Chapter IV Using Conjoint Analysis for Aesthetic Research

1) Introduction

Aesthetic research uses visual stimuli that have been manipulated to vary on specific aesthetic dimensions. In the past, this variation has not always been performed systematically and in a controlled fashion. Moreover, the relative importance of different stimulus dimensions on aesthetic judgments is frequently not evaluated. Thus, even though past research has identified relevant dimensions, it is not known how important they are for the overall aesthetic judgment. The present research uses conjoint analysis to systematically construct visual stimuli and investigates the relative importance of different aesthetic stimulus properties and their variations for aesthetic judgments. Furthermore, results of conjoint analysis were used to establish an "optimal" rank order of the stimuli, which can serve as external standard against which an individual's aesthetic judgment can be evaluated.

2) Theoretical Considerations

Aesthetic research uses visual stimuli for investigating aesthetic judgments that range from simple polygonal figures to real-life objects (e.g., Bamossy et al., 1983; Berlyne, 1974b; Berlyne et al., 1974; Birkhoff, 1933; Eysenck & Castle, 1970; Götz et al., 1979). Researchers typically vary some properties[1] of the aesthetic stimuli to investigate their impact on aesthetic judgments. When real-life objects are used, however, these variations have in the past not always been done systematically and controlled. Eysenck, for example, asked a painter to create paintings that differ only on one dimension (harmony, Eysenck, 1983; Götz et al., 1979). The author hoped that this would result in stimuli that vary only on this one dimension, hoping that the painter would be able to keep all other aspects constant. However, because the different versions of the painting were created by the painter quite intuitively, there are chances that variations on the harmony dimension created changes in other dimensions as well. In other words, the dimension of harmony was not varied in a controlled way. In contrast, the present study describes the creation of stimuli following principles of conjoint analysis, allowing for a systematic variation of stimuli.

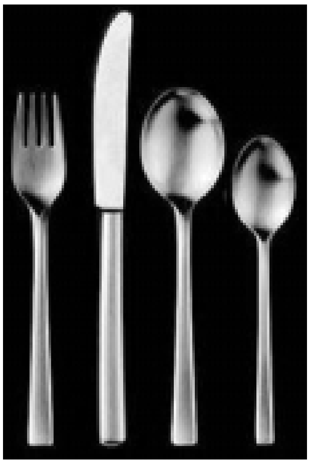
---

[1] Note that in the present text an aesthetic "property" refers to a property of a specific object such as the handle of a fork or knife. An aesthetic "dimension" refers to the relationship properties have to each other. Furthermore, the relationships among properties can be changed resulting in perceptions such as harmony or proportion.

In other research *multiple* dimensions of stimuli were changed simultaneously (e.g., Berlyne, 1970, 1974b). Obviously, when several dimensions of stimuli are changed at the same time, one cannot precisely evaluate which particular dimension affects the aesthetic judgment or whether some dimensions are more important than others. Thus, to investigate the impact of combinations of multiple dimensions on the aesthetic judgment, a systematic variation of the different dimensions is mandatory (e.g., Karpowicz Lazreg & Mullet, 2001). More precisely, when stimuli are simultaneously varied on different relevant aesthetic dimensions, such as proportion and harmony, one would have to examine the relative importance of each dimension for the overall judgment, rather than assuming that both equally affect the judgment. Even though commonly used data analysis techniques such as analysis of variance (ANOVA) can provide a measure for this relative importance using effect size (e.g., η^2), this statistic is usually not reported in the aesthetic literature. Past research has basically investigated which dimensions significantly *impact* the aesthetic judgment using ANOVA, factor analysis or regression (e.g., Berlyne, 1970; 1974a; Eysenck & Castle, 1970; Ritterfeld, 1996; Tobacyk, Bailey, & Myers, 1979) without explicitly focusing on the relative *importance* of each aesthetic dimensions (see Martindale, Moore, & Borkum, 1990, for exceptions). Furthermore, even if the relative importance of each aesthetic dimension is known, it is still unclear how the *combination* of different attribute levels within a stimulus might influence the overall aesthetic judgment. Imagine, for example, pieces of cutlery varying on the dimension of proportion (e.g., proportion distance from head to handle for three kinds of instruments - fork, knife and spoons) and the dimension of harmony (e.g., how well the three instruments match). One stimulus photo might show cutlery in which the handles are relatively short and the instruments match well (Figure 7). Another stimulus photo might show cutlery in which the handles are relatively long and the instruments do not match well because the knife is much longer than the fork and the spoon (Figure 8). In this case, knowing that the proportion dimension is overall more important for the aesthetic judgment than the harmony dimension (based on, e.g., effect size estimates in ANOVA or beta-weights from regression) would still not allow a researcher to infer which *combination* of different levels of the two dimensions is preferred over others. In other words, a rank order for the different objects (stimuli) cannot be built as long as the aesthetic preferences for different combinations of the dimensions are unknown. For instance, the medium level of the proportion dimension might be judged as more aesthetic than the high and the low levels, whereas the high level of the harmony dimension might be judged as more aesthetic than the medium and the low levels. The present study proposes the use of conjoint analysis to identify the relative importance of different aesthetic dimensions and their combinations. This will allow researchers to determine the differential impact of combinations of dimensions for aesthetic judgments.

