Establishing a semantic differential on product prototype aesthetics : a research approach

SUSE BACHINGER[1], EVA WIESE[2] & JOHANN HABAKUK ISRAEL[3]

[1)] *University of Amsterdam*
[2)] *Berlin University of Technology, Prometei Graduate School*
[3)] *Fraunhofer Institute for Production Systems and Design Technology, Berlin*

Keywords: visual aesthetics, product prototype, semantic differential, questionnaire construction

Abstract

So far, studies on general judgments of products mainly focused on usability. Only recently, product aesthetics is considered as well. In this context, both information on the visual aesthetics of products and product prototypes are important issues. Several scales exist which measure visual aesthetics of products. However, no scale on the aesthetics of product prototypes has been established yet. In this article, the construction of a valid and reliable German version of a semantic differential on the visual aesthetics of product prototypes is suggested. An initial pool of items shall be developed in an expert workshop. For validation of the scale, 200 participants shall judge product prototypes with the item pool. By statistical analysis preferentially about 10 to 15 items will be chosen for the final scale. Validity will be assessed by looking at correlations of the scale with a 1-item question on beauty and the ability of the scale to differentiate between products differing in aesthetics.

1. Introduction

Visual aesthetics of products attracted interest of researchers only in the past decades. Mainly, focus lies on its influence on usability and general consumer judgment or satisfaction with products. However, systematic research on the aesthetics of products is still scarce (Carbon & Leder 2007; Liu 2003). Additionally, most studies on product aesthetics refer to websites. Although it is, for example for the approach of user-centered design, of special interest to gain information on the aesthetics of product prototypes, this subject has not been systematically researched at all. This

issue shall be discussed in detail within the next paragraphs. A research approach to take a first step to fill that gap shall be put forward.

1.1 Product aesthetics and consumer choice

What could be the reason for you to buy precisely this canapé, or that chair? Its comfort? The amount of money that you have to pay for it? Or its outward appearance? Marketing research is interested in consumer choice for quite some years (Liu, 2003). In this context, ergonomics or human factors mainly focused on aspects of usability and usefulness of products (Carbon & Leder 2007; Lavie & Tractinsky 2004; Liu 2003; Thüring & Mahlke 2007). Only recently, the importance of systematically researching product aesthetics is stressed (e.g. Carbon & Leder 2007; Liu 2003). Research indeed demonstrates the importance of product aesthetics for consumer choice. Hassenzahl (2004) for example mentions beauty as a good, and often the best predictor of products' overall impression or general user satisfaction. Strong associations between product usability, aesthetic judgment and overall product user impression and satisfaction were found in several studies (e.g. Ben-Bassat et al. 2006; Hassenzahl 2004; Thüring & Mahlke 2007; Tractinsky et al. 2000), resulting in different models regarding the causality of those relations. On the one hand, Tractinsky (2000) declares that beauty influences users' impression of usability: he postulates "What is beautiful is usable". Hassenzahl's (2003, 2004) results on the other hand contradict this theory: aesthetic quality and usability were found to be independent. Both contributed to the perception of the overall judgment of a product. Based on his findings, Hassenzahl (2003) proposes a model for interactive products in which users' overall judgments of products are influenced by two different attribute groups of products: "pragmatic" and "hedonic" attributes. Pragmatic attributes are connected with the achievement of behavioral goals and therefore include utility and usability. Hedonic attributes on the other hand are related to the user's self. They can be subdivided into "stimulation" and "identification". "Stimulation" describes the degree of novelty and challenge of the product and "identification" refers to the ability to express yourself through the product. However, no clear empirical support for one particular theory is found. So, the reasons for you to buy some product are probably formed by a variety of attributes- but which they exactly are, how important each of them is and how they interact still remains vague.

It would be interesting to gain knowledge on whether people differing in personality traits judge aesthetics differently – and if this is the case, which patterns exist. Knowledge on this subject would improve the understanding of aesthetics in general. However, research on associations between aesthetic perception and personality traits do not reveal many clear results. Thielsch (2007), for example, found that the older the person, the more positive the judgments on the visual aesthetics of a website. Bloch's (2003) findings indicated that high aesthetic sensitivity was associated with a more positive judgment of a product. Contrawise, no associations of aesthetic judgments and gender, education, profession, internet experience, Big-Five personality traits or aesthetic sensitivity were found by Thielsch (2007). The contrary findings on a possible association of aesthetic sensitivity and visual aesthetics may be due to the fact that websites are judged differently from other products. A considerable part of research on product aesthetics focuses on the perception of websites (e.g. Pandir & Knight 2006; Schaik & Ling 2009; Thielsch 2007; Tractinsky et al. 2006) and products in general at the moment. Research on judgments of product prototypes is lacking. This surprises because it seems a relevant subject, especially in the light of user centered design (UCD).

