Not worth the extra cost? Diluting the differentiation ability of highly rated products by altering the meaning of rating scale levels

MARTIN MEIßNER1*,†, STEFANIE LENA HEINZLE2,† and REINHOLD DECKER1

1Bielefeld University, Department of Business Administration and Economics, Universitätsstraße 25, 33615 Bielefeld, Germany
2University of St. Gallen, Institute for Economy and the Environment, Tigerbergstraße 2, St. Gallen, Switzerland

ABSTRACT

Over the last decade, the use of rating scales has grown in popularity in various fields, including customer online reviews and energy labels. Rating scales convey important information on attributes of products or services that consumers evaluate in their purchase decisions. By applying multidimensional scaling, this paper reveals that the meaning of the levels of a rating scale can be altered by manipulating the labeling of the rating scale levels. The study reveals that consumers perceive product attributes as being more similar if the labels share similar or identical linguistic or visual characteristics. In addition, two choice-based conjoint studies examine whether the way consumers make their choices among products can be influenced by changing the labeling of rating scale levels. The results show that a manipulation of the meaning of rating scale levels diminishes both the importance of the rating scale information and consumers’ willingness to pay a premium for a rating upgrade. Copyright © 2013 John Wiley & Sons, Ltd.

INTRODUCTION

Rating scales have become an integral part of consumer review platforms, online stores, and online travel agencies like imdb.com, amazon.com, and expedia.com. By evaluating important attributes of products or services, such as the quality of products, movies, or hotels, rating scales commonly allow consumers to assess the performance of the respective product or service. Another important area of application of rating scales are energy efficiency labels for household appliances, cars, and buildings. Those labels indicate the energy efficiency of a product along a rating scale to allow comparisons of the energy efficiency between products. Usually, such rating scales employ a purely uniform scale with successive integers of numerical grades (e.g., 1–2–3–4–5) or an ordered set of letters (e.g., A–B–C–D–E–F–G). Rating scale applications, however, also provide evidence of rating scale designs that do not follow such a pattern: the European Union, for instance, just recently decided to revise the European energy efficiency label, which was originally based on a seven-point scale ranging from A to G, by expanding the scale to a maximum of three tiers (A+, A++, and A+++). The study of the seven-point A–G rating scale by adding new classes A+, A++, and A+++ revealed that the proposed extension of the seven-point A–G rating scale by adding new classes A+, A++, and A+++ causes consumers to weight energy efficiency less in their decisions.

The aim of the present article is threefold: first, building on the primary effect shown by Heinzle and Wüstenhagen (2012), it aims to provide an explanation, based on psychological research on the concept of similarity, as to why such an alteration of the labeling of the levels of a rating scale can influence consumers’ perceptions. By applying multidimensional scaling (MDS), the paper investigates whether consumers perceive products being rated with scale levels that are labeled with similar or identical visual or linguistic characteristics (e.g., A++, A++) as being more similar with respect to the underlying dimension of the scale (e.g., the energy efficiency of the product) compared with products being rated with scale levels that are labeled with distinct characteristics (e.g., A, B). Second, the present paper aims to explore the robustness of the phenomenon as discovered by the aforementioned authors outside of the energy efficiency context. In a typical consumer decision context (holiday trips), we investigate in two choice-based conjoint (CBC) studies whether the importance consumers attach to hotel quality rating information as well as their willingness to pay (WTP) are influenced when similar or identical linguistic or visual characteristics are used for the labeling of two hypothetical hotel rating scales. Third, the paper contributes to previous research on the effect of the labeling of rating scale levels on respondents’ interpretation and use of the scale levels (e.g., Wildt and Mazis, 1978; Schwarz et al., 1991; Weijters et al., 2005; Weijters et al., 2010). The application of rating scales has its origin in the social sciences, where rating scales are generally used in surveys in order to determine attitudes, opinions, or behaviors of respondents. Schwarz et al. (1991), for instance, showed that research results can be affected by the numerical values attached to the labels. Their study revealed that when the numerical values used for labeling the rating scale levels were modified from “0 to 10” to “−5 to 5,” consumers’ response behavior was significantly influenced. The alteration of the rating scale resulted in a less significant utilization of the lower scale levels when those were assigned a negative label (−5 to −1). The aim of the present paper is therefore to contribute to previous research regarding the more general question of how respondents make use of information conveyed by specific rating scales’ labels.