Figure 1. Variation of cutlery stimulus

Figure 2. Variation of cutlery stimulus



Aesthetic research often investigates how much an individual's aesthetic judgment deviates from an external standard. In the past, this so called "objective" aesthetic value of stimuli has mainly been established either by: (a) experts (e.g., Berlyne, 1971; Child, 1962), (b) calculation of averages of aesthetic judgments across participants for the stimulus, or (c) a combination of both (see Eysenck, 1988). There have been several studies showing that the group-mean approach for non-experts results in similar results as the expert approach (see Eysenck, 1947) and that the degree of agreement tends to be constant over different kinds of material (Child, 1962; Eysenck, 1947). In the case of expert judgments, experts judge how aesthetic the stimuli are. Stimuli for which the experts agree in their judgment are then used for further research. However, this approach does not reveal the criteria that experts use for establishing the aesthetic values of objects. A similar problem exists with the average aesthetic judgment approach. In this approach simply the average of all participants' judgments for a stimulus is used as criterion. Again, it remains unclear what criteria the group used for their aesthetic judgments. Moreover, deviations from an average aesthetic judgment describe nothing more than to which extent the person agrees with the average. There are no theoretical reasons to assume that the average judgment represents a valid *external* criterion for the aesthetic value of the object. Accordingly, while potentially useful for particular research questions, expert or averaged participant ratings are of limited use for providing external standards. The present study proposes the use of conjoint analytic results for establishing an external standard. This standard is independent of specific reference groups. Furthermore, dimensions relevant for aesthetic judgments of the given stimulus classes were identified in previous studies using expert interviews (Study 4, Chapter 2) and multidimensional scaling (Study 5, Chapter 2). The present study, then, uses these dimensions for the construction of a set of stimuli that vary systematically on them. Thus, the provided standard is explicitly based on relevant aesthetic dimensions inherent in the stimuli. It therefore uses much more explicit criteria to evaluate an individual's aesthetic sensitivity than previous standards.

In sum, the purpose of the present study was to create a set of systematically varied real-life stimuli that differ only on relevant aesthetic dimensions. Using conjoint analysis the relative importance of each aesthetic dimension and its variations for the overall aesthetic judgment of stimuli was evaluated. The conjoint analytic results also provide an independent external standard for aesthetic judgments of the stimuli.

3) Study 9 - Conjoint Analysis

a. Method

Conjoint analysis is a multivariate technique that is used for modeling how individuals develop preferences for objects that differ in two or more attributes[2]. It is based on the assumption that individuals evaluate the total value of an object by combining the separate amounts of values each attribute provides. The theory underlying this approach is axiomatic conjoint measurement developed by Luce and Tukey (1964) and by Krantz et al. (Krantz, Luce, Suppes, & Tversky, 1971a). Conjoint analysis was initially applied in the field of consumer psychology (P. E. Green & Rao, 1971). Conjoint analysis provides alternative methods to model preference data (vector model, ideal-point model, part-worth function model and mixed model), to collect data (two-factor and full-profile method) and to construct sets of stimuli (fractional factorial design, random sampling from multivariate distribution). It also allows for different kinds of measurement scales for the dependent variable (e.g., paired comparison, rank order, rating scales) and uses different estimation methods in dependence of the used measurement scale. Further details of the basic principles of conjoint analysis can be

