Research Article

The Interplay between Usability and Aesthetics: More Evidence for the "What Is Usable Is Beautiful" Notion

Kai-Christoph Hamborg,[1] Julia Hülsmann,[1] and Kai Kaspar[2]

[1] *Institute of Psychology, University of Osnabrück, Seminarstraße 20, 49074 Osnabrück, Germany*
[2] *Social and Media Psychology, Department of Psychology, University of Cologne, Richard-Strauss-Straße 2, 50931 Cologne, Germany*

Correspondence should be addressed to Kai-Christoph Hamborg; khamborg@uni-osnabrueck.de

With respect to inconsistent findings on the interplay between usability and aesthetics, the current paper aimed to further examine the effect of these variables on perceived qualities of a mobile phone prototype. An experiment with four versions of the prototype varying on two factors, (1) usability (high versus low) and (2) aesthetics (high versus low), was conducted with perceived usability and perceived beauty, as well as hedonic experience and the system's appeal as dependent variables. Participants of the experiment ($N = 88$) were instructed to complete four typical tasks with the prototype before assessing its quality. Results showed that the mobile phone's aesthetics does not affect its perceived usability, either directly or indirectly. Instead, results revealed an effect of usability on perceived beauty, which supports the "what is usable is beautiful" notion instead of "what is beautiful is usable." Furthermore, effects of aesthetics and of usability on hedonic experience in terms of endowing identity and appeal were found, indicating that both instrumental (usability) and noninstrumental (beauty) qualities contribute to a positive user experience.

1. Introduction

Via a long period, research and practice in human computer interaction (HCI) was predominantly concerned with the so-called pragmatic qualities such as utility and usability of interactive systems [1]. Since the early 2000s, additional criteria for design and evaluation addressing the user experience (e.g., aesthetics, beauty, and hedonic quality) of interactive systems have begun to be established [2, 3]. These criteria are also designated as noninstrumental as distinct from the criteria of instrumental quality like usability or pragmatic quality, respectively [4]. Today, besides usability, noninstrumental qualities are considered constituent parts of an enriched model of product quality supposed to affect product appeal and user experience during system usage "beyond the instrumental" [5, page 92 ff.] including affective and emotional reactions [6]. In conjunction with the establishment of aesthetics as a criterion of noninstrumental quality of interactive systems, the question about the interplay between instrumental and noninstrumental qualities of interactive software came up. Lavie and Tractinsky [7] pointed out that in the current field of research the term "aesthetics" is often used to

describe a beautiful or pleasing appearance of an interactive system (see also [8–10]).

In a seminal study on the question of the interplay between instrumental and noninstrumental qualities, titled "What is beautiful is usable," Tractinsky et al. [8] found a high correlation between pre- and postexperimental (after hands-on experience) perceptions of the aesthetics of a computer simulated automatic teller machine (ATM) and its perceived ease of use. Furthermore, the manipulated aesthetics of the ATM layout was the only factor affecting postexperimental perception of its usability and aesthetics as well. The reported findings support the assumption of a direct influence of a system's aesthetics on the perceived usability.

The findings of Tractinsky et al. [8] inspired follow-up studies on the interplay between aesthetics and usability. Hassenzahl and Monk [4] reviewed fifteen studies dealing with this issue. Most of the included studies were correlational in nature. Results showed a huge variation in the obtained correlations, with coefficients ranging from .00 to .92, with a median of .49. The authors assumed that the variability of results may be due to the studies' inconsistency in method and analysis. Hassenzahl and Monk [4] themselves argued

that beauty will play an important role as a starting point for judging products because its primarily sensory nature makes it one of the most immediately available qualities. But in a series of correlational studies with different websites, they found no direct relation between perceived beauty and perceived usability (see also Hassenzahl [11] as well as van Schaik and Ling [12]). Instead, the relationship between perceived beauty and perceived usability was found to be wholly mediated by goodness. The authors concluded that the relationship between beauty and usability has been overplayed [4, page 235]. In a follow-up paper van Schaik et al. [13] could replicate the indirect link between perceived beauty and perceived usability while the effect of beauty on hedonic quality was primarily direct. With respect to the different effects of perceived beauty on the perception of perceived usability and hedonic quality, van Schaik et al. [13] assumed two different ways of inference: (1) probabilistic consistency as an inference rule concerning the primarily *direct* effect of perceived beauty on hedonic quality and (2) evaluative consistency as an inference rule concerning the primarily *indirect* effect of perceived beauty on pragmatic quality, mediated by goodness. The authors argued that individuals follow the latter rule if product qualities, like usability (even with hands-on experience), are not easily available and are not conceptually or causally linked to an available quality (i.e., beauty). In this case, people infer from the general evaluation of the product (goodness) on its usability, according to the rule of "I like it, it must be good on all attributes" [13, page 18]. Goodness itself is assumed to be directly linked to beauty, the starting point of the whole inference process [13, page 4].

In discussing their findings, van Schaik et al. [13] proposed that the task for further research is to find evidence for the different inference rules and suggested experimental manipulations, for example, of design characteristics, that contribute to the "objective" usability of a product. Moreover, Hassenzahl and Monk [4] claimed that more experimental studies are needed to test causal models of the interplay between instrumental and noninstrumental qualities and their effect on the perception of a system's quality because a principle limitation of correlational studies is that they cannot prove causal relationships.