1.2 Judgments of product prototypes

Hall (2001) suggests UCD as a good approach for successful product design. It requires designers to analyze and foresee how users are likely to use and judge a product or interface (Gulliksen et al. 2003). Therefore, designers have to test the validity of their assumptions with regard to user behavior and preferences. Revelations of valid and reliable judgments on visual aesthetics of product prototypes are of particular importance because those form a necessary tool to improve products according to the user's needs. It is of special interest to find out whether the outer design will please users as it can relatively easily be changed. Thereby, it would provide an option to compare different prototypes. The design which is accepted by most users could be identified.

Another reason for interest in product prototypes comes along with the point addressed above: the general possibility to monitor and improve product prototypes in progress. The judgments of prototypes differing in information or level of preciseness could be compared and one could for example establish how much details are needed in order to be able to make valid judgments about a product generally. Surprisingly, so far no standardized questionnaire is established which provides the opportunity to gain information on the aesthetics of product prototypes.

There are some further problems with the current state of affairs such as the often vague or inconsistent definition of the concept visual aesthetics. This will discussed subsequently.

2. Assessment of visual aesthetics

2.1 Definition of visual aesthetics

What becomes evident when reviewing literature on aesthetics is the amount of distinct concepts that all refer to the description of the visual appearance of products: e.g. "attractiveness", "beauty", "goodness", "creativity", "pleasure" and "aesthetics" (e.g. Hassenzahl 2004; Lavie & Tractinsky 2004). Concepts are not used consistently across researches and are rarely defined specifically (Thielsch 2007). Considerable overlap between them can be assumed. This is seldom explicitly addressed. Often, visual aesthetics is assumed to be identical to visual beauty (Thielsch 2007). The concepts visual beauty and visual aesthetics can be assumed to correlate highly.

In this study, visual product aesthetics is defined, in line with Thielsch (2007) and Leder et al. (2004), as the subjective positive impression you get when looking at the product which is followed by positive emotional and cognitive reaction. Likely, it is a multidimensional concept (Lavie & Tractinsky 2004; Thielsch 2007). This means that it is probably formed by separate, underlying variables. Examples could be color, complexity, texture etc. Yet, these variables have not been identified conclusively so far.

2.2 How to measure visual product aesthetics

Visual aesthetics of products is obviously an important issue. But how can this concept be measured? The current solution often seems to be a single item on aesthetics or beauty, such as "How beautiful is this pattern/product?" (Tinio & Leder, 2009). Also, some scales on the issue exist. Hassenzahl developed, based on his theory men-

tioned above, a semantic differential on pragmatic and hedonic attributes and attractiveness of interactive products which is widely used: the AttracDiff 2 (Hassenzahl 2003). With 28 contrary adjectives, this instrument assesses attractiveness as well as pragmatic and hedonic attributes of interactive products. Aesthetics is not explicitly referred to in this model but might be found in the hedonic subscales. A semantic differential (developed by Osgood et al. 1957, cited by Judd et al. 1991) comprises contrary pairs of adjectives judged on a seven-point rating scale. Originally, the scale was developed in order to measure the meaning of an object to an individual. An example of an original pair of adjectives would be "good-bad". Additionally, there are a number of questionnaires which quantify specifically the visual aesthetics of websites (Lavie & Tractinsky 2004; Visual aesthetics of Website Inventory (VAWI) developed by Moshagen 2005, cited by Thielsch 2007).
In sum, there exist several standardized questionnaires on the visual aesthetics of products and websites. However, no standardized questionnaire on visual aesthetics of product prototypes is available yet although is of particular interest to gain more knowledge on this subject, e.g. when applying user centered design (UCD). Therefore, the development of a new scale on the visual aesthetics of product prototypes is suggested.

