

The influence of design aesthetics in usability testing: Effects on user performance and perceived usability

Andreas Sonderegger*, Juergen Sauer

Department of Psychology, University of Fribourg, Rue de Faucigny 2, 1700 Fribourg, Switzerland

ARTICLE INFO

Article history:

Received 21 November 2008

Accepted 8 September 2009

Keywords:

Usability test

Aesthetics

Adolescent

Performance

Mobile phone

ABSTRACT

This article examined the effects of product aesthetics on several outcome variables in usability tests. Employing a computer simulation of a mobile phone, 60 adolescents (14–17 yrs) were asked to complete a number of typical tasks of mobile phone users. Two functionally identical mobile phones were manipulated with regard to their visual appearance (highly appealing vs not appealing) to determine the influence of appearance on perceived usability, performance measures and perceived attractiveness. The results showed that participants using the highly appealing phone rated their appliance as being more usable than participants operating the unappealing model. Furthermore, the visual appearance of the phone had a positive effect on performance, leading to reduced task completion times for the attractive model. The study discusses the implications for the use of adolescents in ergonomic research.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Design aesthetics

Research in consumer ergonomics has indicated that product usability may not be the only major determinant of user satisfaction but that other design features also play an important role (Tractinsky et al., 2000; Norman, 2004). Over recent years, this has led to a continual shift in consumer ergonomics, moving from a functional view of usability issues (with a focus on improving efficiency and effectiveness of product usage) towards an experiential perspective, which takes into consideration the whole user experience (Forlizzi and Battarbee, 2004; Brave and Nass, 2008). User experience comprises the entire set of effects elicited by the use of a product, including aesthetic experience, experience of meaning, and emotional experience (Desmet and Hekkert, 2007). This suggests that aesthetics may play an important role in product and systems design.

The issue of aesthetics enjoys a long historic tradition in the research literature, with psychologists and philosophers having carried out theoretical and empirical work in that field. This topic has been the subject of discussions by ancient Greek philosophers such as Plato (beautiful objects incorporate proportion, harmony, and unity among their parts) and Aristotle (universal elements of beauty are order, symmetry, and definiteness). In the domain of

psychology, issues of aesthetics were first raised by Fechner [cited in Liu, 2003] whose aim was to discover the relationships between different design dimensions and perceived attractiveness through systematic manipulations of visual stimuli such as rectangles and ellipses. More recently, these ideas were taken up again to identify the features of stimuli (such as shape, colour, complexity, order, rhythm and prototypicality) that influence the attractiveness of an object (Liu, 2003; Hekkert and Leder, 2007).

In the research literature, the term design aesthetics is employed in two ways: it may refer to the *objective* features of a stimulus (e.g. colour of a product) or to the *subjective* reaction to the specific product features. To make a distinction between the two meanings, in the present study *aesthetics* refers to the objective design aspects of a product, including form, tone, colour, and texture (Postrel, 2003). Conversely, *attractiveness* refers to the individual's reaction to these product features and represents "the degree to which a person believes that the [product] is aesthetically pleasing to the eye" [van der Heijden, 2003; p. 544].

The response to aesthetic design is not only influenced by specific design factors (such as form or surface attributes) but may also be modified by characteristics of the individual, such as age, personality, cultural background or gender (Crilly et al., 2004). Because of its role in product marketing and consumer behaviour research [e.g. Meyers-Levy and Sternthal, 1991], gender may also be of particular interest in consumer ergonomics, though the evidence of the direction of the influence is far from being unequivocal. While some research has concluded that gender has little or no effect on aesthetic judgments [e.g. Lubner-Rupert and Winakor, 1985; Minshall et al., 1982; Morganosky and Postlewait, 1989],

* Corresponding author. Tel.: +41 26 3007630; fax: +41 26 3009712.
E-mail address: andreas.sonderegger@unifr.ch (A. Sonderegger).

there is other work that did find differences [e.g. Holbrook and Corfman, 1984; Holbrook, 1986]. However, since all of the work cited referred to non-interactive products such as clothes, it remains to be seen how gender moderates the effects of aesthetics in the context of operating interactive consumer products.

1.2. Usability testing

Given the role of aesthetics in product development, there is a need to examine the influence aesthetics have in usability testing. Usability testing is considered to be one of the most important and most widely used methods to evaluate product designs (Lewis, 2006). It aims to assess the usability of a product by simulating the user-product interaction under controlled conditions. Usability is defined according to the International Standardisation Organisation as “the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in a particular environment” (ISO, 1998). *Effectiveness* refers to the extent to which a task goal is successfully achieved (e.g., proportion of users that are able to complete a given task). *Efficiency* refers to the amount of resources a user expends to reach a task goal. It can be measured by the deviation from the optimal user behaviour (e.g., task completion time, number of user actions to complete a task). Both effectiveness and efficiency represent different kinds of performance measures. *Satisfaction* can be considered as an attitude towards the product. It is a subjective measure that is typically collected in usability tests by means of questionnaires (e.g. Chin et al., 1988; Lewis, 1995; Kirakowski et al., 1998; Willumeit et al., 1996).