[2] Following the terms usually used in conjoint analysis, the term "attribute" is used instead of the term "dimension" here.

found in the works of Green and Wind (1975) and Green and Srinivasan (S. B. Green & Srinivasan, 1978, 1990). In the following only the methods used in the present research are described in more detail.

i. Construction method

In conjoint analysis the researcher constructs a set of stimuli by combining selected levels of each attribute (S. B. Green & Srinivasan, 1978, 1990). In this sense attributes are properties of a stimulus, such as the painting style used in a picture, and the attribute levels are thus different styles that might possibly be used to paint a picture. The three methods of stimulus construction most widely associated with conjoint analysis are the trade-off, the full-profile, and the pair comparison method. In the present study the most popular of the three methods, the full-profile method, is used. In this approach, a stimulus is constructed by combining selected levels of each attribute using all attributes simultaneously. Thus, any stimulus consists of a combination of one level for each attribute (Hair, 1998).

ii. Preference model

Conjoint analysis allows using different preference models for modeling the data. These models are the vector model, the ideal-point model, the part-worth function model and the mixed model. For modeling the decision process the part-worth function model was used for the present data. Based on the assumption that individuals evaluate the value or utility of an object by combining the separate amounts of values each attribute provides, the model reflects a utility function that defines a different utility value for each attribute level, a part worth. To obtain a value for a specific stimulus, the part-worth utilities belonging to the attribute levels the stimulus contains are summed. The resulting value is simply called utility. In addition to estimating the impact of each level with the part-worth utilities, the relative importance of an attribute can be assessed by calculating the range of part-worth utilities (difference between highest and lowest value). The resulting importance values can be converted to percentages in multiplying it by 100.

iii. Stimulus construction

The stimuli used in the present research were pictures of objects. Object classes chosen for the present study were paintings, offices, car interiors and cutlery. Attributes and levels for each object class are shown in Table 15 (Appendix A). The stimulus attributes were developed based on two sources: First, experts were interviewed and explicitly asked to articulate their judgment criteria, and second, multidimensional unfolding on preference judgments of stimuli of the relevant object classes (reported in Studies 4 and 5, Chapter 2) was used. Through these two methods, two relevant aesthetic attributes were identified for paintings: shape and painting technique. Shape ranges from rather straight geometrical to more curved shapes. Painting techniques represented different techniques that were used for paintings. Office stimuli include arrangement and perceived atmosphere as relevant aesthetic attributes. The arrangement attribute refers to a more or less appealing way of positioning furniture in an office room. The perceived atmosphere in the room ranges from brighter rooms with big windows to rather dark rooms without windows. For car interiors, simplicity, clarity and technology were identified as relevant attributes. The simplicity attribute refers to the overall impression of the interior ranging from a rather simple, classical design to a rather complex design. Clarity refers to the way instruments are set up in the car. Technology refers to the amount of technical equipment in the car interior, ranging from rather technologically advanced to rather basic. Finally, cutlery attributes include proportion, harmony and perceived ergonomics. The proportion attribute depicts the proportion of head to handle for each instrument of a set of cutlery. Harmony refers to how harmonious the different instruments (fork, knife and spoons) of a set of cutlery look together. Perceived ergonomics signifies the impression of how well the instruments might lay in a user's hand, depending on how the handle of the cutlery is shaped.