The paper proceeds as follows: DEVELOPMENT OF RESEARCH HYPOTHESES section reviews psychological...
research on the concept of similarity and its influence on choice behavior and derives testable research hypotheses. STUDY 1 and STUDIES 2 AND 3 sections present the design and results of three empirical studies. CONCLUSIONS AND IMPLICATIONS section finally summarizes key results of the empirical studies and discusses the relevance of the results as well as practical implications.

DEVELOPMENT OF RESEARCH HYPOTHESES

According to Tversky (1977: 327), “the concept of similarity is ubiquitous in psychological theory.” Similarity typically refers to the outcome of a comparison among objects by defining the degree to which options differ from each other. According to Tversky (1977), similarity is determined by contrasting shared and unshared features of two objects, increases with the number of common features, and decreases with the number of distinctive features. For instance, Tversky proposed that the letter “E should be more similar to F than to I because E and F have more common features than E and I” (Tversky, 1977: 11). The law of similarity (Koffka, 1935), likewise, states that stimuli with similar visual characteristics such as shape, size, and color tend to be perceived as belonging to the same group or unit.

On the basis of Tversky’s concept of similarity, we argue that an explanation for the results found in Heinzle and Wüstenhagen (2012) could be formulated as follows: by using similar or identical visual or linguistic characteristics (e.g., the same letter of the alphabet “A” in combination with additional plus signs) for the labeling of only a subset of scale levels of a rating scale, it could be hypothesized that two products that are rated with those similarly labeled scale levels should also be perceived as being more similar with respect to the underlying dimension of the rating scale (e.g., the energy efficiency of a product). This leads to the following hypothesis H1 (tested in study 1):

\[ H1: \text{Consumers perceive products, which are rated with scale levels that are labeled with similar or identical visual or linguistic characteristics, as being more similar with respect to the underlying dimension of the scale (compared with products that are rated with scale levels which are labeled with distinct characteristics).} \]

As described earlier, the second aim of the present paper is to replicate the effects already shown by Heinzle and Wüstenhagen (2012) in a typical consumer decision context (holiday trips). Similar to the results shown by the aforementioned authors, we assume that by using similar or identical characteristics for the labeling of a subset of the scale levels of two hotel rating scales, the importance of the rating scale information in consumers’ decision-making should also be reduced. The following hypothesis H2 is therefore proposed (tested in studies 2 and 3):

\[ H2: \text{The importance of a rating attribute (i.e., the hotel quality rating information) in consumers’ buying decisions significantly decreases if a subset of the scale levels of a rating scale is labeled with similar or identical visual characteristics.} \]

In addition, it can be supposed that respondents are less willing to pay a price premium for an “upgrade” in hotel quality from a lower to a higher rating level if the perception of similarity with respect to quality is influenced by labeling the levels with similar or identical linguistic or visual characteristics. Therefore, hypothesis H3 is derived as follows (tested in studies 2 and 3):

\[ H3: \text{The price premium consumers are willing to pay for a rating upgrade from a lower to a higher rating level (i.e., in hotel quality) is significantly less pronounced when those scale levels are labeled with similar or identical visual or linguistic characteristics (compared with when those scale levels are labeled with distinct characteristics).} \]