1.3. Design aesthetics and perceived usability

The influence of aesthetics on perceived usability has already been addressed in several studies. These studies reported a positive correlation between perceived attractiveness and perceived usability for a range of products, such as computer-simulated cash machines (Kurosu and Kashimura, 1995; Tractinsky et al., 2000), websites (Hartmann et al., 2007; Schenkman and Jönsson, 2000) and computer software (Hassenzahl, 2004). While in these studies design aesthetics (attractive vs. unattractive) was not manipulated experimentally (and hence it cannot be excluded that perceived attractiveness and perceived usability were confounded), there are also studies in which an experimental manipulation of aesthetics was carried out. This includes the variation in colour settings of a webpage (Nakarada-Kordich and Lobb, 2005), the manipulation of the shape of an electronic phonebook-simulator (Ben-Bassat et al., 2006), the variation in the design of a webpage (following mathematical rules and two choices of colour settings; (Brady and Phillips, 2003), and the manipulation of the colour of casing and screen of a mobile phone (Sauer and Sonderegger, 2009). All these experiments confirmed that perceived usability was positively influenced by the aesthetics of the product. With regard to the psychological mechanisms behind this effect, the halo-effect has been put forward as a possible explanation. The halo effect describes the phenomenon that a specific, salient characteristic of a person or an object influences the apperception of other characteristics. This is analogous to the “what is beautiful is good”-stereotype, known from social psychology, that has been postulated to explain the phenomenon that physically attractive persons are considered to possess more positive personality traits than unattractive persons (Dion et al., 1972). Since attributes of physical beauty are obvious and accessible to others very early in the interaction between humans, they are assumed to colour later perceptions of other personal characteristics. Similarly, in usability testing the user's attitude towards a product is formed very rapidly (i.e. in about 50 ms) during user-product interaction (Lindgaard et al., 2006), which exemplifies the

importance of the very first impression. Overall, there is ample evidence of the positive influence of aesthetics on perceived usability.

1.4. Design aesthetics and user performance

While the positive relation between aesthetics and perceived usability has been well demonstrated by empirical research, it is less clear how aesthetics is linked with objective measures of performance in usability tests. Only very few studies have examined the effect of aesthetics on performance measures, albeit with somewhat inconsistent findings. Two studies found evidence of performance decrements when using an aesthetically pleasing product. For example, test participants showed poorer performance using an appealing computer simulation of an electronic phonebook (Ben-Bassat et al., 2006). Similar results were obtained in a study in which the aesthetics of a mobile phone was manipulated experimentally (Sauer and Sonderegger, 2009). However, two other studies found no effect of aesthetics on performance. Hartmann et al. (2007) reported no correlation between perceived attractiveness and task completion time when comparing three different webpages. Thüring and Mahlke (2007) varied the design aesthetics of existing MP3-players, with the results showing no effects of aesthetics on task completion time and error rate.

One may envisage two different effects of aesthetics on performance measures: an “increased motivation”-effect (i.e. increments in performance) or a “prolongation of joyful experience”-effect (i.e. decrements in performance). For the “increased motivation” effect, one may speculate that technology that is aesthetically pleasing might put the user at ease (Lindgaard, 2007) or put the user “in flow” (Csíkszentmihályi, 1997), which both may result in increased performance (e.g. reduced task completion time). In contrast, the “prolongation of joyful experience”-hypothesis would predict decreased user performance because the user enjoys the beauty of the product and therefore concentrates less on the task to be completed. This may lead to longer task completion times due to extended observation times during user-system interaction. The empirical findings reported above provided cautious support for the “prolongation of joyful experience”-explanation while no support has yet been found for the “increased motivation”-effect.

1.5. The present study

The primary research question of this study addressed the influence of aesthetics on central outcome variables of usability testing, such as perceived usability and user performance. For this purpose, two functionally identical mobile phones were manipulated with regard to their visual appearance to make them either aesthetically appealing or unappealing. In all system features other than aesthetic appeal, the two appliances were identical. The mobile phone was chosen as a technical device because it has a stronger affective component than most other interactive consumer products (e.g., vacuum cleaner). This will give additional weight to design aesthetics. The present study was conducted with adolescents as an important group of mobile phone users (Milanese, 2005). In addition to the influence of aesthetics, we have examined the influence of gender as a secondary research question.

Based on the research literature reviewed, the following three hypotheses were formulated: (a) User performance will be better for the more aesthetically pleasing product than for the less pleasing one. (b) Perceived usability will be higher for the aesthetically more pleasing product than for the less pleasing one. (c) The difference in perceived usability between the two conditions will be less pronounced after the usability test than prior to it, due to the diminishing influence of aesthetic after the user had actual

experience with the product. Because of the equivocal pattern for gender, no hypothesis was formulated for the effects of gender.

2. Method

2.1. Participants

The sample of this study consists of 60 participants (52% female). All of them were pupils doing their GCSEs (General Certificate of Secondary Education) at a secondary school in Thun (Switzerland), aged between 13 and 16 years ($M = 14.2$). Self reports showed that they were quite experienced mobile phone users, employing their mobile phone on average 8.7 times per day ($SD = 10.6$). Their self-rated expertise in operating a mobile phone was $M = 65.0$ on a 100 mm visual analogue scale. The ends of the scale were labelled “very little experience”, and “a great deal of experience”, with higher values indicating more experience. The two experimental groups did not differ in their self-rated expertise in mobile phone usage ($t < 1$) and in their stated frequency of daily phone usage ($t = 1.57$, $df = 55.7$, $p > .05$). There was no difference between male and female participants with regard to their perceived expertise in mobile phone usage ($t < 1$) and their reported frequency of usage ($t < 1$).