The stimuli were designed using a full factorial design, with one visual stimulus card realizing each of the possible eight (8) or nine (9) combinations (for stimuli see Appendix D). Prior to conjoint analysis, multidimensional unfolding was used to show that the above dimensions were correctly implemented in the stimuli and were used as common judgment criteria by participants (Study 6, Chapter 2).

b. Participants and Procedure

i. Participants

The stimuli were presented to an online sample of $N = 65$ participants. Thirty-four men and 29 women (two participants did not report their gender) between 21 and 61 years of age ($M = 32.34$, $SD = 9.53$) participated in the study. Participants were recruited online from two sites explicitly designed for conducting online studies. The announcements asked the addressees to participate in an online study about object aesthetics and provided them with the necessary link to the web page where the study was hosted.

ii. Procedure

Each participant was presented with all 34 stimuli. All stimuli in one object class were presented on the screen at the same time in a random order. The first page contained a short introduction to the study; the second page explained the procedure. The third page showed the stimuli of the first object class. On this and each following page participants were asked to rank order the stimuli from most to least aesthetic, assigning the smallest number (1) to the most aesthetic stimulus. Participants could only proceed to the next page if they had ranked all stimuli on a page. Once each participant advanced to the next page, he or she had no opportunity to go back to a previous page and change judgments.

4) Results

The data for all object classes were analysed using SPSS 10.0. Because the present research concerns the aesthetic value of the stimuli and not individual preferences for stimuli, the data were analysed on the aggregate level. That means that the individual part-worth utilities for each level were aggregated by calculating the average part-worth utilities over all individuals after they have been standardized. The results will be reported for each object class in turn. As goodness-of-fit measure for the four conjoint models the correlation between the input versus estimated values of the preferences as expressed by Kendall's tau is reported.

For paintings the goodness-of-fit was $tau = .83$ ($p = .00$), indicating a good fit. Table 16 (Appendix A) shows the average relative importance, part-worth utilities, utilities and rank orders for the painting stimuli. As Table 16 (Appendix A) shows, the relative importance of the two dimensions showed that the "painting technique" was slightly more important for the overall aesthetic judgment than the "shape". To identify which attributes significantly affect aesthetic judgments, Kohli's (1988) significance test was used. This test is closely related to Kruskal and Wallis' (1952) test for identical distribution functions. Kohli provides two versions of the test differing in terms of whether or not preferences for attribute levels can be ordered a priori. Because in the present study preferences for the attribute's levels could not be ordered a priori, the significance of the attributes was tested without constraining preferences for attribute levels. The λ^2 values resulting from the significance test were transformed into *z* – scores using a formula provided by Ferguson & Takane (1989). For the "shape" dimension a significant value of $z = 8.22$ ($p = .00$) resulted, for the "painting technique" dimension a significant value of $z = 9.48$ ($p = .00$) resulted. Thus, both dimensions affect the average aesthetic judgment significantly.

Furthermore, part-worth utilities for both, the "shape" and the "painting technique" dimension were highest for the lower levels of the attributes whereas they were lowest for the higher attribute levels. Thus, the lower levels of both dimensions were most preferred while the high levels were least preferred.

For offices the goodness-of-fit was *tau* = .67 ($p = .01$). Table 17 (Appendix A) shows the average relative importance, part-worth utilities, utilities and rank order for office stimuli. As can be seen in Table 17 (Appendix A) the average relative importance of the two dimensions indicate that "arrangement" was a little less important for the overall aesthetic judgment than "perceived atmosphere". Both dimensions, "arrangement" ($z = 7.76$, $p = .00$) and "atmosphere" ($z = 10.06$, $p = .00$) turned out to significantly affect the average aesthetic judgment. The part-worth utilities for "arrangement" indicate that offices in which furniture were arranged in a more appealing way (high level of arrangement) were most preferred and offices with a medium level of arrangement were least preferred. Unexpectedly, for "atmosphere" part-worth utilities show that dark offices without a window were most preferred whereas bright offices with window were least preferred.

For car interiors the goodness-of-fit was *tau* = .93 (p = .00). As Table 18 (Appendix A) shows, the average relative importance indicate that "simplicity" was most important for the overall aesthetic judgment whereas "clarity" and "technology" seemed to be similarly important. Again, Kohli's (1988) significant test resulted in significant values for all three dimensions in this object class ("simplicity": z = 7.2, p = .00; "clarity": z = - 3.65, p = .00; "technology": z = -5.25, p = .00). The part-worth utilities for all three dimensions show that the high attribute levels were preferred over the low attribute levels.

