [53] for details), thereby avoiding the anchoring effect [54].

Despite its weak points, AHP was found to be a sufficient and good enough technique for the purpose of this study as it can be found on the list of the most popular tools implemented in practical situations [55,56]. That is why it has been implemented with a multiplicative preference evaluation model (entry values being the integer numbers 1 to 9 and their corresponding inverse values). The subjects were given all the cells of the comparison matrix at once (as opposed to cell-by-cell) with permission to revise once selected pairwise comparison values before the final submission of the matrix.

3.3. Applied Measures

As they have been long dominant in the field and are most widely used for measuring the inconsistency degree [38,56,57], we decided to apply three fundamental inconsistency quantification methods for the purpose of our study. First, Saaty's Consistency Index and the the Consistency Ratio were calculated. Second, the Euclidean distance was calculated.

Let $A = (a_{ij})$ denote a multiplicative pairwise comparison matrix of dimension *n*, and let *B* be a matrix defined as $B = (\ln a_{ij})$.

The Consistency Index (CIndex) was defined by Saaty [7] as

$$\operatorname{CIndex}(A) = \frac{\lambda_{max} - n}{n - 1} \tag{1}$$

where λ_{max} is the principal eigenvalue of *A*; CIndex \geq 0.

The Consistency Ratio (CRatio) represents a standardised version of CIndex. It is expressed as a ratio with CRatio in a numerator and *RI* in a denominator. *RI* is a real number determined as the average CIndex of a large amount of randomly generated matrices of size *n*.

$$CRatio(A) = \frac{CIndex(A)}{RI}$$
(2)

The Euclidean distance (EDA) is defined as

$$EDA(A) = \sqrt{\sum_{i,j \in N} (b_{ij} - v_{ij})^2}$$
(3)

where *V* is a priority matrix $V = (v_{ij}) = (w_i - w_j)$ such that w_i is an arithmetic mean weight vector $w_i = \frac{1}{n} \sum_{i \in N} b_{ij}$. The Euclidean distance can be normalised, yielding the Euclidean normalised distance.

3.4. Statistical Analysis

The data were pre-processed using the R statistical package. The main analyses were accomplished in IBM SPSS. First, the incomplete and corrupted records with inconsistency zero were removed and a dataset with 358 subjects was obtained (dataset is available at https://doi.org/10.6084/m9.figshare. 13317458). Further, the data were cleaned in respect of the time the subjects spent filling in the form. An exceedingly long time for filling the form implied that the subject declined to complete the task. The detection of outliers was based on Tukey's interquartile range, since Tukey's technique is less sensitive to extreme values [58]. The filtered dataset contained inconsistency measurements of 276 subjects. The associations between the studied factors were tested with the Pearson chi-squared indicator. Repeated measures ANOVA was conducted to study the inconsistency of decisions in respect of given conditions. Repeated measures ANOVA controls were used for the same subjects participating in more than one condition. The design of the analyses included the type of the problem, such as within-subject factor and size of the problem, choice of criteria and gender as between subject factors. The analyses were executed incrementally by adding the above stated factors. The key assumption for the repeated measures ANOVA is so called sphericity; that is, the equality of variances computed based on differences between factor levels. However, in the present study, each factor consists of only two levels and thus sphericity was not an issue. The post hoc tests involved pairwise assessments of estimated marginal means between experimental conditions and the mean differences (M) and corresponding 95% confidence intervals (CI). In the present study, the results of the statistical tests were considered significant if the p-value < 0.05. The partial eta squared statistic was used to measure the size of the effect. The interpretation of the partial eta squared is the amount of variance explained by the independent variable. Cohen et al. [59] state that the indicative effect sizes can be small, medium or large with values 0.01, 0.06 or 0.14, respectively.

4. Results

The inconsistency of the decisions resulting from the comparison tasks of 276 subjects were analysed. The basic characteristics of the subjects are presented in Table 1. The frequencies across studied factor levels are shown in Table 2. The Pearson chi-squared tests confirmed that there were no associations between size and choice χ (1) = 0.012, p = 0.912, size and gender χ (1) = 0.021, p = 0.884 or choice and gender χ (1) = 1.542, p = 0.214. That means that there are no significant differences between frequencies across factor levels.

