The Impact of Evaluability on Choice: A Structural Choice Modelling Formulation of the Evaluability Thesis

Ann Wallin
4th July 2013
Proliferation of comparison websites in the Australian market (e.g. iSelect.com, comparison.com.au, getprice.com.au etc.)

Governments around the world are using comparison websites to encourage competition and increase consumer market power.

Comparison websites are thought to reduce information search time, allow consumers to receive the best deal and increase competition among suppliers (Laffey and Gandy 2009).

Within a decade the UK comparison website market has grown to be worth more than £1 billion (David 2008).
The focus on comparison websites has shifted the predominant evaluation context from single evaluation (SE) to joint evaluation (JE).

There is mounting evidence which suggests that such changes in evaluation mode may result in consumers making different choices.

Studies have found that changes in evaluation mode can lead decision makers to reverse their preferences (Hsee, 1996; Gonzalez-Vallejo and Moran, 2001), make mispredictions (Loewenstein et al, 2003) and miscalculate experienced utility (Hsee and Zhang 2004).

Such comparisons may lead to choices of an option that yields a worse consumption experience (Hsee and Zhang 2004; Yang et al 2011).
Single Evaluation (SE)

**Front Load 8.5kg Washing Capacity Inox Colour**

- Digital inverter motor with 11 year parts replacement warranty on models purchased from 1 October 2012
- Inox In Colour
- Maximum 1,400RPM spin speed
- Bubble Wash Feature
- My Cycle
- Easy Iron

**RRP $1599**

* Pricing subject to change

Where to buy

---

Joint Evaluation (JE)

**Comparing Washing Machines**

- **Daewoo Electronics DWC-LD1412**
  - 1,399.99
  - Buy now for $938

- **LG WD14030D6**
  - 1,499.99
  - Buy now at 3 stores from $1,100

- **Samsung WF158UH9AGD**
  - 1,599.99
  - Buy now at 2 stores from $969

**MY WISHLIST**

- Add to wishlist

**RELEASED TO MARKET**

- Jul 2019
- Jul 2011
- Aug 2012

**WARRANTY PERIOD**

- 2 years (Drum 5 yrs, Motor 10 yrs)
- 2 years (Motor 8 yrs)

**SPECIAL FEATURES**

- Integrated Dryer
- Sensor System
- Electronic Controls
- Lint Filter
- Stain Stove
- Fabric Softener Dispenser
- Bleach Dispenser
- Delay Start
- Auto Spin Balancing

**GENERAL**

- Load Type: Front Load
- Washload Capacity: 8.5 kg
- Drum Construction: Stainless Steel

**SPIN**

- Max Spin Speed: 1400 rpm
- No Off Spin Speeds: 4

**WATER**

- Water Efficiency Rating: 4.5 Stars
- Water Consumption: 67 L
- Water Level Adjustment: Automatic

**ENERGY**

- Energy Rating: 4 in Stars
- Warm Wash Energy Consumption: 380 kWh/yr
- Cold Wash Energy Consumption: 118 kWh/yr

**DIMENSIONS**

- Width: 600 mm
- Height: 850 mm
- Depth: 645 mm
All consumer decisions are made in one or two basic modes: single or joint.

Early work in the area of evaluability focused on preference reversals, such as Hsee’s (1996) classic study.

<table>
<thead>
<tr>
<th></th>
<th>Dictionary A</th>
<th>Dictionary B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of publication:</td>
<td>1993</td>
<td>1993</td>
</tr>
<tr>
<td>Number of entries:</td>
<td>10,000</td>
<td>20,000</td>
</tr>
<tr>
<td>Any defects?</td>
<td>No, it’s like new</td>
<td>Yes, the cover is torn; otherwise, it’s like new.</td>
</tr>
</tbody>
</table>

This preference reversal phenomena is theoretically distinct from the findings of the choice versus judgement paradigm.
Hsee and Leclerc (1998) replicated the preference reversal between joint and single evaluations of cordless phones, light bulbs and CD changers.

