

# A user interface for consistent AHP pairwise comparisons

Andrés Cimadamore<sup>[0000-0002-6995-7981]</sup>, Alejandro Fernandez<sup>[0000-0002-7968-6871]</sup>  
LIFIA, CICPBA, FI, Universidad Nacional de La Plata, La Plata, Argentina  
Calle 20 y 120, (1900) La Plata, Argentina  
[andres.cimadamore,alejandro.fernandez]@lifia.info.unlp.edu.ar  
web-page: <http://www.lifia.info.unlp.edu.ar/>

Chenhui Ye<sup>[0000-0001-7258-6651]</sup>, Pascale Zarate<sup>[0000-0002-5188-1616]</sup> (corresponding author)  
IRIT, Toulouse Université  
2 rue du Doyen Gabriel Marty, 31042 Toulouse Cedex 9, France  
Pascale.Zarate@irit.fr  
web-page: <https://www.irit.fr/~Pascale.Zarate/>

Daouda Kamissoko<sup>[0000-0003-0079-0885]</sup>  
Ecole des Mines d'Albi  
[Dadouda.Kamissoko@mines-albi.fr](mailto:Dadouda.Kamissoko@mines-albi.fr)

**ABSTRACT:** Decision Makers generally reason on several criteria, aiming to obtain a total consistency or partial order of several alternatives. MultiCriteria analysis is based on the assumption that such ordering exists. Decision Makers are supported by several kinds of approaches or tools. One approach consists in comparing the criteria two by two, i.e. pairwise comparison, to find the relative importance of each criterion. This relative importance, called weight of criteria, is used to find the final order of alternatives. One methodology, developed by Saaty, called Analytical Hierarchical Process (AHP), is based on this principle of pairwise comparison. Having the weights of criteria, the decision makers have then to compare the alternatives two by two for each criterion. Pairwise comparisons are simple to use; however, as the number of items to compare increases, so do the effort of conducting all comparisons and the probability of introducing inconsistencies. In this article we present an innovative approach to conduct pairwise comparisons based on a UI widget that resembles an interactive data plot. It uses the transitivity property of a consistent comparison matrix to infer comparisons. Our hypothesis is that this new approach is more efficient (as it reduces the number of actions the user must conduct to compare all items), more effective (as it limits the sources of inconsistencies), and yields better user satisfaction. We conducted a controlled experiment involving 50 participants. We observed that the proposed widget reduces the effort of making pairwise comparisons, improves the consistency of the comparisons, and leads to a better user experience.

**Keywords:** Pairwise comparisons, Consistency, Transitivity, UI design, Usability, AHP

## 1. Introduction

The AHP method is an analytical approach for supporting decision making following a multi-criteria approach [1]. It has been used in several areas, such as transport planning, rationing of energy, risk management projects, benchmarking of logistics operations, management of quality of services in hospitals, operations management, allocation resources for product portfolio management. It was developed by Thomas Saaty in 1970 and allows the decomposition of a complex problem in a hierarchical system. Alternatives defined by the decision maker provide their relative priorities thanks to a pairwise comparison. Then a synthesis allows decision makers to easily understand what would be the best choice. Classification is performed at several levels which are associated with different criteria. Thus, it is possible to determine the most appropriate alternative, depending on the priority given to each used criteria. Pairwise comparisons (PCs) are a central feature of AHP.

In this article we present an approach to conduct PCs that is easy to use, intuitive, reduces the number of required comparisons, and yields consistent and complete comparison matrices. A visual 2D representation of the comparable items is used to express relative preferences among items. The transitivity property of the AHP matrix is used to infer preferences thus reducing the number of required comparisons. As a result, the method yields more consistent matrices regardless of the number of alternatives considered.

Next, we motivate our work by presenting an overview of key concepts regarding consistency and transitivity in AHP comparisons, and by discussing the role of visualizations. Then, we present our approach based on an innovative pairwise comparison widget. Following, we present the methodology used to evaluate the approach and the results we obtained. To conclude, we offer conclusions and discuss future work.

## 2. Background: Pairwise Comparisons in AHP

Following the construction of the hierarchical model with various levels of criteria and one level of alternatives, PCs are carried out at each level. Different scales can be used to compare items [2]. In this work we focus on the original scale proposed by Saaty, using integer values in [1,9], and their reciprocals. The decision makers' judgments are kept in a matrix model called the Judgments Matrix. The main objective is to compare the relative importance of all elements belonging to the same level.

Transitivity in multiple criteria decision making is also called ordinal consistency [3]. If a decision maker prefers alternative  $x_1$  to alternative  $x_2$  and  $x_2$  to  $x_3$ , then transitivity requires that he/she also prefers  $x_1$  to  $x_3$ , as otherwise, cycles would exist in the preferences. Tversky [4] considered transitivity to be the cornerstone of normative decision theory. Preference transitivity is a basic principle in most major rational, descriptive decision models [5].

Benitez et al. [6] propose a method to achieve consistency in AHP through optimisation. This method has the major advantage of depending on just the decision variables – the number of compared elements – and so is less computationally expensive than other optimisation methods, and can be easily implemented in virtually any existing computer environment.