Attribute	Min	Max	Mean	Std. dev.
Age	19	30	21.12	1.716
Year of study	1	4	1.96	0.955

Table 1. Descriptive statistics.

Size	Choice	Female	Male	Total
5	Random	23	46	69
	Selected	28	45	73
	Total	51	91	142
7	Random	20	46	66
	Selected	27	41	68
	Total	47	87	134

4.1. The Effect of Problem Type

Repeated measures ANOVA with problem type as within-subject factor revealed that there is a significant effect of problem type on the inconsistency of comparison in all inconsistency coefficients (see Table 3). The table shows that the significance is p < 0.001 for all coefficients and also the effect sizes measured by the partial eta squared point to large effects. The pairwise comparisons with estimated marginal means show that students exhibit significantly higher levels of inconsistency in assessing alternatives of the subjective problem as opposed to the objective one (see Table 4).

Figure 1 shows the estimated marginal means of inconsistency indices scaled on the max–min range. The error bars are based on the 95% confidence interval.



The effect of the problem type

Figure 1. Estimated marginal means of scaled inconsistencies for selected inconsistency coefficients across different problem types. The error bars represent 95% confidence intervals. Scaling is based on the max–min range.

DF_M	DF_E	F	Sig	Partial Eta Squared
1	275	75.924	0.000	0.216
1	275	69.962	0.000	0.203
1	275	100.598	0.000	0.268
	DF_M 1 1 1	DF_M DF_E 1 275 1 275 1 275 1 275	DF_MDF_EF127575.924127569.9621275100.598	DF_M DF_E F Sig 1 275 75.924 0.000 1 275 69.962 0.000 1 275 100.598 0.000

Table 3. The results of repeated measures ANOVA for selected inconsistency coefficients with problem type as within-subject factor.

Table 4. Estimated marginal means of inconsistency coefficients for objective and subjective problem type with standard error and 95% confidence intervals.

	Objective				Subjective			
Coefficient	Mean	Std. Error	LB	UB	Mean	Std. Error	LB	UB
CIndex	0.140926	0.013101	0.115134	0.166718	0.296861	0.015307	0.266727	0.326994
CRatio	0.116411	0.011402	0.093966	0.138857	0.244548	0.012972	0.219011	0.270085
EDA	0.784867	0.057526	0.671619	0.898114	1.549933	0.064357	1.423238	1.676627

4.2. The Effect of Size and Choice of Criteria

Repeated measures ANOVA reapplied with between-subject factors revealed that there is a significant interaction in the problem size and the choice of criteria. The interaction is significant in all indices, however, only the results of CIndex are reported to avoid redundancy. The results indicate that inconsistency differs depending on the choice criteria and size. The pairwise comparison of estimated marginal means with Bonferoni adjustment (see Figure 2) revealed that in case of the problem size, 7 of the subjects achieved significantly lower levels of inconsistency if employing randomly assigned criteria as opposed to subjects who could select their own criteria (mean difference $M = 0.07 \pm 0.03, 95\%$ [0.01, 0.13], p = 0.032). This contrasts with problem size 5 where the opposite pattern is exhibited; however, the difference is only with reduced significance $M = 0.06 \pm 0.03, 95\%$ [0, 0.12], p = 0.062. Viewed from the perspective across sizes, the subjects who could select their own criteria achieved significantly lower inconsistency in problem size 5 compared to problem size 7. The subjects with randomly given criteria demonstrated a reversed trend that is, however, not significant ($M = 0.05 \pm 0.03, 95\%$ [-0.02, 0.11], p = 0.138). Tables 5 and 6 summarise the within-and between-subject effects for size and choice.

Table 5. Tests of within-subject effects for type of problem in respect of size and choice.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Туре	3.383	1	3.383	76.741	0.000	0.220
Type * Size	0.019	1	0.019	0.439	0.508	0.002
Type * Choice	0.010	1	0.010	0.233	0.630	0.001
Type * Choice * Size	0.127	1	0.127	2.889	0.090	0.011
Error (Type)	11.992	272	0.044			

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Size	0.035	1	0.035	0.529	0.468	0.002
Choice	0.004	1	0.004	0.055	0.814	0.000
Choice * Size	0.541	1	0.541	8.134	0.005	0.029
Error	18.074	272	0.066			



The effect of size and choice



Figure 2. Estimated marginal means of inconsistency as measured by the CIndex coefficient across sizes and choice of criteria. The error bars represent 95% confidence intervals.