STUDY 1

Multidimensional scaling, which has its roots in psychophysics, is an established technique for measuring the perceived similarity of objects. The most widespread MDS procedures are based on spatial distance models (Carroll and Arabie, 1980), which represent stimuli in an n-dimensional space. In these models, the distances between stimuli convey the perceived similarities between them (e.g., Kruskal, 1964). MDS is particularly suitable for testing the perceived similarity between stimuli when they only differ along a small number of continuous dimensions (Tversky and Hutchinson, 1986). Thus, as for rating scales, which only have one underlying continuous dimension (e.g., the quality of the hotel accommodation or the energy efficiency of an appliance), the MDS technique seems to be a natural candidate to capture the similarity between the different rating classes in our study. We therefore applied MDS to test hypothesis H1.
Stimulus material and sample
Eighty-four undergraduate students took part in the study and were randomly assigned to three different groups. In groups 1 and 2, two different rating systems, indicating the quality of the hotel with eight and six rating scale levels, respectively, were presented to the respondents. In group 1, the hotels were rated on a scale ranging from 65 (indicating a rather low quality of the hotel) to 100 points (indicating a very high quality of the hotel) with six scale levels in between (i.e., 70, 75, 80, 85, 90, and 95). Together with this numerical rating, each of the scale levels received a different background color (Figure 2). Eight equidistantly perceived color codes ranging from pure red to pure green were selected on the basis of the Natural Color System, a perceptual color model by the Scandinavian Color Institute. The equidistant numerical rating scale was manipulated by assigning the three scale levels 85, 90, and 95 with the same green background color as the highest level equaling 100 points.

Similarly, in group 2, a purely uniform scale with successive integers of numerical grades ranging from 6 to 1 was replaced by six rating scale levels ranging from 3 (indicating a rather low quality of the hotel) to 1*** (indicating a very high quality of the hotel) (Figure 3).

Finally, in group 3, the new European Union energy label ranging from D (lowest energy efficiency rating) to A+++ (highest energy efficiency rating) was the subject of investigation (Figure 1).

Procedures
In each group, the respondents had to compare two products at a time. The products only differed with respect to the hotel quality (groups 1 and 2) or with respect to the energy efficiency of a household appliance (group 3). Each rating scale level was compared successively with all other scale levels, resulting in (8-7)/2 = 28 trials for group 1, (6-5)/2 = 15 trials for group 2, and (7-6)/2 = 21 trials for group 3. Thus, the stimuli were presented in all possible pairs to the respondents. The subjects were instructed to rate the products’ similarity on a 21-point scale ranging from 0 (not similar at all) to 20 (very similar). Stimulus presentation was carried out in randomized order and accompanied by the following instructions:

Groups 1 and 2: Please imagine that you intend to book a hotel online. You are informed about a rating indicating the overall quality of the hotel. The rating of the hotel is based on the following scale: (altered rating scales as displayed in Figures 2 and 3 were presented). The two hotels only differ regarding their quality rating: (pair of rating scale levels is shown). If there is no other information available, how similar are the two hotels with respect to quality?

Group 3: The European Union is planning to introduce a new label for different household appliances, which indicates how energy efficient the particular appliance is. The label looks like this: (altered rating scale as displayed in Figure 1 was presented). Please imagine that you intend to buy a household appliance. The two appliances only differ regarding their energy efficiency: (pair of rating scale levels is shown). If there is no other information available, how similar are the two appliances with respect to energy efficiency?

Results
The PROXSCAL algorithm performs MDS of the proximity data. Data of each respondent were used to calculate the perceived differences between scale levels on the individual level. The average differences between scale levels were then used to visualize the perceived similarity between them. Each rating scale level was plotted on a map in a one-dimensional space. As there was essentially only one underlying dimension relevant to the individuals (i.e., the quality of the hotel in groups 1 and 2 and the energy efficiency of an appliance in group 3), we decided to implement a one-dimensional solution. Although we also examined solutions with a higher number of dimensions, we rejected those as a meaningful interpretation of the dimensions was not possible. Average stress-1 coefficients of 0.13 in groups 2 and 3 and of 0.17 in group 1 according to Kruskal (1964) and Chen et al. (2008) indicate an acceptable goodness-of-fit for representing the input proximities in a one-dimensional space. Figure 4 shows the perceptual maps of all three groups, visualizing

![Figure 2. Hypothetical example for the alteration of background color of the rating scale levels.](image)

![Figure 3. Hypothetical example of alteration of numerical labeling of rating scale levels.](image)
the proximity of the rating scale levels. The distances separating the distinct rating scale levels physically on the perceptual map reflect the subjective distances perceived by the subjects.

