Changing only the aesthetic features of a product can affect its apparent usability

Andrew Monk and Kira Lelos

Centre for Usable Home Technology (CUHTec), Department of
Psychology, University of York, YO10 5DD, UK

http://www.cuhtec.org.uk
A.Monk@psych.york.ac.uk

Abstract. Three experiments were conducted to investigate the relationship between usability and aesthetics with students and older people. A common mechanical domestic appliance, the can opener, was chosen as a proxy for future digital products. The experiments involved comparing the rated usability of can openers that had been painted to make them more or less aesthetically pleasing. Experiment 1 tested students' ratings of beauty and usability. Experiment 2 similarly tested an elderly population on their ratings before and after use. In general, the products rated more beautiful were rated as more usable. To avoid the possibility that rating a product for its aesthetic qualities could somehow affect its subsequent rating for usability, Experiment 3 repeated Experiment 2 but products were only rated for usability. In Experiments 1 and 3 the manipulation of product features associated only with aesthetic qualities of the product (painting the can openers) also significantly affected ratings of usability. The results are related to Hassenzahl's model of user experience, and interpreted in terms of the holistic evaluation of product features in judgements of hedonic and pragmatic attributes. The results confirm and extend previous findings and highlight the importance of aesthetic considerations as well as usability in all forms of design.

1   Introduction

Home oriented IT products bring to the fore different user requirements to those traditionally considered for work oriented IT products [1]. In particular, requirements related to the aesthetic qualities of objects we bring into our homes are quite different to those in the workplace. This has led to an upsurge in research on beauty as a topic within the Human Factors and Human-Computer Interaction (HCI) communities. One theme within this research has been the relationship between aesthetic quality and the traditional requirement for usability. An early study by Kurosu and Kashimura [2] demonstrated a strong correlation between ratings of

usability and ratings of beauty. They had 156 students rate 26 layout drawings for a bank ATM. Averaging over participants and correlating across layouts they obtained a correlation of 0.6 between ratings of usability and beauty. This finding is potentially important because it questions the traditional conception of usability in Human Factors and HCI which is based on performance measures such as time to completion or learnability. If two products are similar in objective usability but the more attractive product is judged more usable then it would seem that the user has a different concept of usability to that implied by the objective criteria. This questioning of the concept of usability has resonance with a new emphasis within the HCI community on emotional response (e.g., [3]) and user experience (e.g., [4]).

Of course, the Kurosu and Kashimura [2] result is open to criticism, in particular, that the participants rating these products had not actually used them. With little else to go on, it is possible that they fell back on attractiveness when making their ratings of usability. This possibility was addressed by Tractinsky et al. [5]. Nine ATM layouts from the [2] study were selected as high, medium or low aesthetic quality. Screen simulations of ATM functionality were added so that participants could use them to simulate eleven tasks such as withdrawing cash or making an account enquiry. The ATMs were rated for usability and beauty before and after use and the post-usage ratings yielded the same apparent effect of aesthetic quality on usability ratings as the pre-usage ratings.

Both [2] and [5] are open to the further criticism, that one cannot assign a causative interpretation to a correlation [6]. The high aesthetic quality layouts were selected from the complete set on the basis of aesthetics ratings. It could be that this selection of layouts are indeed more usable in terms of their objective ease-of-use and ease-of-learning. The experiments to be described here address this issue by manipulating the appearance of otherwise identical products. A causative link would be demonstrated if this manipulation has an effect on perceived usability.

1.1   Hassenzahl's model

Hassenzahl [7] makes a plea for a more explicit model of what is going on in the studies described above. His model will be used to frame the experiments described below. He sets out three entities: Product Features; Apparent Product Character and Consequences (see Figure 1, A). Product Features are the results of design, content, presentation, functionality, etc. The Apparent Product Character is the user's perception of the products. Finally, the Apparent Product Character leads to Consequences, the product's general appeal ("goodness"), pleasure in its use, and behavioural consequences such as time spent using it. Consequences depend on the situations the users find themselves in, for example, a product that is a pleasure to use at work may not be judged so at home (Hassenzahl's diagrammatic depiction of the effect of context has been left out of Figure 1 for the sake of simplicity).