Preference reversal occurs due to some attributes being easier to evaluate independently, whereas other attributes are more difficult to evaluate without reference to the other options available (González-Vallejo and Moran 2001).

The evaluability hypotheses states: When two stimulus options involve “a trade-off between a hard-to-evaluate attribute and an easier-to-evaluate attribute, the hard-to-evaluate attribute has a lesser impact in single evaluation than in joint evaluation, and the easy-to-evaluate attribute has a greater impact” (Hsee 1996, p250).
An attribute is hard to evaluate independently when people are unable to assess how good a given value on an attribute is when the value is presented.

It appears that in joint evaluation mode individuals can use one value of an attribute as a reference to evaluate the other, which is unable to be achieved in single evaluation (Hsee and Zhang 2010).

More recent formulations of the theory have widened the frame of the evaluability hypothesis from just preference reversal to value sensitivity.
The evaluability hypothesis has found consistent support but is characterised by several limitations. The theory can be moved forward by applying new techniques to achieve greater insight into evaluability.

**Contribution**

(1) testing the evaluability hypothesis in conditions characterised by complex realistic scenarios and repeated decision making; and

(2) using structural choice modelling to understand the whether the structure of peoples preferences is changing under single and joint evaluation modes.

Ultimately, this paper seeks to remove the current constraints/ limitations of the evaluability hypothesis to widen its application to the marketing field.
From the previous findings we believe there will be differences in the means for attributes across the single and joint evaluation modes.

This study focuses on understanding how the heterogeneity in people’s preferences is structured. There are three competing hypotheses about the structure of heterogeneity of the tasks.

**H1a:** The structure of heterogeneity is common across evaluation modes.

**H1b:** The structure of heterogeneity is unique to each evaluation mode.

**H1c:** There are two heterogeneity structures. A structure that is common to both single and joint evaluations modes, and a structure that is unique to each mode.
Method

A discrete choice experiment (DCE) was developed for each task (single and joint).

The design of the experiment started with research into the common features offered within the smartphone category. Attributes common to a majority of smartphones with ranges in specifications were chosen.

Respondents for the DCE’s were a sample of 142 students choosing to buy or not buy 8 Smartphone handsets in the single evaluation tasks and choosing among alternative from 8 sets of Smartphone handsets (or not buy) in the joint evaluation tasks.
## Features and Levels

<table>
<thead>
<tr>
<th>Feature</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2-level Attributes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video</td>
<td>YES</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPS</td>
<td>GPS Receiver</td>
<td>Maps Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Music Player</td>
<td>YES</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bluetooth</td>
<td>YES</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key Pad</td>
<td>Hard Key Pad</td>
<td>Touch Screen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screen Size</td>
<td>3.0 inch</td>
<td>4.5 inch</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4-level Attributes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capacity</td>
<td>8GB</td>
<td>16GB</td>
<td>32GB</td>
<td>64GB</td>
</tr>
<tr>
<td>Camera</td>
<td>0</td>
<td>4MP</td>
<td>8MP</td>
<td>12MP</td>
</tr>
<tr>
<td>Battery Time (talk time)</td>
<td>4 hours</td>
<td>6 hours</td>
<td>8 hours</td>
<td>10 hours</td>
</tr>
<tr>
<td>Operating System</td>
<td>Android</td>
<td>iOS</td>
<td>RIM Blackberry</td>
<td>Symbian</td>
</tr>
<tr>
<td>Price</td>
<td>$250</td>
<td>$500</td>
<td>$750</td>
<td>$1000</td>
</tr>
<tr>
<td>Colour</td>
<td>Black</td>
<td>White</td>
<td>Silver</td>
<td>Red</td>
</tr>
</tbody>
</table>
Stimuli

Would you purchase Phone A?

- [ ] Purchase
- [ ] Not Purchase

Would you purchase Phone A or Phone B?