2.2. Experimental design

A 2×2 mixed design was employed in the experiment, with *aesthetics of design* as a between-subjects variable. Participants were randomly assigned to a group using a prototype of mobile phone with an appealing design or an unappealing one. To determine the effects of product usage experience, some measures were recorded repeatedly during the usability test. This within-participants variable was varied at two levels: prior to the product usage in the usability test and following the usability test.

The influence of gender was examined by using this variable as a covariate. The distribution of gender across conditions was unequal (e.g. 12 females used the unappealing phone while 19 females operated the appealing phone), due to the particular distribution of gender in the participating school classes.

2.3. Measures and instruments

2.3.1. Perceived product attractiveness

The attractiveness of the appliance was measured before and after product usage. The measure (prior to usability test) served as a manipulation check. A one-item scale was used (“the design of the mobile phone is very appealing”), with a seven-point Likert scale (strongly agree, agree, partly agree, neither agree nor disagree, partly disagree, disagree, strongly disagree) as a response format. A single-item scale was chosen, to ensure that participant motivation was maintained throughout the testing session. Since the main goal of the study was to attain an overall assessment, the use of a 1-item measure is justifiable if the item is unambiguous and captures the main concept (Wanous et al., 1997). This type of scales has been employed in previous usability studies (e.g. Tractinsky et al., 2000).

2.3.2. Perceived usability

Similar to the evaluation of the attractiveness of the prototype, test participants were asked to assess the usability of the mobile phone before and after product usage on a one-item scale (“The mobile phone seems to be very usable”). Again, a seven-point Likert scale was used (strongly agree, agree, partly agree, neither agree nor disagree, partly disagree, disagree, strongly disagree).

As a more detailed measure of the system usability comprising several subscales, a German translation of the Post System Study Usability Questionnaire (PSSUQ) (Lewis, 1995) was employed after product usage. This instrument has been widely applied for usability testing in laboratory settings. The questionnaire was slightly modified by removing four items that were irrelevant for the intended application area. The remaining items are presented in Table 1. To improve comprehensibility, items were adapted to the appliance it was used for (e.g., “system” was replaced by “mobile phone”). Users rated the items on the same seven-point Likert scale as the single-item scale above. The PSSUQ comprised the following three subscales: system usefulness, information quality and interface quality. The overall internal consistency of the questionnaire as well as the internal consistency of the subscales was found to be satisfactory (see Table 1).

2.3.3. User performance

Three measures of user performance were recorded. *Task completion time* referred to the time needed to accomplish the task. *Interaction efficiency* is a composite parameter, dividing the optimal number of user manipulations by the actual number of user inputs. Lastly, the *number of error messages* that have been displayed when the user chose a wrong navigation option was recorded.

2.4. Materials

Two functionally identical computer prototypes of a mobile phone were used in this study. One version was aesthetically appealing, the other one not so (see Fig. 1). It is useful to note that users only employed the navigation buttons in the top section of the interface for task completion (i.e. they did not need to use the numeric keys). The buttons in the top section were of the same size for both appliances. The functionality of the two appliances was exactly the same. This was because the overlaid event triggers (in the form of invisible push buttons) were exactly of the same size for both appliances while only the form (but not the size) of the visible shell differed slightly between appliances. To control for objective usability differences between the two appliances, we calculated the

Table 1

Adapted version of Post Study System Usability Questionnaire (PSSUQ) (Lewis, 1995).

| Post Study System Usability Questionnaire (Cronbach's $\alpha = .88$) | |
|--|---|
| Subscale “system usefulness” (Cronbach's $\alpha = .91$) | Overall, I am satisfied with how easy it is to use this mobile phone. It was simple to use this mobile phone. I was able to complete the tasks and scenarios quickly using this mobile phone. I felt comfortable using this mobile phone. It was easy to learn to use this mobile phone. I believe I could become productive quickly using this mobile phone. |
| Subscale “information quality” (Cronbach's $\alpha = .68$) | The mobile phone gave error messages that clearly told me how to fix problems. Whenever I made a mistake using the mobile phone, I could recover easily and quickly. The information provided by this mobile phone was clear. It was easy to find the information I needed. The information was effective in helping me complete the tasks and scenarios. The organisation of information on the mobile phone's display was clear. |
| Subscale “interface quality” (Cronbach's $\alpha = .87$) | The interface of this mobile phone was pleasant. I liked using the interface of this mobile phone. |
| Overall satisfaction | Overall, I am satisfied with this mobile phone. |



Fig. 1. Two prototypes employed in experiment: (a) aesthetically appealing design; (b) aesthetically unappealing design.

average time per click. The results of the analysis showed that there was no difference between the appealing ($M = 3.0$; $SD = 0.8$) and the unappealing design ($M = 3.1$; $SD = 0.8$) ($F < 1$). This suggests that task difficulty for the two appliances was the same.

The two designs were developed, using graphic design software (Photoshop). The designs were based on previous research, which identified a number of factors that determine object attractiveness, such as colour, texture, symmetry, and clarity (Ngo et al., 2003; Postrel, 2003).

The dialogue structure of the mobile phone was based on the functionality of a SonyEricsson™ SE W800i. Compared to the original appliance, the functionality of the prototype was limited. Only for the task-relevant menu items, the dialogue structure was modelled in full depth. For functions that were irrelevant for task completion, only the two top levels of the dialogue structure were represented. An error message was displayed (“wrong path, please go back”) when a user selected a function that was not simulated in the prototype (i.e. more than two clicks away from the optimal dialogue path). The computer simulation of the dialogue structure was developed using Powerpoint™. Both computer simulations (appealing and unappealing) were installed on a Toshiba Portege™ M200 TabletPC. For the interaction with the prototype, a computer mouse (Logitech Pilot Optical) was used.