A.

B.
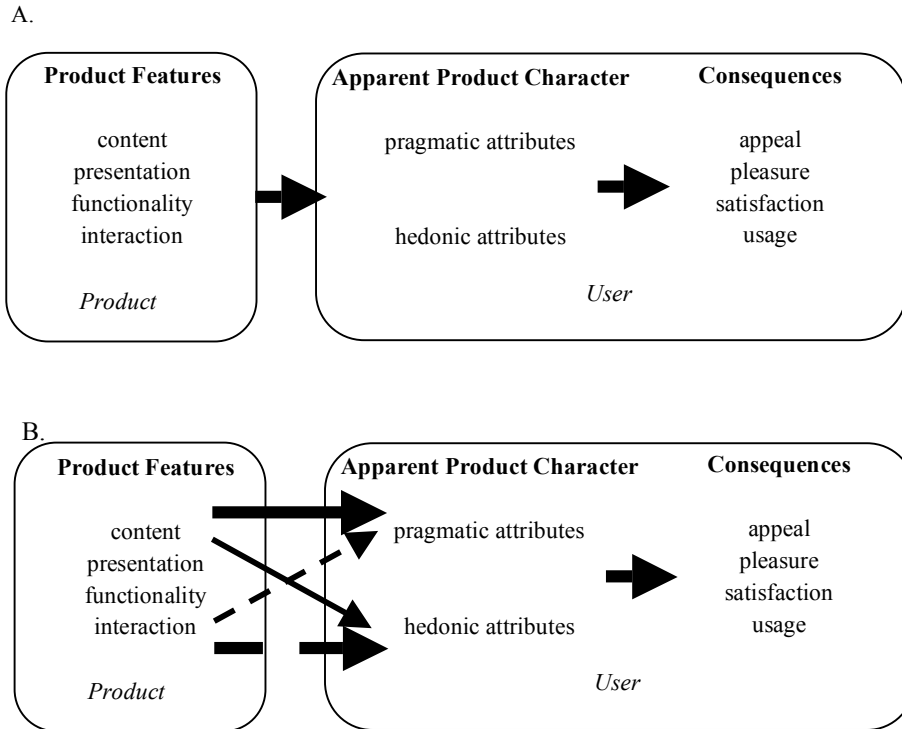
Fig 1. A. Hassenzahl's (2003) user experience model. B. Proposed causative relationships, dotted arrows represent a prediction of the effect of the aesthetic manipulation used here, solid arrows represent the effect of Tractinsky et al.'s (2000) manipulation of usability.

Hassenzahl's [7] model assumes that when a participant in an experiments is asked to rate a design they imagine themselves in a particular situation and make some evaluation of Apparent Product Character and then what the Consequences of this judgement might be. He has encapsulated this theory in a questionnaire (see [6]) measuring different aspects of Apparent Product Character in terms of hedonic and pragmatic quality. Pragmatic quality corresponds to usability. Hassenzahl [6] provides support for his model using data from a study using a range of MP3 player skins. Support for the independence of three elements of product character, hedonic-stimulation, hedonic-identity and pragmatic quality, is provided by identifying four skins that vary quite differently on these three dimensions.

The experiments described below directly manipulate Product Features associated only with aesthetic qualities of a product, that is, its hedonic attributes. The hypothesis is that this manipulation will also affect its pragmatic attributes supporting a holistic view of the evaluation of hedonic and pragmatic attributes in the perception of Apparent Product Character (see dotted arrows in Figure 1, B). Note participants in these experiments were asked to rate the extent to which they agreed with the statement "this can opener is easy to use" rather than the seven semantic differentials suggested by Hassenzahl to measure pragmatic quality. However, Hassenzahl's pragmatic quality score has been shown to correlate highly

with "hard to use - easy to use" and in Hassenzahl's terms is part of the Product Character rather than a Consequence.

The approach of directly manipulating specific product features was inspired by the Tractinsky et al. [5] study which manipulated usability. Some of the screen simulations were made less usable by imposing a 9 second system delay and keys that only worked the second time they were clicked. These simulations were rated as less beautiful demonstrating a causative relationship between usability and beauty ("usable is beautiful") but have nothing to say about the alternative causative relationship ("beautiful is usable"). The explanation of this finding is also depicted in Figure 1 , B.