Indeed, on comparing correlational studies to experiments dealing with the interplay between usability and aesthetics, we have to consider a substantial methodological distinction. Following a correlational approach, van Schaik et al. [13] (see above), for example, let their participants evaluate websites with a given usability and beauty. Consequently, effects concerning the interplay between *perceived* aesthetics and *perceived* usability resulting from this or any other study with a comparable design, that is, without *experimental manipulation* of the usability and the aesthetics factors, may only be traced back to interindividual differences concerning the perception or evaluation of product qualities independent of their factual manifestation in terms of aesthetics and usability.

In contrast, experiments manipulating aesthetics and usability allow us to assess the impact of these variables on users' reactions. Consequently, Tuch et al. [10], like Hassenzahl and Monk [4] as well as van Schaik et al. [13], claimed that the independent manipulation of aesthetics and

usability is crucial for drawing causal conclusions about the aesthetics-usability relation.

Actually, apart from the Tractinsky et al.'s [8] study, there are still very few experiments on the interplay between aesthetics and usability as well as on its effects on perceived aesthetics (i.e., beauty), on perceived usability, and on broader evaluations like hedonic quality or appeal. Hence, currently there is inconclusive evidence on this issue.

Experimental support for the "what is beautiful is usable" hypothesis was found by Ben-Bassat et al. [14] and Lee and Koubek [15], as well as by Sonderegger and Sauer [16]. Ben-Bassat et al. [14] found a computerized phone book with high aesthetics to be perceived as "slightly more usable" than that with low aesthetics. Remarkably, the authors also reported a positive effect of the usability factor on perceived aesthetics. Findings of an experiment by Lee and Koubek [15] using four versions of a website with different levels of usability and aesthetics revealed a significant effect of manipulated aesthetics on perceived usability, and, corresponding to the results of Ben-Bassat et al. [14], a positive effect of manipulated usability on perceived aesthetics. Moreover, Lee and Koubek [15] found a significant effect of aesthetics, but not for the usability factor, on user preference before actual use while preference for the website after use was significantly affected by both factors. Finally, Sonderegger and Sauer [16] reported a positive effect of manipulated visual appearance of a mobile phone on its perceived usability, while Sonderegger et al. [17] found in a longitudinal field experiment that the positive effect of manipulated aesthetics of a mobile phone on its perceived usability wanes over time.

In contrast, no support for the "what is beautiful is usable" hypothesis was found by Thüring and Mahlke [18] as well as by Tuch et al. [10]. In two of three experiments in which the usability and the aesthetics of a portable digital audio player were varied, Thüring and Mahlke [18] did not find an effect either (however, a "trend" with a P value $< .10$ in the second experiment) of manipulated aesthetics on perceived usability or of manipulated usability on perceived visual aesthetics. However, usability affected the overall judgment of the audio player in both experiments and manipulated aesthetics in one of them. In a recent experiment, Tuch et al. [10] found neither an effect of manipulated aesthetics nor an effect of manipulated usability of an online shop on perceived usability at the preuse phase. Even after interacting with the online shop, interface aesthetics had no influence on users' perception of its usability. However, perceived aesthetics and a subcomponent of hedonic quality (hedonic quality identity, AttrakDiff questionnaire [19]) were not only affected by interface-aesthetics but also, and according to Ben-Bassat et al. [14] and Lee and Koubek [15], by interface usability. The authors concluded that the "what is beautiful is usable" notion can be reversed to a "what is usable is beautiful" under certain circumstances [10], characterizing a further variant of the interplay between aesthetics and usability.

To sum up, the reviewed experiments showed mixed results regarding the interplay between usability and aesthetics as well as its effects on perceived aesthetics (i.e., beauty), on perceived usability, and on broader evaluations like hedonic quality and appeal. While some findings supported

the "what is beautiful is usable" hypothesis [8, 14–16], others did not [10, 17, 18]. Remarkably, Ben-Bassat et al. [14], Lee and Koubek [15], and Tuch et al. [10] found perceived beauty to be affected by manipulated usability. Hence, in three of the summarized experiments the reversed effect to the "what is beautiful is usable" hypothesis was found ("what is usable is beautiful").

Knowledge about the interplay between aesthetics and usability is highly relevant in the field of usability engineering and prototyping. For example, concerning the validity of usability tests, it is of special interest whether the aesthetics of a prototype may influence the rating of its usability in a post-use evaluation, and vice versa. Therefore, further evidence on the interplay between instrumental and noninstrumental qualities seems to be essential.

2. The Present Study

With respect to the inconclusive evidence on the interplay between aesthetics and usability of interactive systems, the current study aims to further clarify the effect of the usability of a system on its perceived aesthetics [10, 14, 15] as well as the effect of aesthetics on a system's perceived usability [8, 14–16]. Moreover, research on this issue is extended by scrutinizing the effect of both aesthetics and usability not only on hedonic quality (see [10]) but also on the evaluation of a system's perceived appeal. Finally, with reference to the results of Hassenzahl and Monk [4] as well as van Schaik et al. [13], the indirect (mediated) influence of aesthetics/beauty on perceived usability will be investigated. Importantly, in the following we distinguish between the term "manipulated aesthetics" denoting different experimentally induced manifestations of a system's aesthetics and its subjective valuation as "(perceived) beauty." This extends to the term "manipulated usability" and its subjective counterpart "perceived usability."