2.5. Pilot study

In a pilot study, different design alternatives for the prototype of the mobile phone were compared. 10 participants (aged between 14 and 16 yrs) evaluated the attractiveness of these designs (one aesthetically appealing and two aesthetically unappealing ones). The participants were recruited from the same population as the sample

of participants of the main study. The two aesthetically unappealing designs differed in form and colour setting compared to the appealing one. Both unappealing prototypes consisted of a disharmonious facia of different colours (blue, yellow, pink, and grey). The buttons were either purple or looked like if they were made out of wood. On a scale ranging from 1 to 10, ratings of perceived attractiveness differed remarkably between the appealing and the two unappealing prototypes: (a) $M = 8.1$ ($SD = .88$); (b) $M = 2.0$ ($SD = 1.33$); (c) $M = 2.6$ ($SD = 1.50$). The prototype with the highest score (Fig. 1a) and the one with the lowest score (Fig. 1b) in the attractiveness rating were selected for the main study.

2.6. User tasks

For the usability evaluation, two tasks had to be completed by test users. These tasks were chosen because they represent typical activities in mobile phone usage. The first task (“text message”) involved sending a prepared text message to another phone user. This task could be completed with a minimum number of 9 clicks. In the second task (“phone number suppression”), test users had to change the mobile phone settings in such a way that one's own phone number is suppressed when making a call. To complete this task, a minimum number of 16 clicks were necessary.

2.7. Procedure

The study was conducted in a computer lab of the school. Participants were recruited from different classes on a voluntary basis and within each class, participants were randomly assigned to one of the experimental conditions. Any difference in age or ability between experimental groups is expected to be balanced by the

procedure of randomly allocating participants. Participation in the study took about 20 minutes. All participants were tested individually. After being welcomed by the experimenter, participants were informed that they would take part in a usability test and would have to operate a computer-simulated prototype of a mobile phone. Prior to operating the prototype, participants were asked to rate their previous experience with mobile phones and to rate attractiveness and usability of the mobile phone on the two single-item scales. Then, participants completed the two experimental tasks. Immediately after task completion, the two single-item scales and the PSSUQ were presented. The experiment was concluded with a debriefing session, in which the participant was given the opportunity to give further feedback about the prototype and the testing procedure.

2.8. Statistical analysis

To examine the impact of design aesthetics and product usage on subjective evaluations of attractiveness and usability, a two-factorial analysis of variance was used. For the analysis of the performance data a one factorial analysis of variance was carried out. In both cases, the influence of gender was examined by entering this factor as a covariate.

3. Results

3.1. Perceived product attractiveness

The data of the attractiveness evaluation of the two prototypes before and after product usage is presented in Fig. 2. Representing a manipulation check, the data confirmed that the aesthetically appealing prototype was rated significantly more attractive than the unappealing prototype ($M_{\text{appealing}} = 5.3$ vs. $M_{\text{unappealing}} = 3.15$; $F = 39.8$; $df = 1, 58$; $p < .001$). Furthermore, an interaction between prototype and product usage was found ($F = 4.7$; $df = 1, 57$; $p < .05$), showing an increase in the perceived attractiveness rating of the aesthetically appealing prototype after product usage whereas the

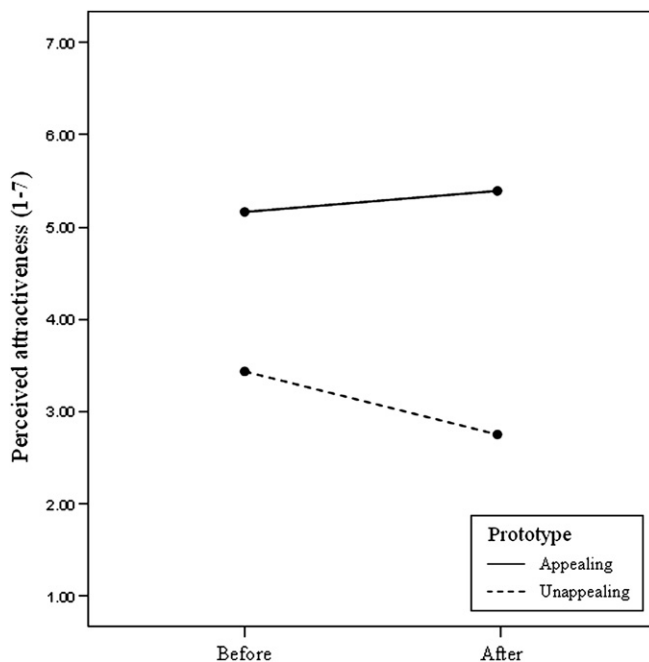


Fig. 2. User ratings of perceived attractiveness (1–7) of the prototype before and after product usage as a function of design aesthetics.