4.3. The Effects of Size and Choice of Criteria including Gender

In order to further study the interaction between size and choice of criteria, the analyses have been extended to include gender as another between-subject factor. The results indicate that there is a significant interaction between size, choice of criteria and gender F(1,268) = 5.955, p = 0.015, partial eta squared = 0.022 (small effect). That means that the interaction between size and choice of criteria varies in respect of gender. The subsequent pairwise comparison with Bonferoni adjustment revealed that in the case of a random choice of criteria, males achieved significantly lower inconsistency in problem size 7 compared to problem size 5 (mean difference M = 0.103 ± 0.037, 95% [0.03, 0.177], p = 0.006). The subjects with selected criteria have higher inconsistency in size 7 compared to size 5 (mean difference $M = -0.096 \pm 0.039$, 95% [-0.172, -0.019]). Females, on the other hand, have an increasing pattern between sizes 5 and 7 in both random and selected choice of criteria that was, however, not significant (see Figure 3). Tables 7 and 8 summarise the within- and between-subject effects for size and choice.

Table 7. Tests of within-subject effects for type of problem in respect of size, choice and gender.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Туре	2.836	1	2.836	64.777	0.000	0.195
Type * Size	0.048	1	0.048	1.096	0.296	0.004
Type * Choice	0.020	1	0.020	0.456	0.500	0.002
Type * Gender	0.093	1	0.093	2.133	0.145	0.008
Type * Size * Choice	0.063	1	0.063	1.448	0.230	0.005
Type * Size * Gender	0.051	1	0.051	1.159	0.283	0.004
Type * Choice * Gender	0.042	1	0.042	0.957	0.329	0.004
Type * Size * Choice * Gender	0.069	1	0.069	1.580	0.210	0.006
Error	11.733	268	0.044			

Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
0.102	1	0.102	1.589	0.209	0.006
0.028	1	0.028	0.434	0.511	0.002
0.049	1	0.049	0.764	0.383	0.003
0.242	1	0.242	3.757	0.054	0.014
0.131	1	0.131	2.035	0.155	0.008
0.226	1	0.226	3.505	0.062	0.013
0.384	1	0.384	5.955	0.015	0.022
17.278	268	0.064			
	O.102 0.028 0.049 0.242 0.131 0.226 0.384 17.278	Type III Sum of Squaresdf0.10210.02810.04910.24210.13110.22610.384117.278268	Type III Sum of SquaresdfMean Squares0.10210.1020.02810.0280.04910.0490.24210.2420.13110.1310.22610.2260.38410.38417.2782680.064	Type III Sum of SquaresdfMean SquareF0.10210.1021.5890.02810.0280.4340.04910.0490.7640.24210.2423.7570.13110.1312.0350.22610.2263.5050.38410.3845.95517.2782680.064	Type III Sum of SquaresdfMean SquareFSig.0.10210.1021.5890.2090.02810.0280.4340.5110.04910.0490.7640.3830.24210.2423.7570.0540.13110.1312.0350.1550.22610.2263.5050.0620.38410.3845.9550.01517.2782680.0641

Table 8. Tests of between-subject effects of size, choice and gender.



Figure 3. Estimated marginal means of inconsistency as measured by the CIndex coefficient. The figure shows different interactions of size and choice of criteria depending on gender. The error bars represent 95% confidence intervals.

5. Discussion

In this study, we investigated inconsistency from the perspective of decision science in general and multi-criteria decision making in particular. We consider this a topic that needs to be studied and dealt with as a decision-making task with multiple attributes requiring quantitative analysis and formal representation. In the controlled experiment, inconsistency in decision making of participants was derived from empirical data.

5.1. Contribution

The experiment provides several findings.

