Are inconsistent decision better?
An interactive experiment with pairwise comparisons

Portsmouth Business School
Richmond Building
Portland Street
PO1 3DE

Alessio Ishizaka
Kyriacos Antoniades
Which fruit do you prefer?

How many times?

2 times

3 times
Which fruit do you prefer?

Which fruit do you prefer?

3 times

How many times?

4 times
Is this inconsistency possible?

2 times

2 times?

3 times
\[ \text{Comparison matrix} \]

\[
\begin{array}{ccc}
1 & a_{12} & \\
& a_{21} & 1 \\
& 1 & a_{ij} \\
& 1/a_{ij} & 1 \\
\end{array}
\]

\[ \text{CI} = \frac{\lambda_{\text{max}} - n}{n - 1} \]

where \( \lambda_{\text{max}} \) is the maximal eigenvalue

\( n \) is the dimension of the matrix
Consistency Ratio

- CR = CI/RI  < 10%

where CR is the consistency ratio
RI is the random index

Saaty (1977) calculated the following random indices:

<table>
<thead>
<tr>
<th>n</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RI</td>
<td>0.58</td>
<td>0.9</td>
<td>1.12</td>
<td>1.24</td>
<td>1.32</td>
<td>1.41</td>
<td>1.45</td>
<td>1.49</td>
</tr>
</tbody>
</table>

**Critic:** 10% is an arbitrary value
**Inconsistent matrices are not better!**

- 18 graduate students compare five different compact cars in global terms, and also in terms of their aesthetics.
- When intransitivities are automatically removed, the preferences of decision makers are not better represented.

Why automatic correction does not work?

How to travel to Gatwick airport?

- coach
- train
- personal car
- car sharing
- taxi
Rankings produced for each participant

- **Original Ranking** ($R_O$), where the priorities are calculated by the eigenvector method without any inconsistency correction.

- **Automatic Ranking** ($R_A$), where inconsistencies are corrected automatically using the goal programming method.

- **Interactive Ranking** ($R_I$), where the software indicates to the participant the most inconsistent pairwise comparison to the least one and invite her/him to change them.
**Experimental procedure (1)**

1. The decision problem is explained to the participant.

2. The participant pairwise compares the five alternatives.

3. The **Consistency Ratio**, the **Original Ranking** and the **Automatic Ranking** are calculated.

4. If the consistency ratio is acceptable, i.e. **below 10%**, the experiment terminates otherwise the consistency error \( \varepsilon_{ij} \) of each pairwise comparison, is calculated with \((\text{Saaty, 2003})\):

\[
\varepsilon_{ij} = \max \left( a_{ij} \cdot \frac{p_j}{p_i}, a_{ji} \cdot \frac{p_i}{p_j} \right)
\]
Experimental procedure (2)

5. Possibility to revise the most inconsistent comparison, i.e. the comparison with the highest $\varepsilon_{ij}$. If they decline, they are asked if they want to revise the next most inconsistent comparison. When they revise, the process restarts from point 4 until the inconsistency falls below 10% or the participant has considered all entries.

6. The final Interactive Ranking is calculated.

7. The participant is asked which of the three rankings (without knowing how they have been calculated) represents their preference.
Results: problem order influence

• Sixty-two participants.
  – The first thirty-one participants solved the problem with the subjective criterion first and then the problem with the objective criterion.
  – The next thirty-one participants solved the problems in the reverse order.

• Both samples produced statistically identical outcomes.
  – Order did not have any influence on the results
Consistency improvement with the interactive method

<table>
<thead>
<tr>
<th>Problem with criterion</th>
<th>Final matrix with improved consistency</th>
<th>Final matrix consistency improved not improved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem with subjective criterion</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>Problem with objective criterion</td>
<td>34</td>
<td>0</td>
</tr>
</tbody>
</table>

In 100% of the cases, the interactive method improved the consistency.
### Alternatives ranking

<table>
<thead>
<tr>
<th>City</th>
<th>Original Ranking</th>
<th>Interactive Ranking</th>
<th>Automatic Ranking</th>
<th>Normalised true distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiff</td>
<td>0.186 ± 0.069</td>
<td>0.190 ± 0.069</td>
<td>0.226 ± 0.081</td>
<td>0.162</td>
</tr>
<tr>
<td>London</td>
<td>0.260 ± 0.053</td>
<td>0.261 ± 0.053</td>
<td>0.267 ± 0.058</td>
<td>0.267</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>0.046 ± 0.040</td>
<td>0.046 ± 0.041</td>
<td>0.068 ± 0.056</td>
<td>0.022</td>
</tr>
<tr>
<td>Southampton</td>
<td>0.423 ± 0.104</td>
<td>0.419 ± 0.093</td>
<td>0.338 ± 0.087</td>
<td>0.471</td>
</tr>
<tr>
<td>Liverpool</td>
<td>0.085 ± 0.033</td>
<td>0.084 ± 0.032</td>
<td>0.101 ± 0.036</td>
<td>0.078</td>
</tr>
</tbody>
</table>