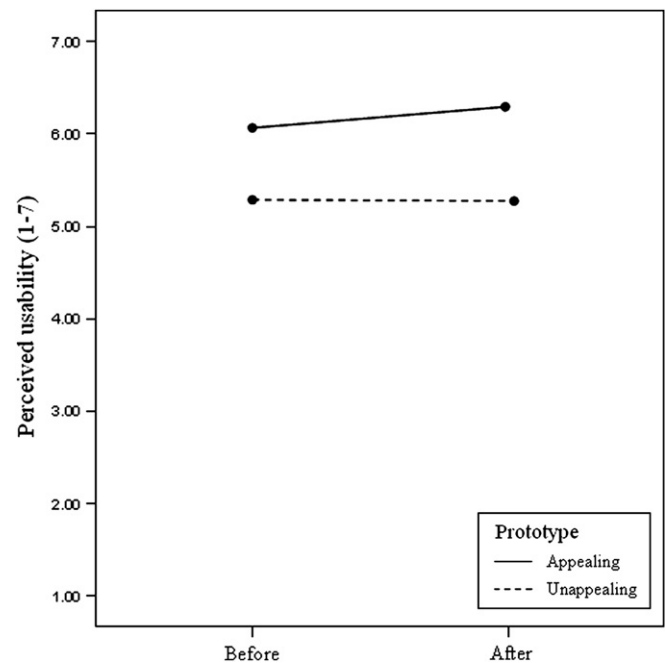


Fig. 3. User ratings of perceived usability (1–7) on the one-item scale before and after product usage as a function of design aesthetics.

attractiveness-rating of the unappealing prototype decreased after product usage. The main effect of product usage (before vs. after) was not significant ($F = 1.5$; $df = 1, 57$; $p > .05$). The covariate *gender* was not related to the perceived product attractiveness, neither before nor after user-product interaction (all $F < 1$).

3.2. Perceived usability

Perceived usability was measured prior to task completion and after task completion by the 1-item scale as well as with the PSSUQ after task completion. The ratings on the 1-item scale differed significantly between the two prototypes ($M_{\text{appealing}} = 6.14$ vs. $M_{\text{unappealing}} = 5.32$; c.f. Fig. 3). The appealing prototype was rated more usable than the unappealing one ($F = 9.8$; $df = 1, 57$; $p < .01$). The actual use of the prototype did not influence the user's usability rating ($F < 1$) and also the interaction between prototype and product usage was not significant ($F < 1$). Gender was not related to the perceived usability, neither before nor after user-product interaction (all $F < 1$).

The perceived usability ratings after product usage on the PSSUQ are similar to the ratings on the 1-item scale (c.f. Table 2). The analysis revealed that overall ratings were significantly higher for the appealing prototype than for the unappealing one ($F = 20.8$; $df = 1, 57$; $p < .001$). A separate analysis of the three subscales confirmed the same pattern for system usefulness ($F = 13.6$; $df = 1, 57$; $p < .001$), information quality ($F = 7.2$; $df = 1, 57$; $p < .01$) and

Table 2

Perceived usability (1–7) on the Post System Study Usability Questionnaire as a function of design aesthetics.

| | Prototype | |
|---------------------------------|---------------------|-----------------------|
| | Appealing prototype | Unappealing prototype |
| | M (SD) | M (SD) |
| Overall scale (item 1–15) | 6.13 (0.48) | 5.19 (0.91) |
| System usefulness (item 1–6) | 6.29 (0.62) | 5.32 (1.22) |
| Information quality (item 7–12) | 6.27 (0.56) | 5.66 (0.87) |
| Interface quality (item 13–14) | 5.30 (1.38) | 3.60 (1.69) |

Table 3
Measures of user performance as a function of design aesthetics and gender.

| | Appealing prototype | Unappealing prototype | Overall |
|----------------------------------|------------------------|--------------------------|-----------------|
| | M (SD) | M (SD) | M (SD) |
| Task completion time (s) | 147.7 (58.0) | 198.5 (83.6) | 173.1 (76.2) |
| Female | 147.2 (57.8) | 230.1 (102.4) | 179.3 (86.8) |
| Male | 148.5 (64.0) | 177.4 (62.9) | 166.4 (63.8) |
| Interaction efficiency index (%) | 59 (17) | 46 (19) | 53 (19) |
| Female | 58 (16) | 42 (20) | 52 (19) |
| Male | 61 (19) | 48 (18) | 53 (19) |
| Number of errors (per trial) | 2.0 (2.0) | 4.8 (5.0) | 3.4 (4.0) |
| Female | 2.2 (2.1) | 7.0 (6.4) | 4.1 (4.8) |
| Male | 1.5 (1.6) | 3.3 (3.1) | 2.7 (2.8) |

interface quality ($F = 14.5$; $df = 1, 57$; $p < .001$). Gender showed no relationship with perceived usability (all $F < 1$).

3.3. User performance

Task completion time. The analysis of the data of task completion time revealed a significant difference for the two designs (c.f. Table 3). It showed that participants using the appealing prototype needed less time to complete their tasks than the participants using the unappealing prototype ($F = 8.9$; $df = 1, 57$; $p < .01$). The covariate gender was not related to task completion time ($F = 1.9$; $df = 1, 57$; $p > .05$).

Interaction efficiency. Similar to the findings of task completion time, the analysis of the data on interaction efficiency (optimal click number divided by actual number of clicks; c.f. Table 3) indicated a significant effect of design aesthetics ($F = 8.8$; $df = 1, 57$; $p < .01$). Participants using the appealing prototype needed fewer clicks to complete their tasks than participants using the unappealing one. Gender was not related to interaction efficiency (all $F < 1$).

Errors. The analysis of errors that occurred during task completion (c.f. Table 3) revealed that participants using the attractive prototype committed significantly fewer errors than the participants using the unappealing prototype ($F = 12.0$; $df = 1, 57$; $p < .001$). This shows that all three performance measures indicate better performance when operating an appealing prototype. Gender was related to the error rate ($F < 5.1$; $df = 1, 57$; $p < .05$), indicating that female participants committed more errors than male ones.