1.2   The can opener as a proxy for future digital products

As stated above, the purpose of the experiments described here is to demonstrate that the direct manipulation of aesthetic Product Features affects ratings of usability when the product is otherwise unchanged. All of the studies described above used screen simulations of products, ATM designs rendered as line drawings and decorative graphic designs in the form of MP3 player skins. Hassenzahl [6] criticises the use of ATM designs because they are not objects owned by the user. He also criticises the use of engineering students as being unrepresentative of the general public. To give a strong test of our hypothesis that the manipulation of aesthetic Product Features can affect the experience of usability, we needed a product where aesthetics and usability were both important to the participants. To make aesthetics important it needed to be a product that they might own and keep in their homes. To make usability important it needed to be a product whose function could be understood readily by the participants, which in these experiments were psychology students and members of the general public attending a drop in centre for older people. It was hard to find an electronic product that (a) met these requirements and (b) could easily be manipulated (rather than selected) to look more or less visually appealing. For this reason we chose to use can openers as a proxy for the many portable electronic devices that are gradually making their way into our homes (e.g., phones, music players and hand held web browsers). Two of each of four models of can opener were purchased ranging from the simplest and cheapest to more expensive ergonomically designed models (see Figure 2 at end of paper).

Enamel paint was applied to alter the aesthetic qualities of the can openers. Pre-rating of the can openers by a small sample of students showed that, when unmodified by painting, Model 1 was rated very low given the statement "this can opener is appealing to look at" whereas Model 4 was rated very highly. Accordingly, the handles of the modified version of Model 1 (row 1 column 2 in Fig. 2) were painted red (Metallic Deep Red) to make it more attractive compared with the unpainted version. Models 2, 3 and 4 (rows 2-4 column 1)were painted a rather unpleasant blotchy green (Pea Green) to make them unattractive. Model 4 (row 4 column 2) was left unpainted in the attractive condition, Models 2 and 3 (rows 2 and 3 column 2) were painted red. It is difficult to see how this manipulation could have affected the objective ease-of-use of the can openers.

2. EXPERIMENT 1

The general procedure in each of the three experiments described here is similar. Each participant rated either the four attractive, or the four unattractive, can openers. This way the participant saw only one version of each model and was never asked to directly compare the attractive and unattractive conditions. This was to minimise the possibility of the participant guessing the purpose of the experiment. To test the effectiveness of the manipulation participants rated the statement "this can opener is appealing to look at". They were also asked to try the can openers with some washed empty food tins before rating the critical statement "this can opener is easy to use".

2.1   Method (Experiment 1)

Participants
These were 20 male and female undergraduate students from the University of York studying various subjects, and ranging in age from 19 to 26. They were not rewarded for their participation.

Procedure
Each participant was randomly assigned to one of the two groups Ugly, rating the four ugly can openers, and Pretty rating the others. Each of the four models was presented to each participant in a randomised order. Half of the participants rated aesthetics first followed by usability, while the other half rated usability first followed by aesthetics. For aesthetic ratings, the participants were able to hold and study the can opener, while for usability ratings the participants were instructed to use the can opener on a tin by turning the handle and partially opening it. For the aesthetic ratings they were read the statement "this can opener is appealing to look at" and given a card with a five item Likert scale printed on it where "1" was labelled "strongly disagree" and "5" strongly agree". For the usability ratings the statement read was "this can opener is easy to use".