The perceptual axes show almost equidistantly perceived distances between the scale levels 1, 2, and 3 in group 2 and between the scale levels A, B, C, and D in group 3. The perceived differences between scale levels that share similar or identical visual or linguistic characteristics, that is, A+++ to A, A++ to A+, and A and 1*** to 1**, 1* and 1, are significantly smaller. The statistical significance is tested by comparing the average perceived difference between scale levels 1 to 3 (A to D) with the average perceived difference between scale levels 1*** to 1 (A+++ to A). A two-sided Mann–Whitney U-test shows that the differences are significant for both group 2 (U = 6, Z = −6.33, p < 0.01) and group 3 (U = 18, Z = −6.13, p < 0.01). In group 1, the average difference in the perceived distance between the scale levels that share the same background color (i.e., 100, 95, 90, and 85) is also significantly smaller than the average difference between levels that were assigned a different background color (group 1: U = 184, Z = −4.34, p < 0.01).

These results clearly support H1, postulating that respondents perceive products (i.e., hotels or household appliances) that are rated with scale levels which share similar or identical visual or linguistic characteristics, as being more similar with respect to the underlying dimension of the scale (i.e., the hotel quality or the energy efficiency of the appliance).

However, we only presented the respective altered rating scales to the respondents (e.g., an efficiency rating scale ranging from A+++ to D) and did not test the perceived similarity between the original rating scale levels (e.g., an efficiency rating scale ranging from A to G). We therefore have to acknowledge that the result of non-equidistance of similarly labeled rating scale levels cannot be attributed with certainty to the use of the manipulated scale levels.\(^1\) We therefore conducted studies 2 and 3 to investigate whether the altered rating scales also influence consumers’ buying decisions.

### STUDIES 2 AND 3

In order to measure how important consumers value the different attributes of the product in their decision-making and, subsequently, to investigate their WTP a premium for an upgrade in hotel quality rating, we apply a CBC analysis. An advantage of CBC is that it measures consumers’ preferences in quasi-realistic purchase decisions (Orme, 2010). Consumers are presented with several product alternatives that are typically described as profiles of multiple attributes and are asked to choose the alternative that best suits their preferences. Choosing between the presented alternatives implies that consumers have to make tradeoffs between different levels of product attributes (Orme, 2010).

#### Stimulus material and samples

Both studies were conducted with university students, namely, 142 in study 2 and 192 students in study 3. In both studies, a between-subjects design was used, which split respondents into two groups.

The respondents were asked to indicate their preferences for hotels in 16 CBC choice tasks in study 2 and 20 CBC choice tasks in study 3 in which they had the task to book a hotel for a 1-week summer holiday. The respondents were told that they needed to make a decision for one hotel from a set of four (study 2) hotel alternatives with each alternative defined on two attributes (rating of the quality of the hotel and price per week per person). Because of the fact that we

\(^{1}\)We would like to thank one anonymous reviewer for this comment on the paper.
included a higher number of scale levels in study 3 (i.e., eight rating scale levels in study 3 versus six rating scale levels in study 2), we decided to increase the number of concepts per choice task for study 3 by one additional hotel alternative. That way, both designs were at a comparable level of accuracy with respect to utility estimation. To assure realistic prices, a pre-study with 50 students was conducted to determine the WTP for a 1-week stay in a hotel. The pre-study resulted in a price range from €289 to €489 (study 2) and €289 to €499 (study 3). Six price levels (with an increment of €40) resulted in study 2, and eight price levels (with an increment of €30) resulted in study 3.

The quality rating information was presented on a six-point rating scale ranging from 1 (indicating a very high quality) to 6 (indicating a rather low quality) to the respondents in group 1 of study 2. A six-point rating scale ranging from 1*** to 3 was presented to the respondents of group 2 (Figure 2). In study 3, an eight-point quality rating scale ranging from 65 (indicating a rather low quality of the hotel) to 100 (indicating a very high quality of the hotel) with six scale levels in between (i.e., 70, 75, 80, 85, and 90) was presented to all respondents participating in this study. Whereas in group 1 each rating scale level was assigned an individual background color, in group 2, the four highest scale levels (85, 90, 95, and 100) were assigned the same background color (Figure 3).