4. Discussion

The findings showed that perceived usability was higher for appealing products than for unappealing ones, even though there was no difference between the two appliances in the objective quality of usability. This pattern was observed for the one-item scale as well as for the more elaborate instrument PSSUQ on all its subscales. These results provide further confirmation of the positive influence of aesthetics on perceived usability observed in previous work (Nakarada-Kordich and Lobb, 2005; Ben-Bassat et al., 2006; Brady and Phillips, 2003; De Angeli et al., 2006). This tendency which was consistently observed across different adult user populations was also applicable in the case of adolescent users. Furthermore, it is noteworthy that the actual completion of the experimental tasks did not change perceived usability as one would expect. If there was an influence of aesthetics prior to using the appliance, one would expect this influence to decrease in size as the

user becomes more familiar with the appliance. However, the ratings remained stable (if anything, the difference widened rather than narrowed as visual inspection of the data suggests). This stability in ratings observed before and after the usability test was also observed in an experiment with adult users, employing a similar experimental set-up (Sauer and Sonderegger, 2009). This suggests that the observed effects are consistent across age groups. Furthermore, similar to social psychology, where the “what is beautiful is good”-stereotype seems to represent a cross-cultural phenomenon (Chen et al., 1997), the cross-cultural quality of the effect also appears to apply to judgements of perceived attractiveness on technical artefacts since similar findings obtained with the Swiss sample in the current study were reported from studies conducted in Japan (Kurosu and Kashimura, 1995), Israel (Tractinsky et al., 2000), and Germany (Thüring and Mahlke, 2007).

While the effects on perceived usability were in line with previous work, the influence of aesthetics on user performance was in contrast to previous findings. The present study provided support for an “increased motivation”-effect, with users showing better performance with the appealing prototype. Previous work, however, found support for the “prolongation of joyful experience”-effect, with users taking more time to complete a data entry task (Ben-Bassat et al., 2006) and to operate a mobile phone (Sauer and Sonderegger, 2009) when using the more appealing version of the technical artefact. These differences may be due to inherent domain characteristics (leisure vs work context). One may assume that the “increased motivation” effect would be more likely to occur in a work context while the “prolongation of joyful experience”-explanation would be more likely to be observed in a leisure context. As the present study was carried out in a school setting (which most pupils would not consider a leisure-oriented environment), a stronger performance-orientation may have ensued from this, resulting in a higher motivation to complete the tasks as fast as possible. Interestingly, this effect was opposite to the one observed in a previous study (Sauer and Sonderegger, 2009), which used a similar experimental set-up with a mobile phone being operated but in a leisure-oriented context. In such a context, the focus may be less on performance but more on fun and enjoyment which supports the mechanism of the “prolongation of enjoyable experience” effect.

While the results clearly demonstrated that the manipulation check was successful (since the two mobile phones were rated very differently with regard to their perceived attractiveness), more interesting was the observation that the difference in perceived attractiveness between high and low aesthetics widened after the usability test. This observation may be interpreted by referring to the attitude polarization effect (Lord et al., 1979). The initial attitude (which is formed very early during user-product interaction; (Lindgaard et al., 2006)) may have become more extreme due to biased information assimilation (MacCoun, 1998). The occurrence of attitude polarization among adolescents was also demonstrated in the context of reasoning about religious affiliations (Klaczynski and Gordon, 1996). It showed that adolescents’ reasoning was systematically biased to protect and promote pre-existing beliefs. Overall, the present findings may suggest that usability has little influence on perceived attractiveness. Otherwise, one would have expected some narrowing of the difference, as users gained increasing experience with the usability of the product (which was identical for both conditions). This speculative explanation needs to be empirically tested by manipulating product usability and determine its effects on perceived attractiveness, with particular consideration to be given to the long-term effects over repeated practice trials.

Overall, gender had little effect on outcome variables. This is in line with the bulk of the literature (albeit a small number of studies did find an effect), suggesting that the influence of aesthetics is not only observed across cultures and age groups but also across

gender. Although an effect was recorded on a single measure (i.e. suggesting that female users committed more errors with the unappealing prototype than male users), it was difficult to interpret and, given its small effect size, it should not be taken as evidence for a general consideration of gender as a crucial variable that moderates the influence of aesthetics in usability evaluations. Although the unequal distribution of gender reduced the power of the covariance analysis, even with a more balanced distribution, it is unlikely that the effect of gender would have been significant, given the size of the effects.

It is important to note that the results in this study are based on a sample of adolescent test users. The use of adolescents as a separate user group seems to be increasingly relevant, given their growing financial freedom of manoeuvre (e.g. Shim, 1996), their influence in family decision-making (e.g. Beatty and Talpade, 1994; Foxman et al., 1989) and their role as future (adult) customers with whom it is important to establish an early brand relationship (Khadir, 2007). Against this background, it is justifiable and increasingly necessary to carry out research with adolescent users. The current study also provided first hints about possible differences in the effects of aesthetics compared to adults, though we do not know whether these were due to differences in user groups (i.e. adults vs. adolescents) or in usage context (work vs. leisure). We would therefore caution against a generalisation of the findings of the present study to other user groups.