- [ ] Phone A
- [ ] Phone B
- [ ] Purchase Neither
Results

As a starting point the fixed and random models were estimated.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Latent Variables</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M(1) – Fixed</td>
<td>24 $\mu_e$’s</td>
<td>-</td>
<td>-1644.00</td>
<td>3336.01</td>
<td>3456.85</td>
</tr>
<tr>
<td>M(2) – Random</td>
<td>34 $\mu_e$’s; 32 $\sigma$’s</td>
<td>-</td>
<td>-1497.88</td>
<td>3127.75</td>
<td>3460.08</td>
</tr>
</tbody>
</table>
H1a: The structure of heterogeneity is common across single and joint evaluation tasks.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Latent Variables</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M(1) – Fixed</td>
<td>24 $\mu_e$’s</td>
<td>-</td>
<td>-1644.00</td>
<td>3336.01</td>
<td>3456.85</td>
</tr>
<tr>
<td>M(2) – Random</td>
<td>34 $\mu_e$’s; 32 $\sigma$’s</td>
<td>-</td>
<td>-1497.88</td>
<td>3127.75</td>
<td>3460.08</td>
</tr>
<tr>
<td>M(3) – 1 Factor</td>
<td>34 $\mu_e$’s; 32 $\gamma$’s</td>
<td>1</td>
<td>-1492.67</td>
<td>3117.33</td>
<td>3449.66</td>
</tr>
</tbody>
</table>
H1b: The structure of heterogeneity is unique for single and joint evaluation tasks.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Latent Variables</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M(1) – Fixed</td>
<td>24 $\mu_\epsilon$’s</td>
<td>-</td>
<td>-1644.00</td>
<td>3336.01</td>
<td>3456.85</td>
</tr>
<tr>
<td>M(2) – Random</td>
<td>34 $\mu_\epsilon$’s; 32 $\sigma$’s</td>
<td>-</td>
<td>-1497.88</td>
<td>3127.75</td>
<td>3460.08</td>
</tr>
<tr>
<td>M(3) – 1 Factor</td>
<td>34 $\mu_\epsilon$’s; 32 $\gamma$’s</td>
<td>1</td>
<td>-1492.67</td>
<td>3117.33</td>
<td>3449.66</td>
</tr>
<tr>
<td>M(4) – 2 Factor</td>
<td>34 $\mu_\epsilon$’s; 32 $\gamma$’s</td>
<td>2</td>
<td>-1497.38</td>
<td>3126.76</td>
<td>3459.09</td>
</tr>
</tbody>
</table>
H1c: The are two heterogeneity structures. A structure that is common to both single and joint evaluations tasks, and a structure that is unique to each task.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Latent Variables</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M(1) – Fixed</td>
<td>24 $\mu_c$'s</td>
<td>-</td>
<td>-1644.00</td>
<td>3336.01</td>
<td>3456.85</td>
</tr>
<tr>
<td>M(2) – Random</td>
<td>34 $\mu_c$'s; 32 $\sigma$'s</td>
<td>-</td>
<td>-1497.88</td>
<td>3127.75</td>
<td>3460.08</td>
</tr>
<tr>
<td>M(3) – 1 Factor</td>
<td>34 $\mu_c$'s; 32 $\gamma$'s</td>
<td>1</td>
<td>-1492.67</td>
<td>3117.33</td>
<td>3449.66</td>
</tr>
<tr>
<td>M(4) – 2 Factor</td>
<td>34 $\mu_c$'s; 32 $\gamma$'s</td>
<td>2</td>
<td>-1497.38</td>
<td>3126.76</td>
<td>3459.09</td>
</tr>
<tr>
<td>M(5) – 2 Factor with correlation</td>
<td>34 $\mu_c$'s; 32 $\gamma$'s; 1 $\phi$</td>
<td>2</td>
<td>-1484.88</td>
<td>3103.76</td>
<td>3441.12</td>
</tr>
</tbody>
</table>

$r = 0.79$
M5 means (t-values)