To sum up, in both studies only the visual or linguistic characteristics for labeling of the levels of the hotel quality rating attribute varied between the groups, either by altering the numerical grades or the background colors assigned to the scale levels. The two studies also differed with respect to the number of attribute levels included.

**Procedures**

As usual in SAWTOOTH SOFTWARE, the most often applied tool for conducting CBC interviews, the complete enumeration algorithm was used to design several choice tasks for each respondent in the empirical studies (Sawtooth SOFTWARE, 2008). By using the same starting seed and keeping the number of questionnaire designs constant for both groups, identical randomized designs were generated for both groups in CBC. In both studies, the respondents were randomly assigned to one of the two groups and one of six questionnaire designs per group.

In both studies, the respondents were asked to complete a paper-and-pencil questionnaire. After a short explanation of both attributes, the choice scenario was explained. The respondents then answered the subsequent choice tasks as shown in Figures 5 (study 2) and 6 (study 3).

**Estimation and data quality**

Respondents’ part-worth utilities were measured at the individual level by using hierarchical Bayes estimation (Huber and Train, 2001). In order to provide support for hypothesis 2, attribute importances were derived from the range of the part-worth utilities as common in conjoint analytic preference measurement (Sawtooth Software, 2008). For the rating attribute, the part-worth type of utility estimation function was chosen. The effect of the price attribute was estimated as a single linear coefficient to reflect the slope of price. A comparison of utilities across samples is difficult as respondents usually do not make choices with the same consistency; consistent respondents are characterized by larger scale parameters than uncertain or inconsistent respondents.

![Figure 5. Example of a choice task in study 2 for groups 1 (top) and 2 (bottom).](image-url)
(Swait and Louviere, 1993; Islam et al., 2007). Thus, there is an increasing concern regarding the role of the “scale” when utilities across different samples are compared (e.g., Swait and Louviere, 1993; Louviere and Eagle, 2006). There might be a risk that “distributions of preference parameters will be confounded with distributions of scales” (Louviere and Eagle, 2006: 214). By expressing conjoint utilities in monetary terms, it is possible to cancel out the scale effect as WTP is not affected by the scale parameter and thus allows for comparison between different models (Louviere and Islam, 2008). Because of the fact that consumer preferences are heterogeneous, consumers’ WTP were calculated on the basis of the individual part-worth utilities measured using the hierarchical Bayes approach. An alternative solution would be to estimate a scale-adjusted latent class model which captures both preference parameters and a scale factor simultaneously (Bateman et al., 2011). Therefore, in order to provide support for hypothesis H3, the results are expressed in terms of implicit WTP when the part-worth utility coefficients are converted into monetary units.

The approach described in Sattler et al. (2010) was used to calculate consumers’ WTP. Applying this approach, the number of iterations that were done before convergence was assumed amounted to 20,000 preliminary iterations. Ten thousand additional iterations were then computed using data from all choice tasks, and each 10th replicate was saved for each respondent. Thus, for each respondent, 1000 utility draws were available for analysis. We divided the difference in attribute level utility between adjacent attribute levels of the rating attribute by the corresponding 1000 values of the part-worth distribution of the price vector (for a similar computation, see Sattler et al., 2010). We then calculated the mean of the price premium distribution at the individual respondent level.

The average root likelihood (RLH) value is an often used quality indicator of the goodness-of-fit of choice models. Accordingly, the achieved data quality can be rated as rather good (RLH (study 2, group 1) = 0.78; RLH (study 2, group 2) = 0.73; RLH (study 3, group 1) = 0.63, and RLH (study 3, group 2) = 0.65), as the estimated model predicted about 3.1 or 2.9 (study 2) and 3.2 or 3.3 (study 3) times better than the random model (with RLH = 0.25 for study 2 and RLH = 0.2 for study 3). Data from 21 respondents in study 2 and 17 in study 3 were excluded from further analysis because they either seemingly misunderstood the direction of the rating scale (despite this point being highlighted at the beginning of the questionnaire) or did not exceed a RLH threshold of 0.4. Furthermore, those respondents who indicated an implausible (i.e., positive) price coefficient were also excluded from further analysis. Finally, 129 (175) respondents remained for further analysis in study 2 (study 3).