2.3 Possible strategies for the construction of the scale

Two strategies for the construction of the scale were taken into consideration. One could, on the one hand, use a well-researched questionnaire on visual product aesthetics as starting point, adapt it and then test its generalizability onto product prototypes. On the other hand, one could construct an entirely new scale on visual product aesthetics, based on an expert workshop for generating new items which specifically match the specific characteristics of product prototypes.
Two widely used scales were taken into account as possible starting points by the authors: the AttrakDiff2 (Hassenzahl, 2003) and the scale on visual aesthetics of Lavie & Tractinsky (2004). The former is developed based on the theory of Hassenzahl mentioned before. Each subscale (stimulation, identification and pragmatic) was, in accordance with the theory, established and validated separately. To our knowledge, the total structure of the scale was examined assuming the verity of the theory. No inspection was performed on whether the theoretically assumed subscales could really be found in the data e.g. by performing a factor analysis on all adjectives and not only the subscales. His theory is, however, not satisfactorily validated yet and it could therefore be biased. This would lead to a biased scale. Also, the scale does not exclusively focus on the aesthetics of products. As a consequence, the authors decided that it would be problematic to use this scale as a starting point. The latter was constructed independently of any theory and would thereby form a better option as starting point. However, this scale focuses on websites. Many items are mainly addressing aspects of websites and the generalizability to other products is not confirmed yet. Possible usage for product prototypes remains questionable.
In sum, the addressed issues led the researchers to choose for the construction of a new scale on the aesthetics of product prototypes rather than the adaptation of an existing scale. Adjectives established in an expert workshop shall provide an initial pool of items. In the next step, about 200 persons shall be asked to judge products with this pool online. Based on the correlations of these judgments with the one item on aesthetics, factor analysis and reliability analysis, a certain amount of pairs of adjectives (aimed are about 10 to 15) will be chosen for the final scale. The proce-

dure of item generation, -selection and –verification and the proposed statistical method is described in detail bellow.

3. Method

3.1 Item acquisition

Items established in an expert workshop will be used as initial item pool for further validation.

3.1.1 Expert workshop

An expert workshop with five experts on designing will provide an item pool for the scale. Four product designers, one professor on product designing and one web designer with more than four years of working experience in the field will take part in the workshop. Experts on designing were chosen because of their knowledge on visual aesthetics. Bipolar adjective pairs will be established in a workshop of approximately three hours. First, a definition of aesthetic will be provided. In the next phase, each expert will produce semantic differentials individually. Those will be read out loud and immediate feedback can be given. This should lead to adaptation and generation of new items. In the final phase, all items are read out and each participant can use veto if the item doesn't seem good to him/her. Only the items without veto will be used for the initial version of the scale. Examples for the established items would be: "intuitiv-rational" (intuitive-rational), "fragil-stabil" (fragile-stable) and "interessant-langweilig" (interesting-boring).

3.2 Stimulus material

In a former study (Wiese et al., 2009), 12 student designers created different types of lamps in a student project. In a pilot study, those lamps are judged with regard to their aesthetics by 10 participants (judged on two single items on beauty and aesthetics on a seven point scale). The most and least aesthetic lamps and one neutral lamp are chosen as stimulus material (*see for examples figure 1*).

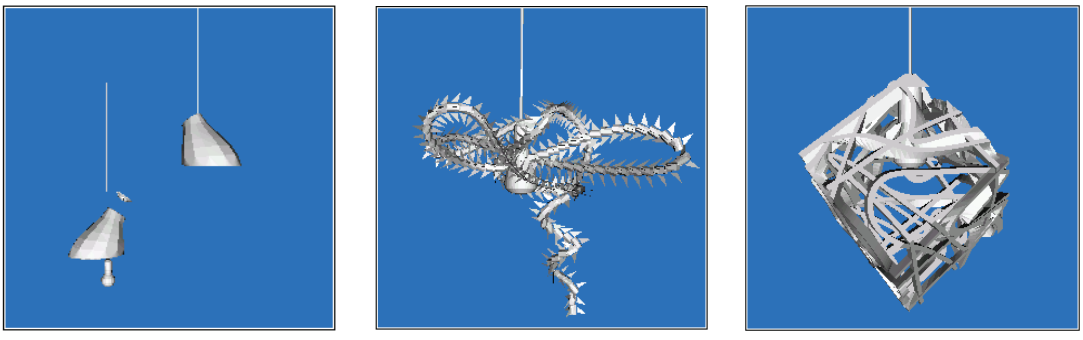

Figure 1: possible stimuli

3.3 Online version of the scale

The online version of the initial scale will be provided with the program "limesurvey" (open source application provided by "The LimeSurvey Project Team", 2003). Participants shall be asked to take part in this research via email (with a link to a

website with the scale). The time to answer all questions should not exceed 15 minutes. First, demographical information will be asked, then the lamps will subsequently be shown and judged with the adjectives and one-item questions on beauty and aesthetics. The one item questions will be used to ensure the validity of the questionnaire. For the final scale, only items will be chosen which correlate highly with aesthetics in the first place. The final scale will be correlated with the item on beauty for convergent validity- beauty is expected to correlate highly with aesthetics. If the 15 minutes are not exceeded, the 11 item scale on aesthetic sensitivity (CVPA-g; Bloch et al. 2003, German version by Moshagen & Thielsch 2007) will be administered as well.