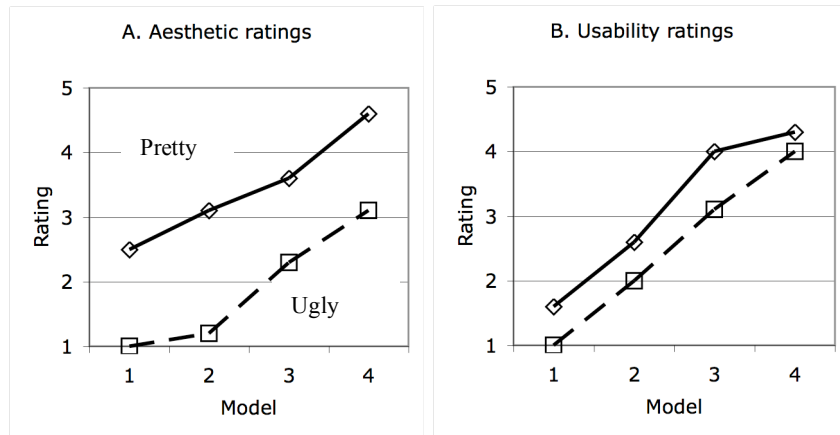### 2.2   Results (Experiment 1)

Manipulation check - aesthetics ratings
If the manipulation of Product Features has been successful one would expect the aesthetics ratings of the Pretty Group to be higher than those of the Ugly Group. This is a strong test of the manipulation as the participants were not able to directly compare the two versions of each model. The results confirm that painting the can openers had the desired effect. The overall mean aesthetic rating of the Ugly Group was 1.90 (Std. Dev. 0.13) and that for the Pretty Group 3.45 (Std. Dev. 0.13). A split-plot analysis of variance where the between subjects effect was Group (2 levels) and the within subjects effect Model (4 levels) showed a significant main effect of Group ($F_{(1, 18)} = 75.209$, $p < 0.05$) and Model ($F_{(3, 54)} = 36.818$, $p < 0.05$) but no significant interaction.

Criterion variable - usability ratings

The critical comparison comes from the usability ratings. As predicted, these closely mirror the aesthetic ratings. The difference between the overall means for the two groups is smaller but still very reliable ($F(1, 18) = 13.157$, $p < 0.05$). The overall mean rating of the Ugly Group being 2.53 (Std. Dev. 0.12) and that for the Pretty Group 3.13 (Std. Dev. 0.12). Again, there was a significant main effect of Group ($F(1, 18) = 75.209$, $p < 0.05$) and Model ($F(3, 54) = 88.105$, $p < 0.05$) and no significant interaction. Manipulating the aesthetic Product Features of the can openers alone, i.e., leaving other Product Features normally associated with usability unchanged, had a direct effect on the Apparent Product Character usability as measured by our rating scale. It would appear that the predictions of the model depicted in Figure 1, B (dotted arrows) are supported. The experience of usability is influenced by aesthetic Product Features as well as those Product Features associated with objective definitions of usability such as time to completion or task completion.

Another notable feature of these results is the way that the usability and aesthetics ratings strongly parallel one another. In both cases there is a large effect of Model with the cheapest, least sophisticated models being rated lowest and the more expensive and more sophisticated models being rated highest (see Figure 3, A and B). Again, this is consistent with a holistic judgement of Apparent Product Character where Product Features interact when a judgement is made. Alternatively, it may be that we have just happened upon a set of products with Product Features that co-vary in this way.

Fig. 3. Experiment 1, students' ratings of Aesthetics and Usability for each model from Pretty and Ugly groups

3.   EXPERIMENT 2

In order to test the generality of the results from Experiment 1 a new user population was selected. To make this as different as possible from the students used in

Experiment 1, and most experiments in this area, this was people visiting a drop-in centre for the over 60s. We expected this to make the usability features more salient. The more expensive models were designed for people with poor grip and hence older people were expected to appreciate them more. In addition, it was decided to add a pre-use rating of the can openers to make the experiment directly comparable to [5].

3.1   Method (Experiment 2)

Participants
Participants were 32 citizens from the St. Sampson's Drop-In Centre located in York. They were a mixture of male and female (22 female and 10 male) and various ages from 60 years to above 80 years (modal range 60 to 65). They were not rewarded for their participation. Permission was gained in writing from the manager of the centre and verbally from the participants.

Procedure
The procedure used was the same as in Experiment 1 except that participants rated the aesthetics and usability of each of model, then used them, then rated them again for both aesthetics and usability. The order of rating (usability-aesthetics or aesthetics-usability) and the order the models were presented were both counterbalanced.