The problem type (subjective vs. objective) has a significant effect on the inconsistency of the decision making. This result (higher inconsistency level for subjective decision-making problems) confirms the findings reported in [21]. This effect is independent of the number of criteria as well as gender. Interpretation can be based on an argument that there exists one perfectly consistent objective solution that can be reached or one can get close to. Moreover, finding this single solution has a cognitive base. In contrast to the objective problem, there may not exist a perfectly consistent subjective solution, which is probably based mostly on preferences and attitudes. As humans are capable of bearing mismatching preferences and attitudes, it may be more challenging to decrease the overall inconsistency of decision making, which involves preferences and attitudes. Thus, in case of the subjective task, reaching a good enough solution can be considered as sufficient and acceptable.

A novel element of this work is the exploration of the influence of free/unfree criteria choice on the inconsistency of the decision making. This impact has not been previously detected.

The method of choosing the criteria (free selection vs. random assignment) was shown to have an impact on the inconsistency. Although not significantly, female decisions were more inconsistent for larger criteria sets, regardless of whether the criteria choice method was free selection or random assignment. At this time, we can only make suppositions as to why the free criteria selection leads to higher inconsistency for a bigger set of criteria. Probably the respondents preferred to select criteria that were (positively) important to them and subsequently it was difficult to order them on the scale available. If that was the case, the ordering task became more difficult as the size of the set of criteria increased.

For males, the inconsistency in decisions increased with the size of the problem when the criteria were freely selected, and it decreased with the size of the problem when the criteria were randomly assigned. This increasing/decreasing difference is statistically significant. Hence, the mode of choosing the criteria has been identified as another factor that has an impact on the inconsistency of the decisions (other factors include, e.g., explanation of inconsistency [60] and information sharing in a group [61]).

The data set attached to this paper can be considered as an additional contribution. Since few experimental studies in inconsistency research exist, real decisions of nearly 280 participants from a controlled experiment give an open opportunity for further real decision analysis.

5.2. Limitations of Our Work

The results of our work are valid under the settings of the experiment. From the point of view of basic descriptive indicators (gender, age), the analysed sample of subjects is considered representative with respect to the defined population. However, experimental studies of university students are difficult to generalise. For generalisation purposes, experiments with distinct target groups have to be designed and performed.

It is also necessary to mention that although the participants were asked to make decisions in a domain presumably familiar to them, the level of their expertise in this particular domain remained unknown.

A final limitation is the entry type used in comparison matrices. As was recently shown [22], the resulting inconsistency is biased by the preference evaluation model.

5.3. Future Work

The experimental procedure was defined in such a way that it could be replicated by other authors. Further research may focus on different dimensions of matrices, a different type of subjective task or a larger sample of females.

Furthermore, this study was based on the calculation and comparison of three selected measures. Although the acquired results are consistent across all of the studied coefficients, other available measures can be applied in order to verify achieved results. As Brunelli [38] found out, there is a kernel of measures within the set of indices with a high degree of similarity. Thus, this verification would be interesting with calculation of outliers (e.g., Singular Value Decomposition [62], Kulakowski's E index [63] or Barzilai's RE index [64]) as consistent comparisons of kernel measures can be anticipated.

Inconsistency in female decisions does not allow statistically sound conclusions. It is worth noting that for females, the free criteria selection seems to have a different effect on the inconsistency of the decisions, when comparing smaller and larger criteria sets. An experiment performed with a larger female cohort is needed to confirm or to disprove this effect.

As inconsistency is influenced by the preference evaluation model [22], a combined effect of the preference evaluation model and criteria selection method may be worth exploring.