The first research question (RQ1) addresses the direct effect of manipulated aesthetics and manipulated usability on beauty and perceived usability, respectively, as follows:

H1.1: manipulated aesthetics affects perceived usability;

H1.2: manipulated usability affects beauty.

The second research question (RQ2) extends previous experimental research on the interplay between usability and aesthetics with respect to additional noninstrumental product evaluations, especially hedonic quality and appeal. Two hedonic attributes may be distinguished: stimulation (HQS) and identity (HQI) [11]. Different studies revealed hedonic quality identity to be correlated to perceived beauty [4, 11, 13]. Consistent with these findings, Tuch et al. [10] reported an effect of aesthetics on HQI but not on HQS. Additionally, Tuch et al. [10] found an effect of usability on HQI. Appeal is considered to be a global judgment about a product (e.g., good versus bad [21]). Hassenzahl [11] reported that both perceived hedonic and pragmatic qualities significantly contribute to the evaluation of the general appeal of products. Concerning this research question, all hypothesized effects are supposed to be direct effects:

FIGURE 1: The effect of beauty on perceived usability mediated by goodness.

H2.1: manipulated aesthetics affects perceived hedonic quality identity;

H2.2: manipulated usability affects perceived hedonic quality identity/stimulation;

H2.3: manipulated aesthetics affects perceived appeal;

H2.4: manipulated usability affects perceived appeal.

The third research question (RQ3) focuses on the *indirect* effects of aesthetics and beauty on perceived usability. With respect to given evidence, different indirect effects are conceivable. Hassenzahl and Monk [4] and van Schaik et al. [13] pursued the assumption that beauty may affect subsequent judgments of other product qualities because of its immediate sensory availability. Both found in their correlational studies an indirect instead of a direct effect of perceived beauty on perceived usability ("evaluative consistency as an inference rule, see above), mediated by goodness (see Figure 1). In the present study, we firstly aim to replicate this effect under experimental conditions. Secondly, we aim to expand research on this issue by analyzing the mediating role of goodness with respect to the effect of not only perceived beauty on perceived usability but also of manipulated aesthetics as well. This mediator model represents a variant of the evaluative consistency rule (see above, [13]):

H3.1: the effect of (perceived) beauty on perceived usability is mediated by goodness;

H3.2: the effect of manipulated aesthetics on perceived usability is mediated by goodness.

Additionally, a further indirect effect of aesthetics on perceived usability may be hypothesized in light of evidence from social psychology regarding the impact of a beautiful outer appearance in humans on the perception of other qualities, like happiness or intelligence (the so-called halo effect). Good-looking men and women, for example, are judged as being more intelligent than those who are not [22]. Transferred to the interplay between aesthetics and usability, we assume that the effect of the given usability of a system on its perceived usability will be moderated by its manipulated aesthetics (see Figure 2). To test this assumption, the effect of manipulated aesthetics on perceived usability is viewed as a moderator effect:

H3.3: the effect of manipulated usability on perceived usability is moderated by manipulated aesthetics.
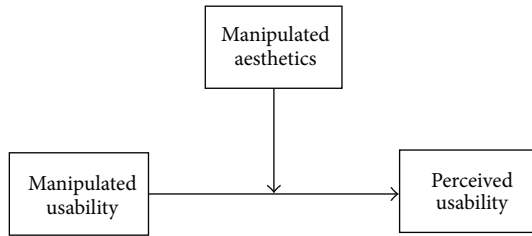
FIGURE 2: Moderating effect of manipulated aesthetics on the effect of manipulated usability on perceived usability.

3. Method

3.1. Design. To test the hypotheses, an experiment with a 2×2 factorial design with usability and aesthetics of a mobile phone prototype as independent variables and perceived usability, perceived beauty, hedonic qualities, and appeal as dependent variables, as well as goodness as a mediating variable, was used.

3.2. Materials. The subject of the experiment was a computer-based prototype of a mobile phone. In usability engineering prototypes are commonly used to evaluate human-system interfaces. Prototypes in this field may take different forms including paper and cardboard mock-ups, software simulations or fully working prototypes [3, 23]. Particularly, if design involves hardware implementations, using software tools for design simulation provides an economic alternative to hardware prototypes [24]. According to this approach, the mobile phone prototype used in the current experiment was realized as an interactive software simulation.

Four different versions of the prototype were designed, one for each combination of factor levels (high versus low aesthetics and high versus low usability). The manipulation of the aesthetics factor was accomplished in a two-step procedure.

First, a sample of 12 participants with a mean age of 23.9 years (SD = 2.69; eight females) was interviewed. They were asked to name attributes of mobile phones that they perceive as beautiful or ugly, respectively. Subsequently, a content analysis of interview responses was carried out to identify visual attributes (e.g., the color of the hardcover and its shape) that constitute the beauty of a mobile phone but that are unrelated to its usability (e.g., the size of the screen or the kind of key panel). Based on the results of this analysis, a professional designer created seven paper-prototypes varying in beauty.