Decision support software packages such as Super Decisions [7] and Expert Choice [8] offer alternative modes to elicit user preferences as PCs. A frequent strategy to elicit comparisons is to present them in a matrix. Each value in the cell compares the item represented by the row, to the item represented by the column. Entering a value in a cell, automatically updates the value if its inverse. This method requires users to get accustomed to the direction of the comparisons, and the interpretation of the values (which are both integers and fractions). Super Decisions improves the matrix view by removing the values in the diagonal, keeping only one value for each pair (i.e., removing the inverse comparison), and introducing an arrow that indicates the direction of the comparison (see Figure 2).

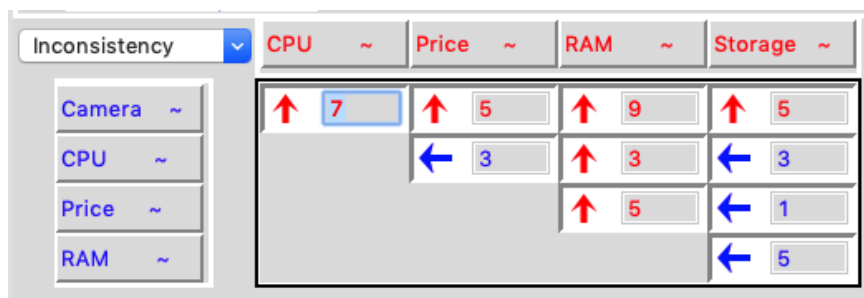


Figure 2: Matrix view in Super Decisions V3.2

The questionnaire view is another common presentation to elicit PCs. Each row represents one pairwise

comparison, with the items to compare on each side. The user must place a mark closest to the item that is considered more important (or preferred). Super Decisions uses radio buttons as markers in its questionnaire view (see Figure 3), while PriEst [9] uses sliders (in PriEst, this view is called equalizer). Placing the marker in the middle indicates that items are equally important. The number of positions between items normally reflect values from 2 to 9 in each direction, plus 1 in the middle position.

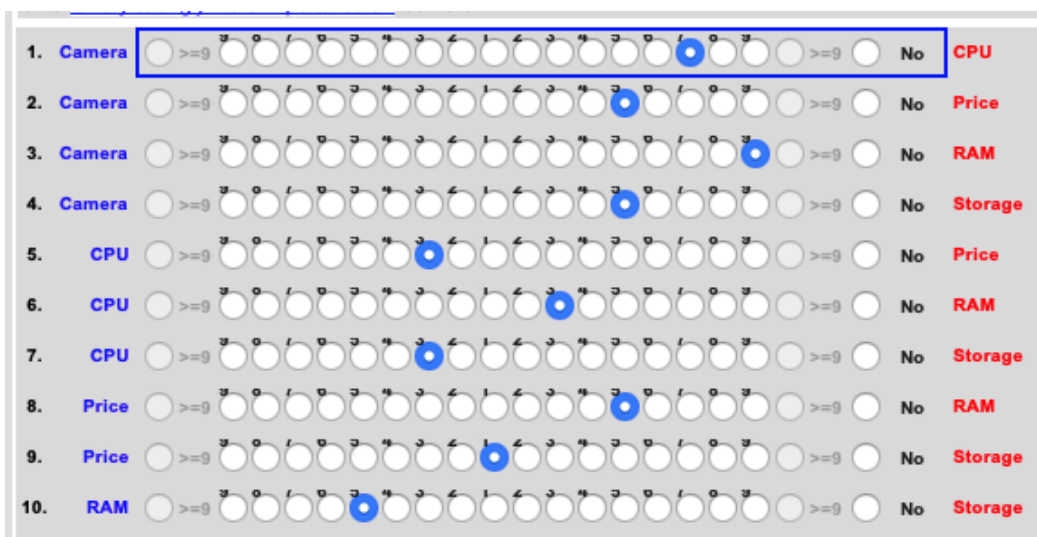


Fig 3: Questionnaire view in Super Decisions V3.2

Both, the matrix presentation and the questionnaire presentation, offer a holistic view of all comparisons. In addition to these holistic presentations, some tools offer visual means to manipulate individual comparisons, for example in the form of an interactive bar or pie chart.

All the aforementioned strategies consider that each pairwise comparison is independent from the rest of them. Independence among comparisons directly correlates to consistency; the more liberty (and the less scaffolding) users have to independently compare items the more likely they are to introduce inconsistency. Super Decisions and PriEsT offer help to identify inconsistency. In addition, PriEsT offers visual aids to observe transitivity.

How to present and to elicit PCs is one of the challenges faced by decision support tools designers; modelling preferences is almost as important as the modelling of the logical structure of the problem [10]. Abel et al. [11] compared the usability of two contrasting approaches to elicit decision priorities namely, PCs and constrained optimizations. Their work focuses on performance and usability as perceived by the user. The authors observed that PCs outperformed constraint optimizations for both efficiency and efficacy. There was little, if any, difference in terms of perceived usability. Millet [12] compared five preference elicitation models in terms of efficacy and ease of use. The results of Millet's research supports the motivation of this work to explore alternative graphical modes to elicit preferences.