**Results**

The importance measures are based on the range between the maximum and minimum part-worth utilities within each attribute, that is, the higher the range in the attribute’s utility values, the higher the relative importance of this attribute (Orme, 2010). In study 2, for respondents of group 1 (“1–6” scale), the quality rating was significantly more important and clearly dominated the purchase decision of the respondents. The relative importance of the rating scale amounted to 70.21%. In contrast, when presenting students with the rating scale ranging from 1*** to 3 in group 2, the importance of the rating scale was much lower, dropping by more than 16 percentage points to 54.04% (Table 1). This difference in the importance of the rating scale was highly significant ($t = 9.666, DF = 127, p < 0.001$).
In study 3, the results also show that the quality rating was more important for group 1 "multi-color scale" compared with group 2 "green color scale". For respondents of group 1, the relative importance of the rating scale amounted to 59.46%. When presenting students with the "green color scale" in group 2, the importance of the rating scale was lower, decreasing to 55.69%. A t-test shows that the quality rating attribute was significantly less important for respondents in group 2 (t<2.091, DF=173, p=0.038). In sum, these results clearly support H2 postulating that the importance of the hotel quality rating information in consumers’ buying decisions significantly decreases if a subset of the scale levels of a rating scale is labeled with similar or identical visual characteristics.

The monetary value (median WTP) of the difference in attribute level utilities between adjacent scale levels was calculated to test H3 (Table 2). For study 2, a two-sided Mann–Whitney U-test shows that the price premium consumers were willing to pay for an upgrade in hotel quality from the second highest to the highest rating level (i.e., from 2 to 1) in group 1 is significantly higher than the price premium consumers were willing to pay for an equivalent upgrade in hotel quality from the second highest to the highest rating level (i.e., from 1** to 1*** in group 2 (p<0.01). In addition, the price premium consumers were willing to pay for an upgrade in hotel quality from the third (fourth) highest to the second (third) highest rating level, that is, from 3 to 2 (4 to 3) in group 1 was also higher than the price premium consumers were willing to pay for an equivalent upgrade in hotel quality from the third (fourth) highest to the second (third) highest rating level, that is, 1* to 1** (1 to 1*) in group 2 (p<0.01).

For study 3, the results are very similar: the differences in premiums consumers were willing to pay for an upgrade in hotel quality were significantly smaller in the case where these scale levels were assigned the same background color (p<0.01).

CONCLUSIONS AND IMPLICATIONS

The first aim of the present paper was to provide an explanation for the primary effect shown by Heinzle and Wüstenhagen (2012) as to why an alteration of the labeling of scale levels can influence consumers’ perceptions of rating scales. The results of study 1 show that the meaning of rating scale levels in consumer choices can be manipulated. MDS revealed that consumers perceive products as being more similar with respect to the hotel quality or the energy efficiency of a household appliance if the labels of the rating scale levels share similar or identical linguistic or visual characteristics.

Moreover, the second aim of this paper was to replicate the results from Heinzle and Wüstenhagen (2012) outside of the energy efficiency context. The results of two subsequent CBC studies show that the importance consumers attach to the hotel quality rating information in their decision-making as well as their WTP for an upgrade in hotel quality rating decreased if the scale levels were labeled as outlined earlier.

Third, the paper contributed to previous research that showed that the labeling of the scale levels largely influences how consumers make use of scale information. In fact, the obtained results have important implications for marketers, public policy officials, and others who are involved in the design (and use) of rating scales: rating scale designers might manipulate the meaning of rating scales and, in consequence, the decision behavior of consumers by altering the labeling of the scale levels in the aforementioned way. This alteration, however, might have unintended consequences: with regard to the new European energy label design, for instance, manufacturers who show technological leadership might get a lower return on their investment in R&D with the introduction of the new classes A+, A++, and A+++.