In a second run, another sample of 28 participants (mean age: 27.53; SD = 8.31; 17 females) ranked pictures of these paper prototypes with respect to their beauty. Resulting mean ranks were subsequently computed to identify the most beautiful mobile (high aesthetics) as well as the ugliest one (low aesthetics). In line with the findings of K. O. Götz and K. Götz [25], the ugliest version was characterized by a mixture of several colors, whereas a plain hardcover was perceived as beautiful. Moreover, a pendant, little gemstones on the hardcover and a smaller form factor significantly reduced the mobile phone's beauty while the size and distance between the

FIGURE 3: High (a) and low aesthetics design (b) of the mobile prototype.

GUI-elements were held constant for both versions. Figure 3 depicts the two hardcover versions.

The usability of the prototypes was manipulated with regard to (a) the depth and width of the hierarchical menu tree (cf. [26]), (b) the position of target functions at a certain menu level, and (c) the labeling of menu categories on the second menu level (i.e., more or less unambiguous labels).

Accordingly, the menus of the low usability prototypes are, in contrast to the high usability versions, more complicated due to their breadth and depth. That is, participants in the low usability conditions had to navigate (vertically) through more menu levels until they could achieve a target item. In addition to the depth of the menu trees, the breadth of menu levels was expanded by providing additional menu options serving as distractors (see Figure 4 for two examples of menu trees for the low and high usability prototypes). Moreover, the target menu items in the low usability versions were placed far at the end of the menu level to hamper fast navigation, and finally, item names were longer as well as less common with regard to the underlying function.

By varying the usability (high/low) and aesthetics of the prototype (high/low) four experimental conditions were accomplished. Finally, the prototypes were implemented in Adobe Flash. Participants interacted with the flash-prototypes via touch screen. None of the participants that were involved in the development of the mobile prototypes participated in the main experiment.

3.3. Measures and Instruments. To measure perceived usability, hedonic experience, and the overall appeal of the prototype the *AttrakDiff2* questionnaire [19] (see as well http://attrakdiff.de/) was employed. This questionnaire comprises four subscales: pragmatic quality (PQ) as a measure for perceived usability (Cronbach's alpha = .88), hedonic quality identity (HQI; alpha = .77), and hedonic quality stimulation (HQS; alpha = .78). Additionally the questionnaire provides a subscale called attractiveness (APPEAL; alpha = .91) that
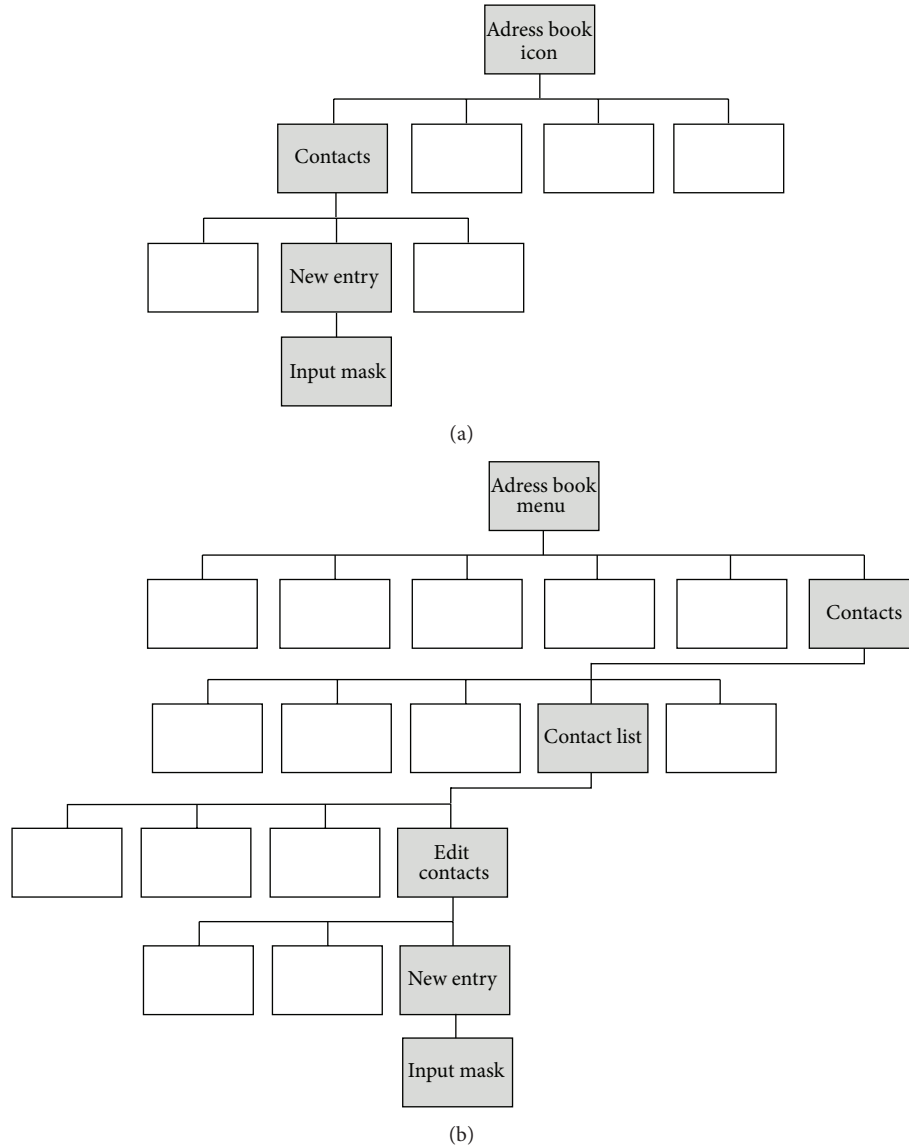
FIGURE 4: Menu trees of the high usable (a) and the low usable prototype (b); for example: enter a new contact into the address book.