Perfectly stating pairwise preferences is seldomly possible for a decision maker. Many factors (such as the number of possible transitive steps) can cause the introduction of inconsistencies in a decision matrix that results from pairwise comparison. Computing the consistency of a decision matrix is a means to assess the decision maker's understanding and experience in a field (which can help value decisions). There are many methods to compute the consistency of a decision matrix [13]. In this work we assess the quality of a given comparison matrix by means of the Consistency Ratio (CR) as proposed by Saaty [14]. It was introduced by Saaty [14] in order to check decision-makers preferences consistency in the AHP methodology. This CR was then analysed by several authors and is one of the main ways to check this consistency. It is the reason why we decided to use it in our study.

Computing the CR is a two step process. First, the Consistency Index (CI) of a matrix A of size n is computed according to Eq. 1, where  $\lambda_{max}$  is the maximum eigenvalue of the matrix. Then, the CR is the ratio between the CI and a real number called the Random Index (Eq. 2). The random index for a matrix of size n (RI<sub>n</sub>) is an estimation of the average CI obtained from a large set of randomly generated matrices of size n [15].

$$\text{Eq. 1 : } CI(A) = \frac{\lambda_{max} - n}{n - 1}$$

$$\text{Eq. 2: } CR(A) = \frac{CI(A)}{RI_n}$$

### 3. Transitive Spatial Comparisons

Our approach to support PCs while maintaining consistency builds on two pillars. Firstly, it proposes a new visual tool (a User Interface widget) to express relative preferences. Secondly, all PCs are updated on every preference update using the transitivity property of an (assumed) consistent AHP matrix. The design of the widget conveys the transitivity of comparisons

The proposed widget is depicted in Figure 4. It resembles a 2D, continuous data plot. The vertical axis is labeled with the expressions that are normally given to the values in Saaty's scale. All items to compare are placed on the horizontal axis. The plot line includes a handle (a small circle) for the value corresponding to each item. The first handle (in this case, Price) is "anchored" to the middle value (representing 1, or equally important). All other handles can be moved upwards or downwards.

At first, all handles are anchored at the middle position indicating that they are equally important. The user moves handles to indicate how a given item compares to the anchored one. For example, Figure 4 shows that the handle for RAM has been moved upwards to indicate that RAM is very strongly better than Price (the anchored item). Moreover, moving a handle to express how the item compares to the anchored one, also indicates how it compares to all other items. That is, moving the handle for RAM expresses how it compares to the anchored item (Price), but also to all other items.

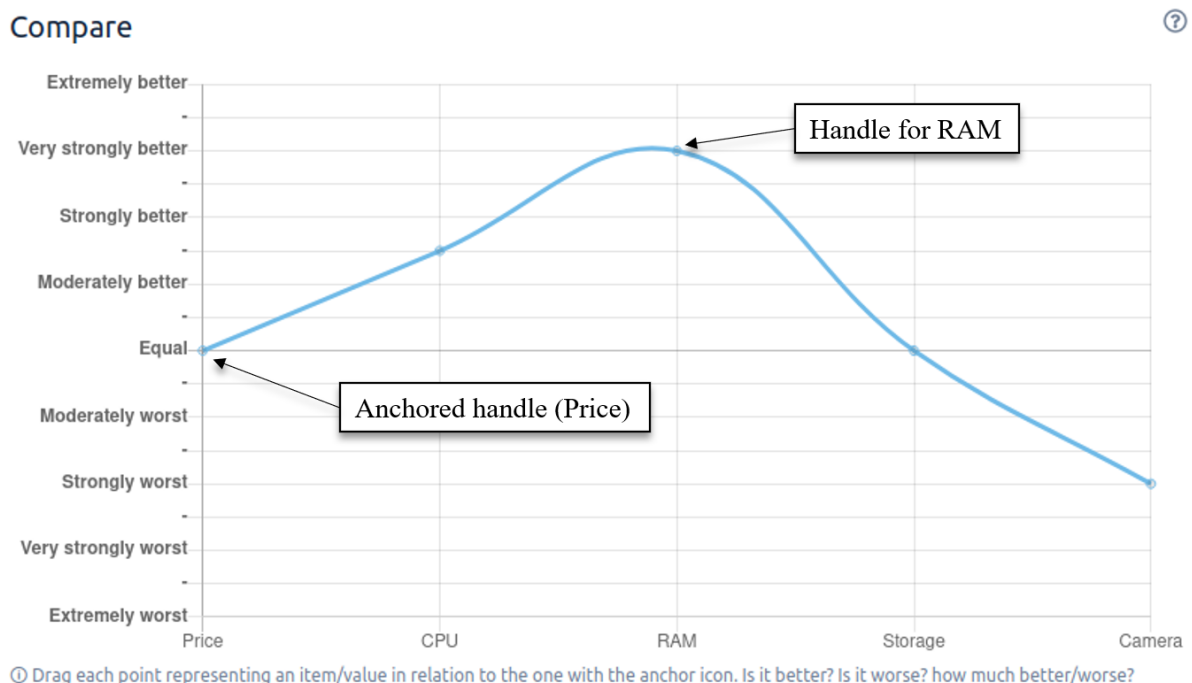


Fig 4: UI widget to present and elicit PCs

The widget offers a complete picture of how items compare to one another; however, labels in the vertical axis are expressed only in relation to the anchored item. Double clicking on a handle (the small circle on the plot), anchors the criterion for that handle to the left, updating the position of all other handles to reflect the change. This feature lets users explore the comparison space from the perspective of each item.