Author Contributions: Conceptualisation, V.B., J.C. and D.P.; methodology, D.P. and P.Č.; software, J.Č.; validation, D.P. and K.M.; statistical analysis, P.Č.; resources, V.B.; data curation, P.Č.; writing—original draft preparation, V.B., J.C., P.Č., K.M. and D.P.; project administration, V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful for the support from the grant project GAČR Project No. 18-01246S and the FIM UHK Specific project 2107 "Computer Networks for Cloud, Distributed Computing, and Internet of Things III". Special thanks go to Hana Švecová for her assistance during the data acquisition stage.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Ma, L.C.; Li, H.L. Using Gower Plots and Decision Balls to rank alternatives involving inconsistent preferences. *Decis. Support Syst.* **2011**, *51*, 712–719. [CrossRef]
- Johnson-Laird, P.N.; Girotto, V.; Legrenzi, P. Reasoning from inconsistency to consistency. *Psychol. Rev.* 2004, 111, 640. [CrossRef] [PubMed]
- 3. Agell, N.; Sánchez, M.; Prats, F.; Roselló, L. Ranking multi-attribute alternatives on The basis of linguistic labels in group decisions. *Inf. Sci.* **2012**, *209*, 49–60. [CrossRef]
- Brunelli, M.; Cavallo, B. Distance-based measures of incoherence for pairwise comparisons. *Knowl. Based Syst.* 2020, 187, 104808. [CrossRef]
- 5. Liu, F.; Yu, Q.; Pedrycz, W.; Zhang, W.G. A group decision making model based on an inconsistency index of interval multiplicative reciprocal matrices. *Knowl. Based Syst.* **2018**, *145*, 67–76. [CrossRef]
- 6. Csató, L. A characterization of The logarithmic least squares method. *Eur. J. Oper. Res.* **2019**, 276, 212–216. [CrossRef]
- Saaty, T.L. A scaling method for priorities in hierarchical structures. J. Math. Psychol. 1977, 15, 234–281. [CrossRef]
- 8. Kułakowski, K. Inconsistency in The ordinal pairwise comparisons method with and without ties. *Eur. J. Oper. Res.* **2018**, 270, 314–327. [CrossRef]
- 9. Brunelli, M.; Canal, L.; Fedrizzi, M. Inconsistency indices for pairwise comparison matrices: A numerical study. *Ann. Oper. Res.* 2013, 211, 493–509. [CrossRef]
- 10. Kendall, M.G.; Smith, B.B. On The method of paired comparisons. *Biometrika* 1940, 31, 324–345. [CrossRef]
- 11. Cook, W.D.; Golany, B.; Kress, M.; Penn, M.; Raviv, T. Optimal allocation of proposals to reviewers to facilitate effective ranking. *Manag. Sci.* 2005, *51*, 655–661. [CrossRef]
- 12. Keeney, R.L. Common mistakes in making value trade-offs. Oper. Res. 2002, 50, 935–945. [CrossRef]
- 13. Miller, G.A. The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **1956**, *63*, 91–97. [CrossRef]
- 14. Bozóki, S.; Rapcsák, T. On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices. *J. Glob. Optim.* **2008**, *42*, 157–175. [CrossRef]
- 15. Saaty, T.L. *Multicriteria Decision Making: Planning, Priority Setting, Resource Allocation;* RWS Publications: Pittsburgh, PA, USA, 1996.
- Chu, M.T. On the optimal consistent approximation to pairwise comparison matrices. *Linear Algebra Appl.* 1998, 272, 155–168. [CrossRef]
- 17. Leung, L.C.; Cao, D. On consistency and ranking of alternatives in fuzzy AHP. *Eur. J. Oper. Res.* 2000, 124, 102–113. [CrossRef]
- 18. Ramík, J.; Korviny, P. Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets Syst.* **2010**, *161*, 1604–1613. [CrossRef]
- 19. Szybowski, J. The Cycle Inconsistency Index in Pairwise Comparisons Matrices. *Procedia Comput. Sci.* **2016**, *96*, 879–886. [CrossRef]
- 20. Bureš, V.; Ponce, D.; Čech, P.; Mls, K. The effect of trial repetition and problem size on The consistency of decision making. *PLoS ONE* **2019**, *14*, e0216235. [CrossRef]
- 21. Bozóki, S.; Dezső, L.; Poesz, A.; Temesi, J. Analysis of pairwise comparison matrices: An empirical research. *Ann. Oper. Res.* **2013**, *211*, 511–528. [CrossRef]
- 22. Cavallo, B.; Ishizaka, A.; Olivieri, M.G.; Squillante, M. Comparing inconsistency of pairwise comparison matrices depending on entries. *J. Oper. Res. Soc.* **2019**, *70*, 842–850. [CrossRef]
- 23. Mazurek, J.; Perzina, R. On The inconsistency of pairwise comparisons: An experimental study. *Sci. Pap. Univ. Pardubic. Ser. D Fac. Econ. Adm.* **2017**, 24, 102–109.
- 24. Moslem, S.; Alkharabsheh, A.; Ismael, K.; Duleba, S. An integrated decision support model for evaluating public transport quality. *Appl. Sci.* **2020**, *10*, 4158. [CrossRef]
- 25. Dehraj, P.; Sharma, A. An empirical assessment of autonomicity for autonomic query optimizers using fuzzy-AHP technique. *Appl. Soft Comput.* **2020**, *90*, 106137. [CrossRef]