3.4 Statistical analysis

3.4.1 General approach

For the construction of a scale, statistics forms a valuable tool to select the best items out of a large item pool and it also gives information on whether those selected items can be expected to form a good scale. In this case, about 10 to 15 items shall be picked for the final scale. Several criteria for choosing the best items are well-established, such as the loadings of the items on latent factors (identified by factor analysis) and reliability analysis such as Cronbachs's alpha, inter-item and item-rest correlation.

First of all, the correlations of the individual items with the item on aesthetics will be examined. Only items which correlate highly (e.g. $r > .5$) will be used for further analysis. Subsequently, the structure of the scale will be looked at with an explorative factor analysis (EFA) with the statistical software Lisrel (Jöreskog & Sörbom 1993). EFA is a statistical method used to describe variability among observed variables in terms of fewer unobserved variables called factors. The observed variables are modeled as linear combinations of the factors, plus "error" terms. The information gained about the interdependencies can be used to reduce the set of variables in a dataset. For the scale, a clear structure of one or more factors is aimed. The factors should be explicitly identified by the items which load highly on them and should be interpretable, which means that it should be obvious which latent variable each of them describes. Therefore, items which load on more than one factor, which load on factor/s which shall not be included in the scale (for example because they are not interpretable), which load only weakly on one factor and which load on more factors can be left out. Crossvalidation will be used to ensure that the found structure is no coincidence: the structure identified with EFA on one half will be checked with a confirmatory factor analysis (CFA) on the second half of the data. Only if the CFA supports the results found with the EFA, good quality of the structure of the scale is probable. To further ensure the quality of the structure of the scale, it should also be the same for all three lamps. Therefore, factor analysis will be performed for each lamp separately as well.

In order to establish the reliability of the scale, Cronbachs's alpha, inter-item and item-rest correlation will be looked at. Cronbach's alpha measures how well a set of variables or items measures a single, unidimensional latent construct. As minimum value, .80 is chosen which indicates good reliability. If more factors are found, the same should apply for the subscales. Aiming at getting Cronbach's alpha as large as possible, items which decrease it can be erased. Also, items which do not have the inter-item and item-rest correlations aimed at can be left out because they fail to measure the same concept as the rest of the scale and/or they fail to differentiate be-

tween different degrees of aesthetics. Mean inter-item correlation should be between .20 and .40 in order to be optimal or between .10 and .50 in order to be acceptable, and item-rest correlations should be above .20.
Regarding convergent validity, Spearman's correlations of the scale total score with the item on beauty will be considered. It should be significant and positive. Furthermore, the ability of the scale to differentiate between the most aesthetic and the least aesthetic lamp will be examined with a t-test. There should be a significant difference between the judgments of the two lamps- otherwise whether the scale measures what we intend to measure should seriously be disputed.

3.4.2 Hypotheses

So, the precise hypotheses for the finale scale are:

1) The individual items should correlate highly with the item on aesthetics.
2) Cronbach's alpha should be very good (if more factors, this should apply for the subscales). As minimum value, .80 is chosen.
3) Mean inter-item correlation should be between .20 and .40 in order to be optimal or between .10 and .50 in order to be acceptable, and item-rest correlations should be above .20.
4) The final scale should have a clear structure (preferably unidimensional, if more dimensions they should be interpretable) which should be verified by confirmatory factor analysis (using crossvalidation).
5) The structure of the scale should be the same for each lamp.
6) Validity should be indicated by: positive significant correlation of the total score of the scale with the item on beauty. The scale should be able to differentiate between the most aesthetic and the less aesthetic lamp.

Exploratively, the correlations of the total score of the scale with aesthetic sensitivity (CVPA-g) and age will be looked at.