The visual representation of the comparisons helps maintain transitivity. By plotting all preferences in a 2D space, they inherit the transitivity properties of the comparison function in real numbers. Moving one handle simultaneously expresses how the item compares to the anchored one, and to all other items. This feature of the widget reduces the number of actions (e.g., clicks) that the user must make to compare all items.

PCs are stored in a matrix using Saaty's scale; rows and columns represent the handles (i.e., the items that handles stand for). The widget is divided in the middle by a line labeled "equally important". In addition, each half is divided by eight lighter lines (that correspond to the labels in the vertical axis).

When the user moves a handle  $h$ , the widget computes which line it is closest to. Then it computes the vertical distance  $d_i$  (rounded to the next integer) from that line, to the horizontal axis. The distance  $d_i$  is used to update the value in the cell that corresponds to the comparison between  $h$  and the anchored item. If  $h$  is above the horizontal

axis, the cell takes the value  $1 / (d_i + 1)$ ; otherwise, it takes  $(d_i + 1)$ .

As described previously, weights in an AHP matrix are consistent if they are transitive. That is,  $a_{ik} = a_{ij}a_{jk}$  for all  $i, j$ , and  $k$ . The widget forces the transitivity property, using it to compute all cell values for rows different from that of the anchored item. Listing 1 outlines the algorithm used to transitively update cells. It iterates only over the cells that are above the diagonal and not in the row that corresponds to the anchored item. For each cell, it sets the expected value (according to the transitivity formula), and it sets the value of its inverse. To deal with rounding errors and extreme values (smaller than  $1/9$  or higher than  $9$ ), the algorithm adjusts values to the closest in Saaty's scale.

```
1: for i = 1 to n-1
2:   for j = i+1 to n
3:     if (i != anchored)
4:       transitively_update(i,j);
5:       set(j, i, 1 / get(i,j));
```

Listing 1: Using the transitive property to update cells not corresponding to the anchored item

In combination, the widget design and the update function limit the sources of inconsistencies to only those cases that reflect extreme comparisons (i.e. the transitive distance between two items is larger than what Saaty's scale can express).

## 4. Evaluation

We argue that the proposed approach represents an improvement in terms of usability and consistency of the resulting comparison matrix. To formally demonstrate such a claim, we conducted a controlled experiment. The general design of the experiment involves comparing the proposed widget presented in section 3 to the widespread questionnaire widget depicted in Section 2, Figure 3 . The experiment aims to answer the following questions:

- Efficiency: Does the new widget reduce the effort of conducting pairwise comparisons?
- Efficacy: Does the new widget improve the consistency ratio of the resulting comparison matrix?
- Ease of use: Is the new widget easier to use than the questionnaire widget?
- Is there a correlation between the use of one widget or the other and the perceived validity of the resulting ranking?

The experiment was designed as a sequence of tasks to be completed, and surveys to complete. Following we describe the experiment protocol in more detail.

### 4.1 Participants

The experiment involved 45 students from a computer and information system Master in Toulouse as the experimental subjects. All 45 students have received at least one year of computer training, and master basic decision-making knowledge and clear logic. In addition, 5 researchers, holding PhDs in computer science from the Toulouse-IRIT laboratory took part. The researchers are not related to the project. All participants in the experiment have a basic understanding and basic knowledge in the field of decision support.

### 4.2 Evaluation Protocol

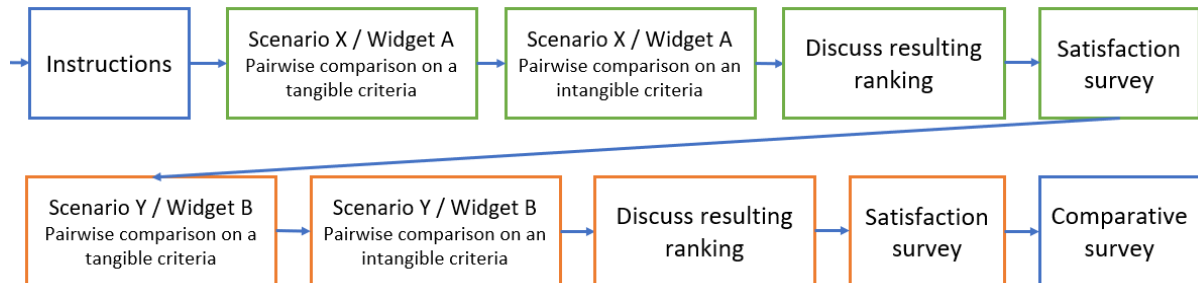
The experiment was carried out individually, one participant at a time. A Within-Subjects design was used for this study, meaning that each subject used both pairwise comparison widgets.

Figure 5 provides an overview of the experiment protocol. First, participants receive instructions related to the tasks they will be asked to perform. Instructions include a brief introduction to AHP and the widget that will be used. Special care was taken to avoid guiding the participants to any expected outcome. Then, each participant conducts comparisons in two different scenarios. Each scenario includes two pairwise comparisons, one of them on a tangible criteria and the other one on an intangible criteria. In one scenario participants compare travel destinations in terms of price (tangible) and attractions (intangible). In the other scenario, participants compare apartments in terms of price (tangible) and characteristics (intangible). Both pairwise comparisons in each scenario must be conducted with one of the two widgets. Both widgets must be used (one for each scenario).