- 26. Gluszak, M.; Gawlik, R.; Zieba, M. Smart and Green Buildings Features in The Decision-Making Hierarchy of Office Space Tenants: An Analytic Hierarchy Process Study. *Adm. Sci.* **2019**, *9*, 52. [CrossRef]
- 27. Saaty, T.L.; Ozdemir, M.S. Why The magic number seven plus or minus two. *Math. Comput. Model.* 2003, *38*, 233–244. [CrossRef]
- Fischer, P.; Reinweber, M.; Vogrincic, C.; Schaefer, A.; Schienle, A.; Volberg, G. Neural mechanisms of selective exposure: An EEG study on the processing of decision-consistent and inconsistent information. *Int. J. Psychophysiol.* 2013, *87*, 13–18. [CrossRef]
- 29. Nikolova, N.L.; Hendry, M.C.; Douglas, K.S.; Edens, J.F.; Lilienfeld, S.O. The Inconsistency of Inconsistency Scales: A Comparison of Two Widely Used Measures. *Behav. Sci. Law* **2012**, *30*, 16–27. [CrossRef]
- Schreiber, P.; Weber, M. Time inconsistent preferences and The annuitization decision. *J. Econ. Behav. Organ.* 2016, 129, 37–55. [CrossRef]
- 31. Dummel, S.; Rummel, J. Take-the-best and The influence of decision-inconsistent attributes on decision confidence and choices in memory-based decisions. *Memory* **2016**, *24*, 1435–1443. [CrossRef]
- 32. Eguchi, M.; Yamashita, T. How Affinity Influences On Decision Making Inconsistent with WOM Trustworthiness And EWOM Trustworthiness. *Psychologia* **2016**, *59*, 19–37. [CrossRef]
- 33. Belton, V.; Gear, T. On a shortcoming of Saaty's method of analytic hierarchies. *Omega* **1983**, *11*, 228–230. [CrossRef]
- 34. Schoner, B.; Wedley, W. Ambiguous criteria weights in AHP: Consequences and solutions. *Decis. Sci.* **1989**, 20, 462–475. [CrossRef]
- 35. Kersuliene, V.; Zavadskas, E.; Turskis, Z. Selection of rational dispute resolution method by applying new step-wise weight assessment ratio analysis (SWARA). *J. Bus. Econ. Manag.* **2010**, *11*, 243–258. [CrossRef]
- 36. Turskis, Z.; Dzitac, S.; Stankiuviene, A.; Sukys, R. A Fuzzy Group Decision-making Model for Determining The Most Influential Persons in The Sustainable Prevention of Accidents in the Construction SMEs. *Int. J. Comput. Control* **2019**, *14*, 90–106. [CrossRef]
- 37. Liang, F.; Brunelli, M.; Rezaei, J. Consistency issues in The best worst method: Measurements and thresholds. *Omega Int. J. Manag. Sci.* 2020, *96*, 102175. [CrossRef]
- 38. Brunelli, M. A survey of inconsistency indices for pairwise comparisons. *Int. J. Gen. Syst.* **2018**, 47, 751–771. [CrossRef]
- 39. Bozóki, S.; Fülöp, J.; Poesz, A. On reducing inconsistency of pairwise comparison matrices below an acceptance threshold. *Cent. Eur. J. Oper. Res.* **2015**, *23*, 849–866. [CrossRef]
- 40. Saaty, T. Fundamentals of Decision Making and Priority Theory; RWS Publications: Pittsburgh, PA, USA, 2000.
- 41. Lin, C.; Gang, K.; Daji, E. An Improved Statistical Approach for Consistency Test in AHP. *Ann. Oper. Res.* **2013**, *211*, 289–299. [CrossRef]
- 42. Lin, C.; Gang, K.; Daji, E. A Statistical Approach to Measure The Consistency Level of the Pairwise Comparison Matrix. *J. Oper. Res. Soc.* **2014**, *65*, 1380–1386. [CrossRef]
- 43. Siraj, S.; Mikhailov, L.; Keane, J. Contribution of individual judgments toward inconsistency in pairwise comparisons. *Eur. J. Oper. Res.* 2015, 242, 557–567. [CrossRef]
- 44. Vargas, L. The consistency index in reciprocal matrices: Comparison of deterministic and statistical approaches. *Eur. J. Oper. Res.* 2008, 191, 454–463. [CrossRef]
- 45. Amenta, P.; Lucadamo, A.; Marcarelli, G. On The transitivity and consistency approximated thresholds of some consistency indices for pairwise comparison matrices. *Inf. Sci.* **2014**, *507*, 274–287. [CrossRef]
- Aguaron, J.; Moreno-Jimenez, J. The geometric consistency index: Approximated thresholds. *Eur. J. Oper. Res.* 2003, 147, 137–145. [CrossRef]
- 47. Vargas, L.G. Reciprocal matrices with random coefficients. Math. Model. 1982, 3, 69-81. [CrossRef]
- 48. Alonso, J.A.; Lamata, M.T. Consistency in the analytic hierarchy process: A new approach. *Int. J. Uncertain. Fuzz.* **2006**, *14*, 445–459. [CrossRef]
- 49. Mazurek, J. Some notes on The properties of inconsistency indices in pairwise comparisons. *Oper. Res. Decis.* **2018**, *28*, 27–42.
- 50. Duszak, Z.; Koczkodaj, W. Generalization of a new definition of consistency for pairwise comparisons. *Inf. Process. Lett.* **1994**, *52*, 273–276. [CrossRef]
- 51. Choo, E.U.; Wedley, W.C. A common framework for deriving preference values from pairwise comparison matrices. *Comput. Oper. Res.* 2004, *31*, 893–908. [CrossRef]