4. Conclusion

As presented above, a gap in research on product aesthetics is obvious: only few empirically validated scales on the visual aesthetics of products and no scale at all on the aesthetics of product prototypes are established so far. Valid and reliable judgments on the aesthetics of product prototypes are of special interest in order to obtain well-grounded information on prototypes, e.g. for UCD approaches. Also, they might provide the possibility to systematically compare between different prototypes or between different phases of the same prototypes. This might improve the effectiveness of the UCD approach. In this article, a research approach for the construction of a new scale is suggested. A semantic differential which describes the aesthetics of product prototypes shall be established and subsequently validated. This might form a first step to fill that gap by gaining empirically grounded knowledge on judgments of product prototypes. In order to generalize results, the psychometric properties of the scale will have to be examined for other products than lamps. This shall form a necessary next step to ensure the quality of the scale.

References

Ben-Bassat, T., Meyer, J., and Tractinsky, N. (2006). Economic and Subjective Measures of the Perceived Value of Aesthetics and Usability. *ACM Transactions on Computer-Human Interaction, 13 (2),* 210–234.

Bloch, P.H., Brunel, F.F., & Arnold, T. (2003). Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement. *Journal of Consumer Research, 29,* 551-565.

Carbon, C.-C., & Leder, H. (2007). Design Evaluation: From Typical Problems to State-of-the-Art Solutions. *Thexis, 2,* 33-37.

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, A. (2003). Key principles for user-centred systems design. *Behaviour & Information Technology, 22,* 397-409.

Hall, R.H. (2002). Prototyping for usability of new technology. *Int. J. Human-Computer Studies, 55,* 485-501.

Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human–Computer Interaction, 13(4),* 481–499.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Zwillus (Eds.), Mensch & Computer 2003. *Interaktion in Bewegung* (pp. 187-196). Stuttgart, Leipzig: B.G. Teubner.

Hassenzahl, M. (2004). The interplay of beauty goodness and usability in interactive products. *Human Computer Interaction, 19,* 319–349.

Hassenzahl, M., Schöbel, M., & Trautmann, T. (2008). How motivational orientation influences the evaluation and choice of hedonic and pragmatic products: The role of regulatory focus. *Interacting with computers, 20,* 473-479.

Hekkert, P. (2006). Design aesthetics: principles of pleasure in design. *Psychology Science, 48 (2),* 157–172.

Jöreskog K., & Sörbom D. (1993). *LISREL 8 Structural equation modelling with the SIMPLIS command language.* Erlbaum, Hillsdale, NJ.

Jordan, P.W. (1998). Human factors for pleasure in product use. *Applied Ergonomics, 29 (1),* 25–33.

Judd, C.M., Smith, E.R., Kidder, L.H. (1991). *Research Methods in Social Relations (6ths ed.).* Harcourt Brace Jovanovich College Publishers, USA.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of Web sites. *International Journal of Human-Computer Studies, 60,* 269–298.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95,* 489-508.

Liu, Y. (2003). Engineering aesthetics and aesthetic ergonomics. Theoretical foundations and a dual-process research methodology. *Ergonomics, 46,* 1273–1292.

Pandir, M., & Knight, J. (2006). Homepage aesthetics: The search for preference factors and the challenges of subjectivity. *Interacting with computers, 18,* 1351-1370.

Schaik, P., & Ling, J. (2009). The role of context in perceptions of the aesthetics of webpages over time. *International Journal of Human-Computer Studies, 67,* 79–89.

The LimeSurvey Project Team. (2003). http://www.limesurvey.org.

Thielsch, M.T. (2007). *Ästhetik von Websites. Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen.* Münster: Verlagshaus Monsenstein und Vanerdat.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology Interaction. *International Journal of Psychology,* 42 (4), 253–264.

Tinio, P.P.L., & Leder, H. (2009). Just how stable are aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica, 130,* 241-250.

Tractinsky, N., Katz, A.S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13(2),* 127–145.

Tractinsky, N. (2004). A few notes on the Study of Beauty in HCI. *Human-Computer Interaction, 19,* 351-357.

Tractinsky, N., Cokhavi, A., Kirschenbaum, M., & Sharfi, T. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. *International Journal of Human Computer Studies, 64,* 1071–1083.

Wiese, E., Israel, J. H., Zöllner, C., Pohlmeyer, A. E., & Stark, R. (2009). *The Potential of Immersive 3D-Sketching Environments for Design Problem-Solving.* In: Tagungsband 13th International Conference on Human-Computer Interaction HCI 2009. San Diego, CA, USA, 485-489.