3.2   Results (Experiment 2)

Manipulation check - aesthetics ratings
Again the overall mean aesthetic rating of the ugly group was significantly lower (2.44, Std. Dev. 0.15) than that for the pretty group (2.98, Std. Dev. 0.15). A split-plot analysis of variance was carried out where the between subjects effect was Group (2 levels) and there were two within subjects effects, Time of Test (2 levels, before and after) and Model (4 levels). This showed a significant main effect of Group ($F_{(1, 30)} = 6.354$, $p < 0.05$). While statistically significant, this effect was much smaller for this population of older people than it was for the students in Experiment 1 (0.54 scale points rather than 1.55 respectively). This could be because the colours used were selected using ratings from students, i.e., they were not colours that were particularly attractive or unattractive to older people. Another possibility is that older people are less influenced by colour in general.

Time of Test was also significant ($F_{(1, 30)} = 29.138$, $p < 0.05$) and Model ($F_{(3, 90)} = 17.347$, $p < 0.05$). The Time of Test effect was due to the second set of ratings after using the product being slightly higher than the before ratings. Much to our surprise the very significant effect of Model was in exactly the reverse direction to that observed with the student raters. Model 1 the cheapest and simplest product was rated highest and model 4, the most expensive and ergonomic lowest (see Figure 3, A). There were no significant two- or three-way interactions.

Criterion variable - usability ratings

Usability ratings again closely mirror the aesthetic ratings, except that there was no significant effect of Group. The overall mean usability rating of the Ugly group was only marginally lower (2.87, Std. Dev. 0.13) than that for the Pretty group (2.98, Std. Dev. 0.13). The analysis of variance applied to the aesthetic ratings was applied to the usability ratings. This showed a non-significant main effect of Group ($F(1, 30) < 1$, n.s.), but significant effects of Time of Test ($F(1, 30) = 28.200$, $p < 0.05$) and Model ($F(3, 90) = 21.434$, $p < 0.05$). There were no significant two- or three-way interactions.

As with the aesthetic ratings the Time of Test effect was due to the second set of ratings after using the product being slightly higher than the before ratings. The effect of Model in the usability ratings closely followed those of the aesthetic ratings, i.e., they were also the reverse direction to that observed with the student raters (see Figure 4 A and B). Model 1 the cheapest and simplest product was rated highest and model 4, the most expensive and ergonomic lowest.
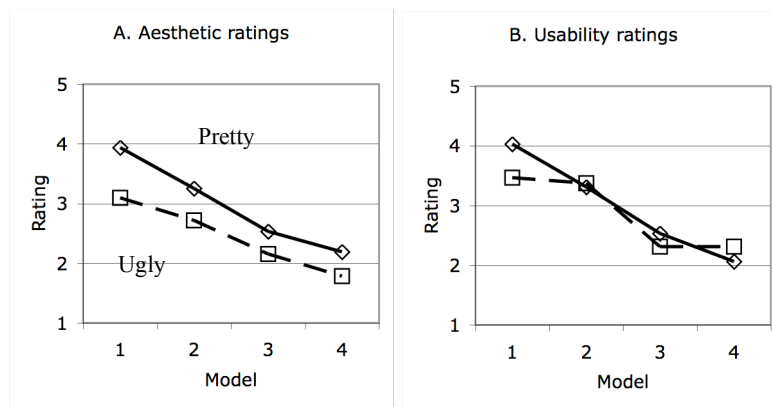
Fig. 4. Experiment 2, older people's ratings of Aesthetics and Usability for each model from Pretty and Ugly groups averaging across before and after.

Experiment 2 then does not replicate the key result found in Experiment 1. There was not a significant effect of Group. Of course, this is a null result. It could be that there is no effect of manipulating aesthetic Product Features on usability ratings with this user population. Alternatively, it may be that the experimental design was not sensitive enough to detect such an effect. As noted above, while still statistically significant, the effect of painting the can openers on aesthetic ratings was very much smaller than in Experiment 1.