- All three rankings ordered the distance of the cities to Portsmouth correctly.
Error between estimated and true distances

<table>
<thead>
<tr>
<th>City</th>
<th>Original priority - true distance</th>
<th>Interactive priority - true distance</th>
<th>Automatic priority - true distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiff</td>
<td>0.087</td>
<td>0.082</td>
<td>0.140</td>
</tr>
<tr>
<td>London</td>
<td>0.057</td>
<td>0.058</td>
<td>0.089</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>0.042</td>
<td>0.040</td>
<td>0.048</td>
</tr>
<tr>
<td>Southampton</td>
<td>0.024</td>
<td>0.024</td>
<td>0.046</td>
</tr>
<tr>
<td>Liverpool</td>
<td>0.023</td>
<td>0.022</td>
<td>0.029</td>
</tr>
</tbody>
</table>

- The automatic ranking is furthest from the true distance.

- The original and interactive priorities are very close, which makes the effort to improve the consistency questionable, if the final result does not improve.
Priorities of transport selection

<table>
<thead>
<tr>
<th>City</th>
<th>Original priority</th>
<th>Interactive priority</th>
<th>Automatic priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>0.187 ± 0.131</td>
<td>0.190 ± 0.132</td>
<td>0.183 ± 0.123</td>
</tr>
<tr>
<td>Coach</td>
<td>0.121 ± 0.104</td>
<td>0.127 ± 0.099</td>
<td>0.131 ± 0.098</td>
</tr>
<tr>
<td>Taxi</td>
<td>0.179 ± 0.142</td>
<td>0.173 ± 0.145</td>
<td>0.167 ± 0.133</td>
</tr>
<tr>
<td>Car sharing</td>
<td>0.176 ± 0.134</td>
<td>0.181 ± 0.137</td>
<td>0.161 ± 0.105</td>
</tr>
<tr>
<td>Own car</td>
<td>0.337 ± 0.219</td>
<td>0.214 ± 0.230</td>
<td>0.357 ± 0.229</td>
</tr>
</tbody>
</table>

• The priorities of the subjective problems were more dispersed, i.e. the standard deviation was higher.

• The “own car” alternative was by far the most preferred transportation mode in the original and interactive ranking. This clear preference for ‘own car’ was faded in the interactive ranking.
Participants’ preferred ranking for the subjective problem

A Chi-square test confirms that the frequencies of participants’ preferences were equally distributed.
Participants’ preferred ranking for the objective problem

<table>
<thead>
<tr>
<th></th>
<th>Interactive</th>
<th>Automatic</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Freq</td>
<td>18</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Expected Freq</td>
<td>(11.3 (.3))</td>
<td>(11.3 (.3))</td>
<td>(11.3 (.3))</td>
</tr>
</tbody>
</table>

Note. $\chi^2 = 8.71$, degree of freedom = 2, significance threshold $p > .05$

- A Chi-square test confirmed that the frequency of the participants’ preferences were not equally distributed.
- If the automatic ranking is ignored, there is no significant difference between the original and interactive ranking with a Chi-square test.
Closest ranking to the true value

<table>
<thead>
<tr>
<th>Rankings closest to the true value</th>
<th>Interactive</th>
<th>Automatic</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed Frequency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td><strong>Expected Frequency (proportion)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed Frequency</td>
<td>6</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Expected Frequency (proportion)</td>
<td>18.5 (3)</td>
<td>18.5 (3)</td>
<td>18.5 (3)</td>
</tr>
</tbody>
</table>

Note. $\chi^2 = 8.71$, degree of freedom = 2, significance threshold $p > .05$

- A Chi-square test confirmed that the frequency of the participants’ preferences were not equally distributed.
- If the automatic ranking is ignored, there is no significant difference between the original and interactive ranking with a Chi-square test.
Conclusions:

- **Consistency improvement:**
  - The interactive and automatic methods *improved* consistencies in pairwise comparisons
  - Revisions *are in agreement* with the best fit for the pairwise comparison

- **Representation of ranking:**
  - Interactive approach *does not* better represent participants’ preferences
  - The original ranking is *closest* to the true value in the objective problem
  - The difference between the priorities of the original and interactive ranking were found to be very small
Main Conclusion

The effort to reduce inconsistencies using the interactive approach is questionable.

Dr Alessio Ishizaka
Reader in Decision Analysis
University of Portsmouth
Portsmouth Business School
Richmond Building
PO1 3DE Portsmouth
United Kingdom

Alessio.Ishizaka@port.ac.uk
Thank you!