evaluates a product's appeal [11]. Each of the subscales consists of seven seven-point semantic differential items with bipolar verbal anchors. The pragmatic quality scale asks participants, for example, to rate a product as "simple versus complicated." The concept of hedonic quality is split into two aspects: identity (HQI) describes the ability of a system to communicate a desirable identity to others (e.g., "isolating versus coupling"), while stimulation (HQS) assesses the amount a product supports striving for personal development (e.g., "original versus conventional"). The attractiveness scale (APPEAL) assesses the global (overall) judgment of the interactive product (e.g., "good versus bad"; "unpleasant versus pleasant"; "ugly versus attractive"; sympathetic versus unsympathetic") [21].

To gauge the perceived beauty of the prototype, participants rated the mobile phone prototype on a 7-point scale ("not beautiful" to "beautiful"). This measure was also applied to check the manipulation of the mobile phone's aesthetics. In addition, perceived beauty was measured with a single item ("ugly-beautiful") taken from the attractiveness scale of the AttrakDiff2 questionnaire.

In accordance to van Schaik et al. [13], perceived goodness was measured by the corresponding single-item "good-bad" of the attractiveness scale of the AttrakDiff2 questionnaire.

The concept of usability is defined in terms of effectiveness, efficiency, and satisfaction [27]. Metrics of efficiency refer to the cognitive as well as physical workload required to achieve a task [24, 28]. In the current study the *NASA-TLX* questionnaire [29] was applied to measure the workload associated with using the mobile phone prototypes. The *NASA-TLX* questionnaire is a multidimensional rating procedure comprised of six single-item scales assessing participants' mental ("*How mentally demanding was the task?*"), physical ("*How physically demanding was the task?*"), and temporal demands ("*How hurried or rushed was the pace of the*

task?"), as well as an assessment of self-performance ("*How successful were you in accomplishing what you were asked to do?*"), effort ("*How hard did you have to work to accomplish your level of performance?*"), and current frustration derived from interacting with a system ("*How insecure, discouraged, irritated, stressed, and annoyed were you?*"). The mean across all scales was calculated for the perceived workload (alpha = .87). Following Pfendler [30], we did not apply a subsequent pairwise comparison to create an individual weighting of subscales because weighted and nonweighted scale values are highly correlated ($r = .94$), while nonweighted scales show a higher reliability in the German version of the NASA-TLX. Hence, all scales were nonweighted, that is, weighted equally.

Before experimental tasks were performed, participants completed a questionnaire assessing mobile expertise with respect to duration and intensity of use.

3.4. Main Study Participants. Eighty-eight participants (69 females) with a mean age of 23.1 years (SD = 6.7) took part voluntarily for course credits in the main study. Gender was counterbalanced across experimental conditions. All participants were mobile phone users, whereby 53.9% had regularly used a mobile phone for at least 5 years and 36.0% for seven or more years. Moreover, 71.9% of them reported using their mobile phone several times a day, 20.2% even several times per hour.

To control the effect of mobile phone expertise we computed an overall score for participants' mobile phone expertise. This score includes the frequency of mobile phone use (five levels from "less than once a week" to "several hours per day"), the length of mobile phone use (five levels from "less than one year" to "more than seven years"), and the intensity of mobile phone use (mean frequency of 23 given mobile phone functions and apps, rated from "seldom" to "very often"). All three measures were additively as well as multiplicatively combined to the overall score for mobile expertise. A 2 × 2 (manipulated usability × manipulated aesthetics) analysis of variance (ANOVA) revealed no significant effects on mobile phone expertise, independent of the combination rule applied, with all $P \geq .276$. Hence, we did not find differences in mobile expertise between the experimental groups as a possible source of experimental bias. Additionally, we checked whether the results of the following hypothesis tests were affected by mobile phone expertise. For this purpose, we included the expertise score to all ANOVAs as a covariate. However, none of the results changed (all significant effects remained significant; all nonsignificant results remained nonsignificant) when including mobile phone expertise.

The study conformed to the Code of Ethics of the American Psychological Association (http://www.apa.org/ethics/code/principles.pdf) and to the national guidelines (http://www.dgps.de/).