Some limitations with regard to the interpretation of the results are acknowledged. While the effects of aesthetics may be due to the mechanisms discussed above, alternative explanations are also possible. The attractive phone might have been perceived as a conventional phone that can be purchased in the shops (and is fully usable) whereas the unattractive one might have been perceived as having a rather unusual design (which is not yet fully developed). Therefore, the more conventional product might have been evaluated more positively, resulting in a confounding effect of familiarity and aesthetics. On the basis of the available data, it is not possible to control for such a confounding effect. However, participant feedback after task completion did not indicate a difference in the perception of prototypicality between the two prototypes. Furthermore, to minimise a possible effect of familiarity, users of a recent SonyEricsson™ mobile phone (upon which our prototypes were based) were not allowed to take part in the study.

As the present study and previous work have demonstrated, there seems to be increasing evidence for the influence of aesthetics beyond subjective parameters such as perceived usability. Indeed, aesthetics may influence performance, with empirical evidence having been observed for both effects (“prolongation of joyful experience”-effect vs. “increased motivation”-effect). This suggests the need for experiments to address the following issues in future research. First, direct comparisons between adolescents and adults should be made. We may predict that adolescents might attach even more importance to the mobile phone's aesthetics (resulting in a stronger effect of aesthetics on usability test outcomes) since they are often prone to extreme self-focus and are excessively concerned with what their peers think of them (Magrab, 2005). Second, different usage contexts such as the domestic and work domain should be compared. We may predict a stronger influence of aesthetics in the domestic (and leisure) domain than in the work domain. Third, it would be of interest to determine to what extent the influence of aesthetics is moderated by the prestige value associated with the product. The prestige value of a mobile phones may be considered high (Dedeoglu, 2004), compared to other products such as a vacuum cleaner or an electric fire. As a concluding remark, we would like to point out that the usage of interactive consumer products should not be considered separately from the purchase decision. As the user is often the buyer of a product,

ergonomic issues become more strongly interwoven with issues pertaining to consumer psychology.

Acknowledgement

We are very grateful to Amadeus Petrig, Eric Bourquard and Karin Schlapbach for their support in completing this study.

References

- Beatty, S.E., Talpade, S., 1994. Adolescent influence in family decision making: a replication with extension. *J. Cons. Res.* 21, 332–341.
- Ben-Bassat, T., Meyer, J., Tractinsky, N., 2006. Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Trans. Comput. Human Interact.* 13, 210–234.
- Brady, L., Phillips, C., 2003. Aesthetics and usability: a look at color and balance. *Usability News* 5.
- Brave, S., Nass, C., 2008. Emotion in human–computer interaction. In: Sears, A., Jacko, J.A. (Eds.), *The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Lawrence Erlbaum & Associates, Mahaw, pp. 77–92.
- Chen, N., Shaffer, D., Wu, C., 1997. On physical attractiveness stereotyping in Taiwan: a revised sociocultural perspective. *J. Soc. Psychol.* 137, 117–124.
- Chin, J.P., Diehl, V.A., Norman, K.L., 1988. Development of an instrument measuring user satisfaction of the human–computer interface. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Washington, pp. 213–218.
- Crilly, N., Moultrie, J., Clarkson, P., 2004. Seeing things: consumer response to the visual domain in product design. *Design Studies* 25, 547–577.
- Csikszentmihályi, M., 1997. *Finding Flow: The Psychology of Engagement With Everyday Life*. Basic Books, New York.
- De Angeli, A., Sutcliffe, A., Hartmann, J., 2006. Interaction, usability and aesthetics: what influences users' preferences? In: *Proceedings of the 6th Conference on Designing Interactive Systems*. ACM, New York, pp. 271–280.
- Dedeoglu, A.O., 2004. The symbolic use of mobile telephone among Turkish consumers. *J. Euromarket.* 13, 143–162.
- Desmet, P.M.A., Hekkert, P., 2007. Framework of product experience. *Int. J. Design* 1, 57–66.
- Dion, K., Berscheid, E., Walster, E., 1972. What is beautiful is good. *J. Pers. Soc. Psychol.* 24, 285–290.
- Forlizzi, J., Battarbee, K., 2004. Understanding experience in interactive systems. In: *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. ACM, New York, pp. 261–268.
- Foxman, E., Tansuhaj, P., Ekstrom, K.M., 1989. Adolescents' influence in family purchase decisions: a socialisation perspective. *J. Bus. Res.* 18, 159–172.
- Hartmann, J., Sutcliffe, A., Angeli, A.D. Investigating attractiveness in web user interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, April 28–May 3, 2007, pp. 387–396.
- Hassenzahl, M., 2004. The interplay of beauty, goodness and usability in interactive products. *Human Comput. Interact.* 19, 319–349.
- Hekkert, P., Leder, H., 2007. Product aesthetics. In: Schifferstein, H.N.J., Hekkert, P. (Eds.), *Product Experience*. Elsevier Science, Amsterdam, pp. 259–286.
- Holbrook, M.B., Corfman, K.P., 1984. Quality and value in the consumption experience: phaedrus rides again. In: Jacoby, J., Olson, J.C. (Eds.), *Perceived Quality: How Consumers View Stores and Merchandise*. D.C. Heath, Lexington, pp. 31–57.
- Holbrook, M.B., 1986. Aims, concepts, and methods for the representation of individual differences in esthetic responses to design features. *J. Cons. Res.* 13, 337–347.
- ISO 9241-11, 1998. *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)-Part 11, Guidance on Usability*. ISO.
- Khadir, F. Marketing and its impact on vulnerable consumer groups like children, adolescents etc. In: *Proceedings of the International Marketing Conference on Marketing & Society*, Kozhikode, April 8–10, 2007, pp. 433–441.
- Klaczynski, P.A., Gordon, D.H., 1996. Self-serving influences on adolescents' evaluations of belief-relevant evidence. *J. Exp. Child Psychol.* 62, 317–339.
- Kirakowski, J., Claridge, N., Whitehand, R. Human centered measures of success in web site design. In: *Proceedings of the Fourth Conference on Human Factors and the Web*, Basking Ridge, June 5, 1998.
- Kurosu, M., Kashimura, K. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Denver, May 07–11, 1995, pp. 292–293.
- Lewis, J.R., 1995. IBM computer usability satisfaction questionnaire: psychometric evaluation and instructions for use. *Int. J. Human Comput. Interact.* 7, 57–78.
- Lewis, J.R., 2006. Usability testing. In: Salvendy, G. (Ed.), *Handbook of Human Factors and Ergonomics*. John Wiley, New York, pp. 1275–1316.
- Lindgaard, G., Dudek, G., Fernandes, G., Brown, J., 2006. Attention web designers: you have 50 milliseconds to make a good first impression! *Behav. Info. Tech.* 25, 115–126.
- Lindgaard, G., 2007. Aesthetics, visual appeal, usability, and user satisfaction: what do the user's eyes tell the user's brain. *Aust. J. Emerg. Tech. Soc.* 5, 1–16.