To simplify the experiment, participants are not asked to compare criteria, only alternatives. Criteria are defined to be equally important. After participants finish both pairwise comparisons in one scenario, the resulting

ranking of alternatives is presented, and discussed. If the resulting ranking does not match the one expected by the participant, the participant is asked to write down an alternative one.

After completing each scenario, participants answer a satisfaction survey. Finally, after completing both scenarios, participants answer a comparative survey. This experiment design results in four different combinations of scenarios and widgets (see Table 1).



**Fig 5:** Overview of the experiment session

**Table 1:** Four possible combinations of scenarios and widgets to avoid learning effect and task bias

| Scenario X                    | Widget A (Tangible criteria)    | Widget A (Intangible criteria)            | Scenario Y                    | Widget B (Tangible criteria)    | Widget B (Intangible criteria)            |
|-------------------------------|---------------------------------|---|-------------------------------|---------------------------------|---|
| Choosing a travel destination | Questionnaire widget (on price) | Questionnaire widget (on attractions)     | Choosing an apartment         | 2D Plot widget (on price)       | 2D Plot widget (on characteristics)       |
| Choosing a travel destination | 2D Plot widget (on price)       | 2D Plot widget (on attractions)           | Choosing an apartment         | Questionnaire widget (on price) | Questionnaire widget (on characteristics) |
| Choosing an apartment         | Questionnaire widget (on price) | Questionnaire widget (on characteristics) | Choosing a travel destination | 2D Plot widget (on price)       | 2D Plot widget (on attractions)           |
| Choosing an apartment         | 2D Plot widget (on price)       | 2D Plot widget (on characteristics)       | Choosing a travel destination | Questionnaire widget (on price) | Questionnaire widget (on attractions)     |

### 4.3 Tools and data collection

Both comparison widgets were deployed as part of an ad-hoc web application. The Plot 2D widget was implemented as depicted in Figure 4 (in the previous Section). The questionnaire widget was implemented as depicted in Fig. 5 below.



**Fig 6:** Questionnaire widget as implemented in the web application used for the experiment

It collects the start time and end time for each scenario, and for each pairwise comparison. It also records the number of mouse clicks to complete each pairwise comparison. This information is used to compute efficiency. In this experiment, efficacy is defined in terms of the CR of each comparison matrix. The application transparently generates the comparison matrix that corresponds to each pairwise comparison. Then, it computes the CR using the procedures presented in Section 2.1 . After each subject completed the experiment, the following data was exported from the application and collected in a spreadsheet for later analysis:

- Title of scenario X
  - Name of 1st criteria (tangible)
    - Name of the widget A
    - Time (ms) to complete this PC
    - Clicks to complete this PC
    - CR of this PC
  - Name of 2nd criteria (intangible)
    - Name of the widget A
    - Time (ms) to complete this PC
    - Clicks to complete this PC
    - CR of this PC
  - Resulting ranking
  - Expected ranking
- Title of scenario Y
  - Name of 1st criteria (tangible)
    - Name of the widget B
    - Time (ms) to complete this PC
    - Clicks to complete this PC
    - CR of this PC
  - Name of 2nd criteria (intangible)
    - Name of the widget B
    - Time (ms) to complete this PC
    - Clicks to complete this PC
    - CR of this PC
  - Resulting ranking
  - Expected ranking

The satisfaction survey that participants complete after each scenario was created on the basis of the Systems Usability Scale survey [16]. It assesses perceived ease of use. It consists of the following 10 questions that can be answered with a value in the range 1 (strongly disagree) to 5 (strongly agree).

1. I think I'll use "this comparison widget" frequently.
2. I find "this comparison widget" unnecessarily complex.
3. I think "this comparison widget" is easy to use.
4. I think I will need the help of a technician to be able to use "this comparison widget".
5. I found that the various functions of "this comparison widget" were integrated well.
6. I think there is too much inconsistency in "this comparison widget".
7. I imagine most people would be able to learn to use "this comparison widget" very quickly.
8. I found "this comparison widget" very cumbersome to use.
9. I felt very confident using "this comparison widget."
10. I need to learn a lot of things before I can use "this comparison widget".

The final comparative survey includes only the following two questions:

1. Which tool allowed you to model more precisely the relative importance of each alternative?
2. Which Tool is easiest to use?

### 4.3 Results

After 35 days of experiment, a total of 50 persons participated in the experiment. After data screening and cleaning, there were 13 sets of data with missing data and incomplete information. Therefore, we conducted data

analysis and mining for the remaining 37 sets of data. Following, we present results one research question at a time.

- Efficiency: Does the new widget reduce the effort of conducting pairwise comparisons?

To compare both tools in terms of efficiency, we considered the time it took to complete each PC, and the number of clicks it required. Table 2 presents the results. Some samples were discarded as they presented invalid data, such as a very short time in the order of a few seconds, which suggested that the participant did not complete the task. Comparing the means for the numbers of clicks of both samples, yields a difference of 5.98 clicks less for the 2D Plot. Given the sample size and the standard deviation of the sample, this indicates that the 2D Plot requires significantly less effort with significance value (P-value) of 0.0001. In terms of time, the Plot 2D did not show a statistically significant improvement over the questionnaire widget.