3.5. Procedure. Participants were randomly assigned to one of the four experimental conditions. As a cover story, they were told to take part in a usability test to improve a prototype for a mobile phone manufacturer. Before the experimental procedure started, participants were asked to complete the questionnaire to gauge mobile phone expertise. After that, the mobile phone prototype was presented on a Tablet PC. Illumination in the laboratory was controlled to keep the visual appearance of the mobile phone prototypes constant. Participants were introduced to the computer-based prototype to become familiar with its interface and to train to navigate through the menu options by using a touch pen. Afterwards, demographic variables were collected and participants were introduced to the experimental tasks. Participants were requested to read task-scenarios prior to the test. If necessary, scenarios were clarified by the experimenter who acted in the experiment as the conductor of the usability test.

At the beginning of the experiment, participants were instructed to complete three tasks addressing different mobile phone functions in the following order: phone book entry, cost manager, and changing the current ringtone. The tasks covered a wide range of more or less frequently used mobile phone functions with varying difficulty. For each task, completion time and the number of navigation errors were logged automatically. Because we were not able to exclude the possibility that subjects could transfer what they had learned about the mobile's interface in one task to the next, the sequence of tasks was held constant for all participants. Thus, in favor for the between-subject design we avoided additional noise in the data by waiving pseudorandomization of tasks.

After task completion participants evaluated the pragmatic quality of the mobile phone as well as its perceived appeal and hedonic quality by means of the AttrakDiff2 questionnaire. Moreover, participants answered the single item (7-point Likert scale) asking for the perceived beauty of the mobile phone. Subsequently participants completed the NASA-TLX to assess subjective workload. Finally, all participants were debriefed and received course credits, if desired.

3.6. Data Analysis. Hypotheses were tested with a 2 × 2 (manipulated usability × manipulated aesthetics) ANOVA and regression models. All statistical tests were performed with a significance level of 5%. ANOVAs were calculated regardless of variance (in)homogeneity because analysis of variance is considered robust to violations of variance homogeneity when the sample sizes are equal [31, 32].

Referring to Cohen [33], effect sizes are reported by partial eta squared η_p^2 (benchmark values: .01 = small, .06 = medium, .14 = large). However, we want to emphasize that these values are quoted as benchmark values for both η_p^2 and the classical eta squared (η^2), whereby the former is typically greater than the latter (cf. [34]). Hence, the respective values should be interpreted with some caution. In the case of t-tests we used Cohen's d as a measure of effect size (0.2 = small; 0.5 = medium; 0.8 = large).

4. Results

4.1. Manipulation Check: Aesthetics. To validate the aesthetics manipulation of the prototype, a 2 × 2 ANOVA (manipulated usability × manipulated aesthetics) was computed. As intended by the manipulation, we found a main effect of manipulated aesthetics on perceived beauty (7-point scale)
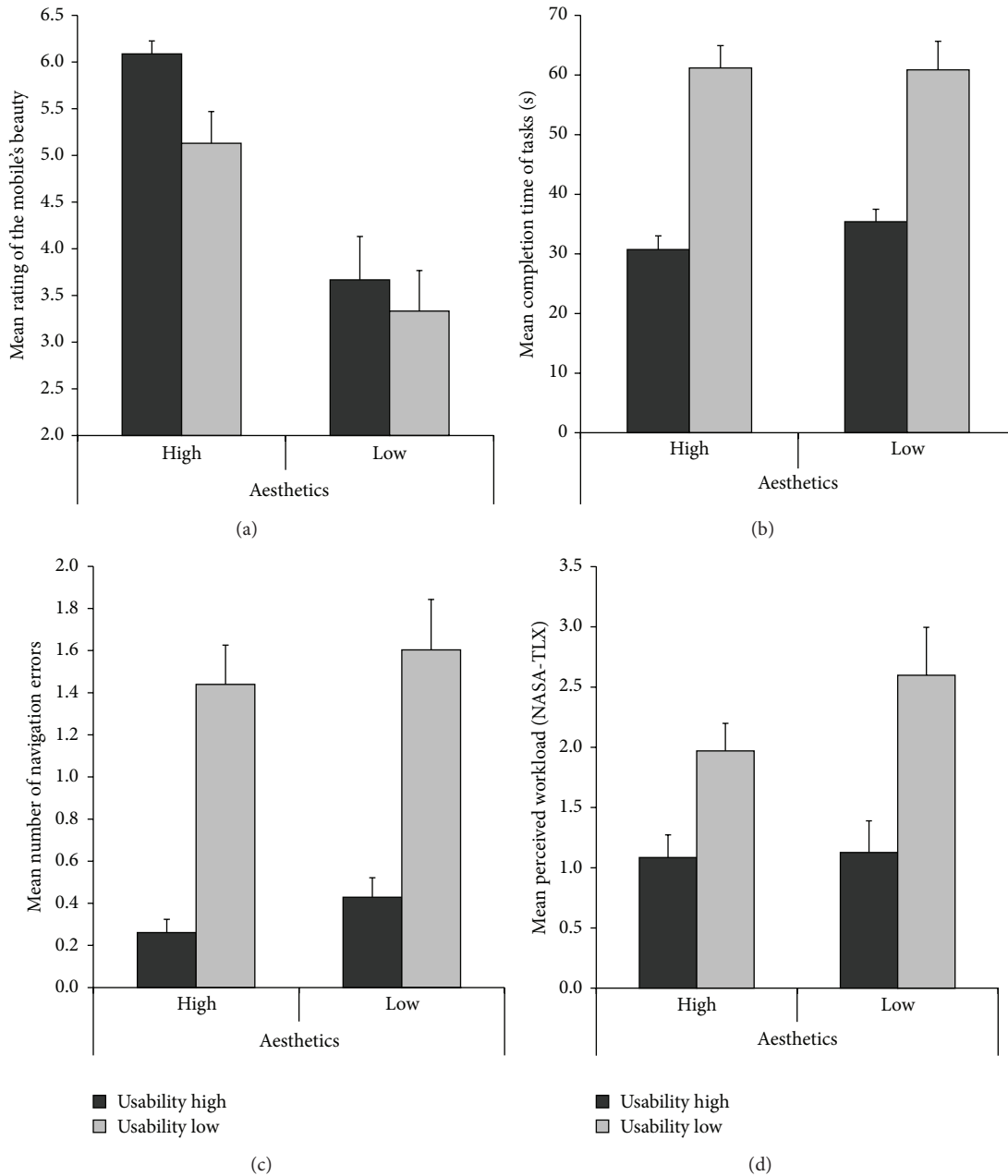
(a)