- Liu, Y., 2003. Engineering aesthetics and aesthetic ergonomics: theoretical foundations and a dual-process research methodology. *Ergonomics* 46, 1273–1292.
- Lord, C.G., Ross, L., Lepper, M.R., 1979. Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *J. Pers. Soc. Psychol.* 37, 2098–2109.
- Lubner-Rupert, J.A., Winakor, G., 1985. Male and female style preference and perceived fashion risk. *Family Cons. Sci. Res. J.* 13, 256–266.
- MacCoun, R.J., 1998. Biases in the interpretation and use of research results. *Ann. Rev. Psychol.* 49, 259–287.
- Magrab, P.R. The adolescent learner and the aesthetic experience: a brief overview. In: *Education through Art – Building Partnerships for Secondary Education*, Proceedings of the Unesco Expert Panel Meeting, Newark, October 27, 2005, pp. 7–11.
- Meyers-Levy, J., Sternthal, B., 1991. Gender differences in the use of message cues and judgments. *J. Market. Res.* 28, 84–96.
- Minshall, B., Winakor, G., Swinney, J.L., 1982. Fashion preferences of males and females, risks perceived, and temporal quality of styles. *Family Cons. Sci. Res. J.* 10, 369–379.
- Morganosky, M.A., Postlewait, D.S., 1989. Consumers' evaluations of apparel form, expression, and aesthetic quality. *Clothing Textiles Res. J.* 7, 11–15.
- Milanese, S., 2005. Adolescent ergonomics. In: Bust, P.D., McCabe, P.T. (Eds.), *Contemporary Ergonomics*. Taylor & Francis, London, pp. 327–330.
- Nakarada-Kordich, I., Lobb, B. Effect of Perceived attractiveness of web interface design on visual search of web sites. In: *Proceedings CHINZ '05*, July 6–8, Auckland, 2005, pp. 25–27.
- Ngo, D.C.L., Teo, L.S., Byrne, J.G., 2003. Modelling interface aesthetics. *Info. Sci.* 152, 25–46.
- Postrel, V., 2003. *The Substance of Style: How the Rise of Aesthetic Value is Remaking Commerce, Culture, and Consciousness*. Harper Collins, New York.
- Norman, D.A., 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, New York.
- Sauer, J., Sonderegger, A., 2009. The influence of prototype fidelity and aesthetics of design in usability tests: effects on user behaviour, subjective evaluation and emotion. *Appl. Ergon.* 40, 670–677.
- Schenkman, B.N., Jönsson, F.U., 2000. Aesthetics and preferences of web pages. *Behav. Info. Tech.* 19, 367–377.
- Shim, S., 1996. Adolescent consumer decision-making styles: the consumer socialization perspective. *Psychol. Market.* 13, 547–569.
- Thüring, M., Mahlke, S., 2007. Usability, aesthetics and emotions in human-technology interaction. *Int. J. Psychol.* 42, 253–264.
- Tractinsky, N., Shoval-Katz, A., Ikar, D., 2000. What is beautiful is usable. *Interact. Comput.* 13, 127–145.
- van der Heijden, H., 2003. Factors influencing the usage of websites: the case of a generic portal in The Netherlands. *Inform. Manage.* 40, 541–549.
- Wanous, J.P., Reichers, A.E., Hudy, M.J., 1997. Overall job satisfaction: how good are single-item measures? *J. Appl. Psychol.* 82, 247–252.
- Willumeit, W., Gediga, G., Hamborg, K., 1996. IsoMetrics: a technique for formative evaluation of software in accordance to ISO 9241/10. In: Ziegler, J. (Ed.), *Ergonomie und Informatik*. Copy Druck, Stuttgart, pp. 5–12.

Andreas Sonderegger is a research assistant at the Department of Psychology, University of Fribourg. He received his Master's degree in Work and Organizational Psychology from the University of Fribourg, Switzerland, in 2003.

Juergen Sauer is Professor of Cognitive Ergonomics at the Department of Psychology, University of Fribourg. He obtained his Ph.D. in Psychology from the University of Hull, UK, in 1997.