**Table 2:** Comparison of time and number clicks to complete PC

|                      | Clicks (Mean) | Clicks (Std) | Time (Mean) | Time (Std) | Data points |
|----------------------|---------------|--------------|-------------|------------|-------------|
| Questionnaire widget | 16.76         | 4.82         | 352.135     | 185.756    | 37          |
| 2D Plot Widget       | 10.78         | 3.61         | 269.324     | 173.828    | 37          |

- Efficacy: Does the new widget improve the consistency ratio of the resulting comparison matrix?

To answer this question, we computed the CR of the resulting comparison matrix after each pairwise comparison. In each case we computed the pass rate (i.e., whether the CR was lower than 0.1).

Using tool A and tool B to make decisions in the same scenario, the ratio of CR in the consistency test results is less than 0.1. For the AHP model, consistency is a very important issue, and it is also one of the limitations of the AHP model. Generally, the AHP model cannot arrange the order of more than ten criteria or more than ten alternatives, because once the ten factors are exceeded, it is difficult to avoid inconsistencies. Especially for untrained decision makers, it is difficult to maintain logic and consistency when comparing multiple factors. A major advantage of the new approach is that it can avoid the problem of inconsistency in the user's ranking.

According to the rules of the AHP model, the CR needs to be less than 0.1. If the CR is less than 0.1, then we consider the consistency test to be passed, if it is greater than or equal to 0.1, then it is deemed that the consistency test fails. Each Scenario has two Criteria, one of them of tangible nature, and the other one of intangible nature. Table 3 presents the results regardless of the type of criteria, and Table 5 and 6 discriminate between tangible and intangible criteria. In all cases, the 2D Plot widget leads to better consistency.

Table 3: CR regardless of the type of criteria

|                      | CR (Mean) | CR (Std) | Pass rate | Data points |
|----------------------|-----------|----------|-----------|-------------|
| Questionnaire widget | 0.21494   | 0.18189  | 28.38%    | 74          |
| 2D Plot Widget       | 0.03669   | 0.06868  | 90.54%    | 74          |

Table 4: CR for tangible criteria

|                      | CR (Mean) | CR (Std) | Pass rate | Data points |
|----------------------|-----------|----------|-----------|-------------|
| Questionnaire widget | 0.18681   | 0.1621   | 32.43%    | 37          |
| 2D Plot Widget       | 0.02408   | 0.0495   | 91.89%    | 37          |

Table 5: CR for intangible criteria



|                      | CR (Mean) | CR (Std) | Pass rate | Data points |
|----------------------|-----------|----------|-----------|-------------|
| Questionnaire widget | 0.24307   | 0.19537  | 24.32%    | 37          |
| 2D Plot Widget       | 0.04931   | 0.08163  | 89.19%    | 37          |

- Is there a correlation between the use of one widget or the other and the perceived validity of the resulting ranking?

At the end of each scenario (i.e., after completing the two required PCs) the application presented a final ranking of alternatives. The participant had to indicate if the presented ranking was the expected one or not. If not, the participant indicated what was the expected ranking. Fig. 7 reports the count of matches and misses by widget. It was not a surprise to discover that misses largely outcount matches for both widgets. Whether or not the final ranking matches the user's expectations depends on multiple factors. Firstly, forcing both criteria to be equally important possibly contradicts the model the participants would have constructed. Second, if two alternatives obtain the same rank the tool randomly decides which one to rank first.

The focus of this experiment is on the effect of different widgets used in the decision making process (in PC in particular), not the whole AHP model and process. The key question is whether both widgets integrated similarly into the whole model. To gain further insight into the relation between the widget and the final ranking, we computed the Levenshtein distance between the observed and the expected ranking. Fig. 8 reports the results for all misses. It can be observed that both widgets obtained a similar distribution. This result indicates that there is no significant correlation between the widget, and between the expected and observed ranking.

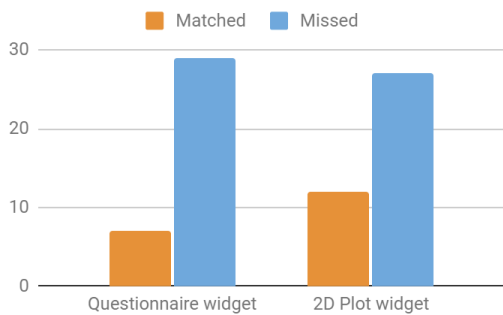


Fig. 7: Correspondence of expected rank vs. observed rank, by widget.

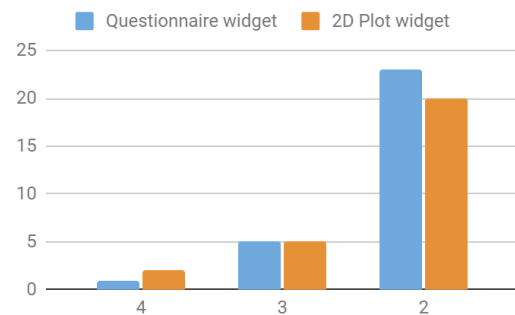


Fig. 8: Levenshtein distance between expected and observed rank, by widget

- Ease of use: Is the new widget easier to use than the questionnaire widget?