(b)

(c)

(d)

FIGURE 5: The impact of the mobile versions differing in manipulated usability (high versus low) and manipulated aesthetics (high versus low).

$(F(1, 84) = 34.489; P < .001; \eta_p^2 = .291)$. The high-aesthetic prototype was rated as more beautiful than the low-aesthetic version (Figure 5(a)). By trend also a main effect of manipulated usability revealed a higher beauty rating for the high-usable prototypes versus low-usable prototypes $(F(1, 84) = 3.226; P = .076; \eta_p^2 = .037)$, but the effect size was considerably smaller than for the effect of manipulated aesthetics on perceived beauty. No interaction between manipulated usability and aesthetics was observed $(F(1, 84) = .753; P = .388; \eta_p^2 = .009)$. The results of the ANOVA did not change when using the beauty item of the AttrakDiff2

questionnaire as dependent variable instead of the 7-point scale (not depicted).

4.2. Manipulation Check: Usability. To check the usability manipulation we first compared mean task completion times with a 2 × 2 ANOVA (manipulated usability × manipulated aesthetics). A significantly longer completion time was found for the low-usable mobile phones in contrast to the high-usable ones $(F(1, 83) = 68.201; P < .001; \eta_p^2 = .451)$ (Figure 5(b)), but neither a main effect of the mobile phone's aesthetics occurred $(F(1, 83) = .409; P = .524; \eta_p^2 = .005)$

nor an interaction between usability and aesthetics was found ($F(1, 83) = .543$; $P = .463$; $\eta_p^2 = .006$). We also checked whether potential outliers, even if they showed realistic task completion times, biased these results. For this purpose, two participants with completion times that were more than two standard deviations above the mean across subjects were excluded from a reanalysis. However, the results of the ANOVA did not change considerably; that is, no change in significance appeared.

The mean number of navigation errors was also analyzed for the prototype versions. A 2×2 ANOVA (manipulated usability $\times$ manipulated aesthetics) revealed a main effect of manipulated usability on the number of navigation errors with more errors when using low-usable prototypes ($F(1, 83) = 54.484$; $P < .001$; $\eta_p^2 = .396$) (Figure 5(c)). We neither found an effect of manipulated aesthetics ($F(1, 83) = 1.081$; $P = .301$; $\eta_p^2 = .013$) nor an interaction effect ($F(1, 83) < .001$; $P = .990$; $\eta_p^2 < .001$).

Additionally, we computed the mean workload score across the six items of the NASA-TLX. A 2×2 ANOVA (manipulated usability $\times$ manipulated aesthetics), which revealed higher demands for the low-usable prototypes compared to the high-usable prototypes ($F(1, 84) = 18.364$; $P < .001$; $\eta_p^2 = .179$) (Figure 5(d)), but neither a main effect of the mobile phone's manipulated aesthetics occurred ($F(1, 84) = 1.484$; $P = .227$; $\eta_p^2 = .017$) nor an interaction effect was found ($F(1, 84) = 1.134$; $P = .290$; $\eta_p^2 = .013$).

Finally, the high-usable mobile phone was perceived as more usable (PQ scale AttrakDiff2), as revealed by a 2×2 ANOVA (manipulated usability $\times$ manipulated aesthetics) ($F(1, 84) = 16.128$; $P < .001$; $\eta_p^2 = .161$) (see Figure 6). We neither found a main effect of manipulated aesthetics nor an interaction effect (see Section 4.3-H1.1 below for statistical results).

To conclude, independent of manipulated aesthetics, the high-usable prototype enabled better task performance, led to less workload, and was perceived as more usable than the low usability versions. Consequently, we may consider the usability manipulation successful.