The reversal of the effect of model and the way it is seen both in aesthetic and usability ratings is noteworthy. One hypothesis is that the older people were strongly influenced by the familiarity of the models considered. Model 1 is commonly found in many homes while the more expensive designs represented by the other models less so. The higher ratings for model 1 may reflect subjective estimates of the objective usability criterion, time-to-learn. Many older people have a quite reasonable scepticism about learning to use unfamiliar tools that would not be a concern to students. Whatever the cause of this reversal of effect it is most

interesting that it is reflected equally in the aesthetic and usability ratings, further bolstering the case for holistic judgements of Apparent Product Character from a variety of Product Features.

4.   EXPERIMENT 3

It is possible that asking for ratings for aesthetics and usability for the same object somehow confused the participants in our experiment. To simplify the procedure we kept the before-after element but only asked for ratings of usability. The same set of can openers were used.

4.1   Method (Experiment 3)

Participants
A new set of 32 citizens were recruited from the St. Sampson's Drop-In Centre who had not participated in the previous study. The participants were a mixture of male and female (19 female and 13 male) and various ages from 60 years to above 80 years old (modal range 66 to 70).

Procedure
The procedure used in this experiment was the same as in Experiment 2, except each participant was only asked about usability and was not questioned about aesthetics.
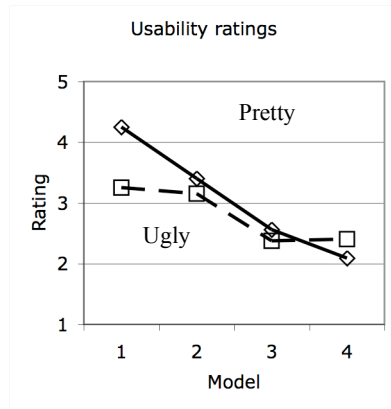
4.2   Results (Experiment 3)

Criterion variable - usability ratings
There were no ratings of aesthetics in this experiment. The usability rating confirm the results from Experiment 1. The overall mean usability rating of the Ugly group was significantly lower (2.80, Std. Dev. 0.08) than that for the Pretty group (3.08, Std. Dev. 0.08). The analysis of variance used in Experiment 2 showed a significant main effect of Group ($F(1, 30) = 6.551$, $p < 0.05$), a non-significant effect of Time of Test ($F(1, 30) = 2.899$, n.s.) and a significant effect of Model ($F(3, 90) = 36.882$, $p < 0.05$). The effect of Model was as in Experiment 2: Model 1, the cheapest and simplest product, was rated highest and model 4, the most expensive and ergonomic, lowest. However, in these data there was a significant Group by Product interaction ($F(3, 90) = 5.510$, $p < 0.05$), see Figure 5. It would seem that the effect of Group is mainly seen in the ratings for Model 1, the simplest and aesthetically preferred can opener for this group. This interpretation is confirmed by a simple main effects analysis which show a significant effect of Group for Model 1 ($F(1, 120) = 19.248$, $p < 0.05$) but none of the other models.

This then is a partial replication of the key effect observed in Experiment 1 for the model showing the largest effect of the aesthetics manipulation as evidenced in the aesthetics ratings in Experiment 2. Manipulating the aesthetic Product Features

of this model and leaving other Product Features normally associated with usability unchanged, had a direct effect on the Apparent Product Character usability. The result is striking given: (i) the limited nature of the manipulation as evidenced by the manipulation check in Experiment 2; (ii) the fact that the participants did not have the two versions of the can opener to compare, and (iii) that they were only asked to rate the can openers for usability.

Fig. 5. Experiment 3, older people's ratings of Usability for each model from Pretty and

Ugly groups averaging across before and after.

5. GENERAL DISCUSSION

Three experiments have been presented comparing ratings of usability and aesthetics for products whose aesthetic Product Features had been manipulated. The results of Experiment 1 conformed closely to the predictions of the model represented in Figure 1, B. The manipulation had the desired effect on aesthetic ratings, and was also reflected in usability ratings. The results of Experiment 2 were less clear cut. The effect of the manipulation on aesthetic ratings was much smaller and was not reflected in a significant effect on usability ratings. However, Experiment 3, which only required ratings of usability, did show a significant effect of the manipulation on usability ratings for Model 1.