Furthermore, in using rating scales where levels are labeled with similar or identical visual or linguistic characteristics, online review platforms and online travel agencies may intend to show that the majority of the products are of high quality. However, if, as a consequence thereof, prices become more important in consumers’ decisions, this behavior may have negative effects on the profitability of the platform provider assuming that they receive different commissions for different (hotel) prices. In addition, designing such a rating scale is also not in the best interest of the producers and manufacturers, who offer a superior quality and could (potentially) lose their competitive advantage over less expensive products with lower ratings. Introducing a rating scale as described earlier therefore results in a decrease in the differentiation potential of high-quality products. We thus conclude that when rating scales become an integral part of eco-labels, consumer review platforms, online stores, or online travel agencies, particular attention must be paid to the labeling of the respective rating scales.

Finally, marketing practitioners could also apply the insights provided in this article in a way to boost sales. For instance, colored price tags could be used to manipulate consumers’ perceptions of price levels. By assigning identical colored price tags to products with different price values, consumers might perceive the differences in costs
Table 2. Monetary value (median WTP) of the differences in attribute level utilities

<table>
<thead>
<tr>
<th>Study 2</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1</td>
</tr>
<tr>
<td></td>
<td>(“1–6 scale”)</td>
</tr>
<tr>
<td>n</td>
<td>60</td>
</tr>
<tr>
<td>1 vs. 2</td>
<td>€21.35</td>
</tr>
<tr>
<td>1*** vs. 1**</td>
<td>€47.27</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>€97.25</td>
</tr>
<tr>
<td>3 vs. 4</td>
<td>€233.03</td>
</tr>
<tr>
<td>1* vs. 1</td>
<td>€55.78</td>
</tr>
<tr>
<td>4 vs. 5</td>
<td>€233.03</td>
</tr>
<tr>
<td>1 vs. 2</td>
<td>€55.78</td>
</tr>
<tr>
<td>5 vs. 6</td>
<td>€233.03</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>€55.78</td>
</tr>
</tbody>
</table>

Note: n reflects the number of respondents in each group.
*p-value calculated on the basis of Mann-Whitney U-tests

*p < 0.05
**p < 0.01
***p < 0.001
between the various products as less significant and hence rather pay attention to other characteristics. At the same time, by using distinct or very dissimilar characteristics for labeling the scale levels of the rating attribute, consumers might focus more on the rating information when making a choice. Therefore, it can be supposed that the importance of price decreases, which in turn should have a positive impact on sales and profit.

ACKNOWLEDGEMENTS

Stefanie Lena Heinzle was supported by a Grant for Prospective Researchers from the Swiss National Science Foundation.

BIографICAL NOTES

Dr. Martin Meißner is an Assistant Professor of Marketing at the Department of Business Administration and Economics, Bielefeld University, Germany. He obtained a Master’s degree in Business Administration and a PhD in Marketing from Bielefeld University. His research focuses on consumer decision behavior and new preference measurement techniques. He has published in the Journal of Marketing Research, Journal of Multi-Criteria Decision Analysis, and the International Journal of Market Research.

Dr. Stefanie Lena Heinzle is currently working for BEUC - the European Consumer Organisation. From 2008 to 2012, she was a PhD student at the Institute for Economy and the Environment at the University of St. Gallen. Her research interests lied in sustainable consumption, energy labeling, and household energy behavior. She holds a degree in International Business Administration from the Vienna University of Economics and Business Administration and a CEMS Master degree of International Management. She was a Visiting Scholar at Bielefeld University and the National University of Singapore.

Dr. Reinhold Decker is a Full Professor of Marketing at the Department of Business Administration and Economics at Bielefeld University, Germany. He holds a Master’s degree in Business Engineering and a PhD in Marketing from the University of Karlsruhe. His research focuses on model-based decision support in marketing, data mining, and online preference measurement and he has published in various scholarly journals and conference proceedings. He was a Visiting Professor at, e.g., the Moscow Academy of Economics, the New University of Lisbon, and the University of Vienna, and he currently serves on the editorial board of Review of Managerial Science.

REFERENCES