The satisfaction (perceived ease of use) survey was completed for both widgets by 30 participants. Responses were used to produce a score in the range 0-100 as suggested by the SUS method. Results were interpreted as relative to the observations made by Sauro[17]. According to the author, who conducted a large number of usability studies and compared the results, any score value under 68 is considered to be below average. Scores above 68 can be grouped in three buckets, each representing the top 30% (grade C, for scores between 68 and 74), top 20% (grade B, for scores between 74 and 83.1), and top 10% (grade A). Figure 9 presents the scores obtained by both widgets, grouping them in the above average buckets. The Plot 2D widget was perceived above average for more participants, and with higher grades in general.

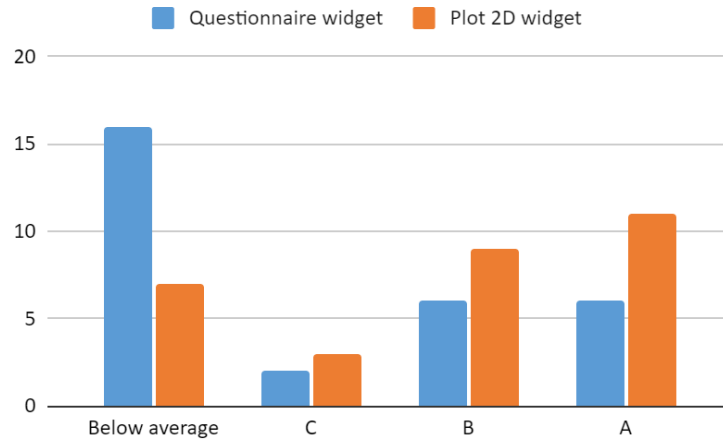


Fig. 9: SUS Scores for both widgets, for 30 responses.

The comparative survey that users completed after finishing the experiment confirms that the users perception favored the 2D Plot widget. As depicted in Figure 10, the 2D Plot widget allowed users to model more precisely the relative importance of each alternative, and was easier to use.

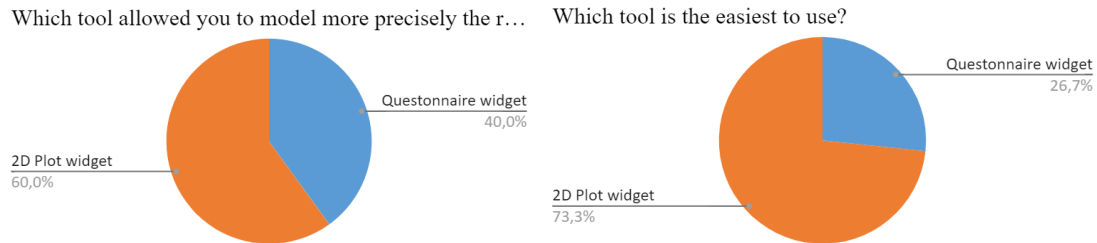


Fig. 10: Results of the comparative survey

#### 4.4 Discussion

Our aim was to show that the proposed approach represents an improvement in terms of usability and consistency of the resulting comparison matrix. The experiment was designed to answer the following questions:

- Efficiency: Does the new widget reduce the effort of conducting pairwise comparisons?

The experiment has shown that the number of clicks and the time are clearly reduced using the 2D Plot Widget (see Table 2).

- Efficacy: Does the new widget improve the consistency ratio of the resulting comparison matrix?

Tables 3, 4, 5 have shown that the CR (that is a representation of consistency) is better, regardless of the type criteria or for each type of criteria, using the 2D Plot Widget. Although the 2D Plot aims to remove the sources of inconsistencies, extreme transitive comparison can still cause the CR to be non-zero. The widget still allows the user to express extreme comparisons such as criterion A being "extremely better" than criterion B (the anchored item), and criterion C being "extremely worse" than criterion B. It can correctly express these comparisons in the underlying matrix as 1/9 and 9 respectively. However, when computing (via the algorithm in Listing 1) the value for the pairwise comparison for A and C, it obtains values that are smaller than 1/9 or higher 9. The algorithm rounds these values to the closest one in Saaty's scale, which causes the CR to be non-zero.

- Ease of use: Is the new widget easier to use than the questionnaire widget?

The Plot 2D widget was perceived above average for more participants, and with higher grades in general.

- Is there a correlation between the use of one widget or the other and the perceived validity of the resulting ranking?

Figures 7, 8 indicate that there is no significant correlation between the widget, and between the expected and observed ranking. Nevertheless, as shown in Figure 10 the 2D Plot widget allowed users to model more precisely the preferences of each alternative, and was easier to use.

This experiment has shown that the 2D Plot Widget is easier to use and allows a better consistency for end-users. Nevertheless, we have to mention that the number of operable data is not consistent for all questions, as an example the satisfaction (perceived ease of use) conclusions are based on only 30 answers as the other conclusions are based on 37 answers.

To consolidate these results, the study must be conducted with more subjects. Another issue is that the subjects were students using computers easily. It could be interesting to compare two kinds of subjects : students not accustomed to using these Widgets and Information technologies students.