Finally, for cutlery the goodness-of-fit was *tau* = .79 (p = .00). As shown in Table 19 (Appendix A), the average relative importance of "proportion" and "perceived ergonomics" indicate that these two dimensions were similarly important for the overall judgment. The "harmony" dimension turned out to be less important. In accordance with this, Kohli's significance test showed that the "proportion" dimension (z = 2.19, p = .01) and the "harmony" dimension (z = -4.6, p = .00) significantly affect the aesthetic judgment. The "perceived ergonomics" dimension however was not significant (z = 0.94, p = .17). The part-worth utilities for all three dimensions indicate that the high attribute levels were preferred over the low attribute levels.

The part-worth utilities can be used to calculate the overall utility for each stimulus. Using a simple additive model, the overall utilities were calculated by simply summing the part-worth utilities. Because part-worth estimates are typically converted to a common scale, a specific constant is added for each model (Hair, 1998). The resulting utilities for each stimulus are shown in Tables 16 -19 (Appendix A). These utilities can then be used to build a rank order of the stimuli within each object class. The resulting rank orders not only take each relevant aesthetic dimension and the relative importance of each dimension for the overall aesthetic judgment into account. They also include the specific value for each attribute level that is used for constructing a stimulus. The last column in Tables 16 - 19 (Appendix A) shows the rank order of the stimuli resulting from the utilities. The outcome is a full rank order of all stimuli of an object class from the most (1) to the least (8 or 9) preferred stimulus. These rank orders can be used as external standards for the aesthetic values of the given stimuli.

5) Discussion

In the research just presented, criteria that were derived from expert interviews and confirmed by multidimensional unfolding studies were used to systematically construct visual stimuli that vary on no other than specific aesthetic dimensions. Using conjoint analysis the average relative importance of each dimension, the value for each variation of a dimension (i.e. the part-worth utility) and for each stimulus (i.e. the utility) was evaluated. The results show that different aesthetic dimensions of an object are differently important for the overall aesthetic judgment. For instance, the most important dimension for car interiors appears to be "simplicity", the second important is "clarity" and the last important is "technology." Conjoint analysis thus revealed a precise picture of the relative importance of dimensions for evaluating the aesthetic value of objects. Furthermore, the part-worth utilities provide the information how the co-occurrence of different attribute levels within a stimulus affect the overall aesthetic judgment for this stimulus.

The calculation of the overall utility for each stimulus results in a specific rank order within each object class. These rank orders can be used as an external standard for evaluating an individual's rank order of the stimuli in future research. For example, the rank order could be used to measure aesthetic sensitivity. The deviation of any individual's rank order from the "optimal" rank order as established in the present study could be evaluated using a Spearman correlation. It could be concluded that the higher the correlation of an individual's rank order with the optimal rank order, the more aesthetically sensitive a person is because the person perceives more differences between stimuli. Because the present stimuli were constructed systematically, differences between individuals in evaluating the aesthetic value of the stimuli can be interpreted as being due to relevant aesthetic dimensions inherent in the stimuli (which individuals may be able to perceive or not), not degree of agreement with experts or non-experts. The external standard proposed here is different from the commonly used criteria in that (a) it is based on knowledge about the properties of stimuli gained from interviews with experts and from multidimensional unfolding studies with non-experts, (b) the relative importance of each aesthetic dimension on the aesthetic judgment is taken into account and (c) it is not a measure that is relative to a certain reference group (such as an average judgment) but rather based on the properties of the stimuli themselves.

Future research must show whether the relative importance and the rank orders identified for the aesthetic dimensions of the different stimulus classes are consistent over different samples. It must also investigate whether the proposed approach results in a valid external standard. Nonetheless, the present research shows that conjoint analysis is a useful technique for aesthetic research. Additionally, when using data on interval level conjoint analysis results could also be used within parametric tests for further analysis, as is widely done in consumer research. For instance, in addition to evaluating main effects and interaction effects between different attributes of a set of stimuli, part-worth utilities could be used in regressions analysis to investigate the relationship between the *importance* of aesthetic properties of stimuli as evaluated by conjoint analysis and individual difference variables such as personality, expertise, and others. In sum, conjoint analysis is a useful tool in aesthetic research.