- 52. Lin, C.C. A revised framework for deriving preference values from pairwise comparison matrices. *Eur. J. Oper. Res.* **2007**, 176, 1145–1150. [CrossRef]
- 53. Charness, G.; Gneezy, U.; Kuhn, M.A. Experimental methods: Between-subject and within-subject design. *J. Econ. Behav. Organ.* **2012**, *81*, 1–8. [CrossRef]
- 54. Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **1974**, *185*, 1124–1131. [CrossRef]
- 55. Kubler, S.; Derigent, W.; Voisin, A.; Robert, J.; Le Traon, Y.; Viedma, E. Measuring inconsistency and deriving priorities from fuzzy pairwise comparison matrices using The knowledge-based consistency index. *Knowl. Based Syst.* **2018**, *162*, 147–160. [CrossRef]
- 56. Aguaron, J.; Escobar, M.; Moreno-Jimenez, J. Reducing inconsistency measured by The geometric consistency index in The analytic hierarchy process. *Eur. J. Oper. Res.* **2021**, *288*, 576–583. [CrossRef]
- 57. Fang, L.; Zou, C.; Li, Q. Deriving priorities from pairwise comparison matrices with a novel consistency index. *Appl. Math. Comput.* **2020**, *374*, 125059.
- 58. Tukey, J.W. Exploratory Data Analysis; Pearson: Reading, MA, USA, 1977; Volume 2.
- 59. Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*; Routledge: Abingdon, UK, 2013.
- 60. Khemlani, S.S.; Johnson-Laird, P.N. Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychol.* **2012**, *139*, 486–491. [CrossRef]
- 61. Mojzisch, A.; Schulz-Hardt, S.; Kerschreiter, R.; Brodbeck, F.C.; Frey, D. Social validation in group decision-making: Differential effects on The decisional impact of preference-consistent and preference-inconsistent information. *J. Exp. Soc. Psychol.* **2008**, *44*, 1477–1490. [CrossRef]
- 62. Gass, S.; Rapcsak, T. Singular Value Decomposition in AHP. Eur. J. Oper. Res. 2004, 154, 573–584. [CrossRef]
- 63. Kulakowski, K. Notes on Order Preservation and Consistency in AHP. *Eur. J. Oper. Res.* 2015, 245, 333–337. [CrossRef]
- 64. Barzilai, J. Consistency measures for pairwise comparison matrices. J. Mult. Crit. Dec. Anal. 1998, 7, 123–132. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).