## 5. Conclusions

PCs are a central feature of AHP. They are simple to use; however, as the number of items to compare increases, so does the effort of conducting all comparisons and the probability of introducing inconsistencies increase as well. We presented an innovative approach to conduct PCs based on a UI widget that resembles an interactive data plot. It uses the transitivity property of a consistent comparison matrix to infer comparisons. Our hypothesis is that this new approach is more efficient (as it reduces the number of actions the user must conduct to compare all items), more effective (as it limits the sources of inconsistencies), and yields better user satisfaction.

Pairwise comparison tools frequently present each comparison independently. In contrast our widget presents multiple pairwise comparisons at once, visually suggesting how they relate to one another (especially via transitivity). Presenting comparisons this way may hinder (psychological) independence of comparisons. Moreover, the curve that connects handles may misguide users to believe that there is something between items (taking intermediate values). Studying the impact of these potential drawbacks is the focus of future work.

In the AHP methodology proposed by Saaty, inconsistencies are addressed thanks to the systematic pairwise comparisons. This way of capturing users' preferences ensures consistency of preferences as these preferences are processed at the deepest level of details. Nevertheless, this approach is very time consuming and we cannot guarantee that at the end of the process the end-user remembers his own choices done at the beginning of the process. Another approach to capture users' preferences is to ask them to directly evaluate the criteria and the alternatives on an ordinal scale. This approach has also shown its limits as it is very difficult for a decision-maker to proceed without any comparison. Thanks to the developed widget, our approach is a proposal to guarantee decision-makers' consistency by proposing global comparisons among criteria and alternatives and at the time the end-user saves time.

## REFERENCES

- [1] T. L. Saaty, *Fundamentals of Decision Making and Priority Theory*, 1st edition. Pittsburgh, PA: RWS Publications, 2000.
- [2] E. Triantaphyllou, *Multi-criteria Decision Making Methods: A Comparative Study*, vol. 44. Boston, MA: Springer US, 2000. doi: 10.1007/978-1-4757-3157-6.
- [3] T. Tanino, 'Fuzzy preference orderings in group decision making', *Fuzzy Sets and Systems*, vol. 12, no. 2, pp. 117–131, Feb. 1984, doi: 10.1016/0165-0114(84)90032-0.
- [4] A. Tversky, 'Intransitivity of preferences.', *Psychological Review*, vol. 76, no. 1, pp. 31–48, 1969, doi: 10.1037/h0026750.
- [5] M. Regenwetter, J. Dana, and C. P. Davis-Stober, 'Transitivity of preferences', *Psychological Review*, vol. 118, no. 1, pp. 42–56, Jan. 2011, doi: 10.1037/a0021150.
- [6] J. Benítez, X. Delgado-Galván, J. Izquierdo, and R. Pérez-García, 'Improving consistency in AHP decision-making processes', *Applied Mathematics and Computation*, vol. 219, no. 5, pp. 2432–2441, Nov. 2012, doi: 10.1016/j.amc.2012.08.079.
- [7] T. L. Saaty, *Super Decisions*. 2004. [Online]. Available: <https://superdecisions.com/>
- [8] *Expert Choice*. Expert Choice, 2021. [Online]. Available: <https://www.expertchoice.com>
- [9] S. Siraj, L. Mikhailov, and J. A. Keane, 'PriEsT: an interactive decision support tool to estimate priorities from pairwise comparison judgments', *International Transactions in Operational Research*, vol. 22, no. 2, pp. 217–235, 2015, doi: <https://doi.org/10.1111/itor.12054>.
- [10] T. Kämpke, F. J. Radermacher, and P. Wolf, 'Supporting preference elicitation: The FAW preference elicitation tool', *Decision Support Systems*, vol. 9, no. 4, pp. 381–391, Jun. 1993, doi: 10.1016/0167-9236(93)90048-8.
- [11] E. Abel, I. Galpin, N. W. Paton, and J. A. Keane, 'Pairwise comparisons or constrained optimization? A usability evaluation of techniques for eliciting decision priorities', *International Transactions in Operational Research*, vol. n/a, no. n/a, Nov. 2020, doi: <https://doi.org/10.1111/itor.12907>.
- [12] I. Millet, 'The Effectiveness of Alternative Preference Elicitation Methods in the Analytic Hierarchy

- Process', *Journal of Multi-Criteria Decision Analysis*, vol. 6, no. 1, pp. 41–51, 1997, doi: [https://doi.org/10.1002/\(SICI\)1099-1360\(199701\)6:1<41::AID-MCDA122>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1099-1360(199701)6:1<41::AID-MCDA122>3.0.CO;2-D).
- [13] M. Brunelli, 'A survey of inconsistency indices for pairwise comparisons', *International Journal of General Systems*, vol. 47, no. 8, pp. 751–771, Nov. 2018, doi: 10.1080/03081079.2018.1523156.
- [14] T. L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York ; London: McGraw-Hill, 1980.
- [15] M. Brunelli, *Introduction to the Analytic Hierarchy Process*. Springer International Publishing, 2015. doi: 10.1007/978-3-319-12502-2.
- [16] J. Brooke, 'SUS - A quick and dirty usability scale', p. 8.
- [17] J. Sauro, *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Denver, CO: CreateSpace Independent Publishing Platform, 2011.