In Experiment 1, the participants were students, whereas in Experiments 2 and 3 they were older people attending a drop-in centre. That these two populations have different values and aesthetics is dramatically demonstrated in the reversal of the effect of model. The cheapest and most familiar model obtained the lowest ratings from the students but the highest ratings from the older people. Serendipitously, this provides further evidence for a holistic judgement of Apparent Product Character as the ratings of usability completely paralleled the ratings of aesthetics.

Implications for Hassenzahl's model:
Hassenzahl's [6] model only contains the two rightward facing arrows depicted in Figure 1, A. No predictions are specified as to the interaction of individual Product

Features with pragmatic and hedonic character. It is however assumed that several elements of Apparent Product Character will interact to determine Consequences (see also [6], Figure 6).

Overall the prediction in Figure 1, B (dotted arrows) are upheld. Taken with [5] demonstration of the effect of a manipulation of usability on aesthetic ratings, there is a strong case that Product Features influence aspect of Apparent Product other than those they are obviously associated with.

Figure 1, B is convenient for considering alternative explanations of the findings. For example, [6] explains the correlation between beauty and usability observed by [5] in terms of an effect in the Consequences column. He states that "ugly-beautiful" has a general evaluative function as a rating scale. That is, the semantic differential "ugly-beautiful" taps into the Consequences as well as hedonic character. The effect of manipulating Product Features normally associated with objective usability on ratings on "ugly-beautiful" then is explained as occurring in the causative link between Apparent Product Character and Consequences.

It is harder to make such a case with the findings presented here as "this can opener is easy to use" would seem much less generally evaluative. However, it would be interesting to repeat the study with Hassenzahl's full questionnaire to directly test this assumption. This would also permit analysis of covariance to test the goodness of fit of alternative models of causation.

Methodological and practical implications:

There are clearly some advantages in sampling from different populations of potential users. The participation of older users extended the generality of previous findings in this area and a new effect of Model was observed. However, the effect of the aesthetic manipulation was small for this sample. Further research is required to determine what governs ratings of hedonic and pragmatic quality in the older population and to repeat the experiment with more successful manipulations of aesthetic features. Given the differences observed with these two user populations, it would also seem to be imperative to repeat the experiment with different products and different imagined or actual contexts of use.

This paper started by contrasting objective measures of usability, such as time to completion or ease of learning, with subject user experience. The results confirm previous findings that highlight the importance of aesthetic considerations in determining many aspects of user experience. The question remains as to how objective measures of usability should be factored into this equation. We have not assessed the objective usability of our can openers and whether this differed for older people and students, but if we had, it is not clear exactly how such data should be interpreted in the general model of user experience. Instantiating Hassenzahl's (2003) model in these experiments has thus exposed important theoretical and practical questions that need to be addressed in future research.

REFERENCES

1. Monk, A.F., User-centred design: the home use challenge, in Home informatics and telematics: information technology and society, A. Sloane and F. van Rijn, Editors. 2000, Kluwer Academic Publishers: Boston. p. 181-190.
2. Kurosu, M. and K. Kashimura. Apparent usability vs. inherrent usability. in CHI 1995. 1995: ACM Press.
3. Norman, D.A., Emotional Design: Why We Love (or Hate) Everyday Things. 2004, New York: Basic Books.
4. McCarthy, J. and P. Wright, Technology as experience. 2004, Cambridge, MA: MIT Press.
5. Tractinsky, N., A.S. Katz, and D. Ikar, What is beautiful is usable. Interacting with Computers, 2000. 13(2): p. 127-146.
6. Hassenzahl, M., The interplay of beauty, goodness and usability. Human-Computer Interaction, 2004. 19(4): p. 319-349.
7. Hassenzahl, M., The thing and I: understanding the relationship between user and product, in Funology: from usability to enjoyment, M. Blythe, et al., Editors. 2003, Kluwer: Dortrecht. p. 31-42.

Fig. 2. The can openers. Those in the left hand column were seen by the Ugly group, those on the right by the Pretty group. The first in each column is model 1 and so on down to model 4. (see text for description). This figure is in colour. Please contact the authors if you only have access to a printed copy of the paper.