4.3. Results on Hypotheses

H1.1: Manipulated Aesthetics Affects Perceived Usability. As shown above, manipulated usability affected the perceived usability of the mobile phone prototype in the intended way. However, in contrast to the term "what is beautiful is usable," a 2×2 ANOVA revealed no a main effect of manipulated aesthetics on perceived usability (PQ) ($F(1, 84) = .497$; $P = .483$; $\eta_p^2 = .006$) (Figure 6) and also no interaction effect between the mobile's usability and its aesthetics ($F(1, 84) = 1.981$; $P = .163$; $\eta_p^2 = .023$). Hence, no evidence for the "what is beautiful is usable" notion was found.

H1.2: Manipulated Usability Affects Perceived Beauty. The results of the manipulation check for the aesthetics factor showed a main effect of manipulated aesthetics on perceived beauty, but no interaction between manipulated usability
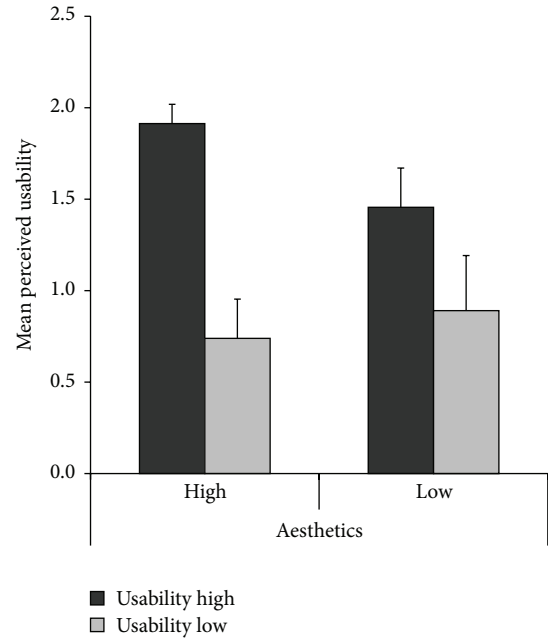

FIGURE 6: Mean rating of the perceived usability of the mobile depending on the mobile's manipulated usability (high versus low) and the manipulated aesthetics (high versus low). Note that the scale ranges from -3 to $+3$. Vertical lines indicate the standard error of the mean.

and manipulated aesthetics was found. Moreover, higher beauty ratings were observed for the high-usable versus the low-usable prototypes (see Figure 5(a)). Accordingly, a 2×2 ANOVA revealed a main effect of manipulated usability on beauty ratings that, however, narrowly missed the significance level of 5%. This result was independent of whether we used the single-item scale measuring beauty ($F(1, 84) = 3.226$; $P = .076$; $\eta_p^2 = .037$) or the item "ugly-beautiful" taken from the APPEAL scale of the AttrakDiff2 questionnaire ($F(1, 84) = 3.053$; $P = .084$; $\eta_p^2 = .035$). However, taking into account that Tuch et al. [10] coined the term "what is usable is beautiful," we can consider this hypothesis as one-sided, licensing halved P values for the corresponding F-values as suggested by McNeill et al. [35]. Given that this leads to P values of $P = .038$ and $P = .42$, respectively, we hence may conclude that these results support the "what is usable is beautiful" notion.

H2.1: Manipulated Aesthetics Affects Hedonic Quality Identity/Stimulation. A 2×2 ANOVA (manipulated usability $\times$ manipulated aesthetics) revealed a large-sized main effect of the phone's manipulated aesthetics on perceived hedonic quality "identity" (HQI) ($F(1, 84) = 24.105$; $P < .001$; $\eta_p^2 = .223$) with higher ratings for the high-aesthetic version in contrast to the low-aesthetic version (Figure 7) while a further 2×2 ANOVA with hedonic quality "stimulation" (HQS) as dependent variable did not show an effect of manipulated aesthetics ($F(1, 84) = .655$; $P = .421$; $\eta_p^2 = .008$). Neither an interaction effect between manipulated usability and
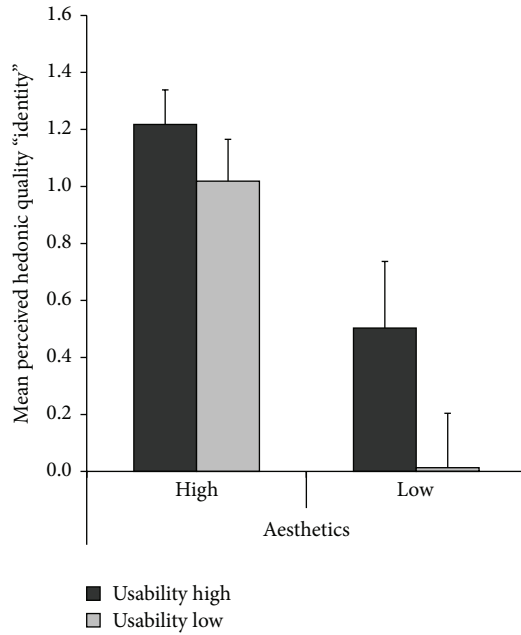
FIGURE 7: Mean rating of HQI, depending on the mobile's manipulated usability (high versus low) and the manipulated aesthetics (high versus low). Note that the scale ranges from −3 to +3. Vertical lines indicate the standard error of the mean.
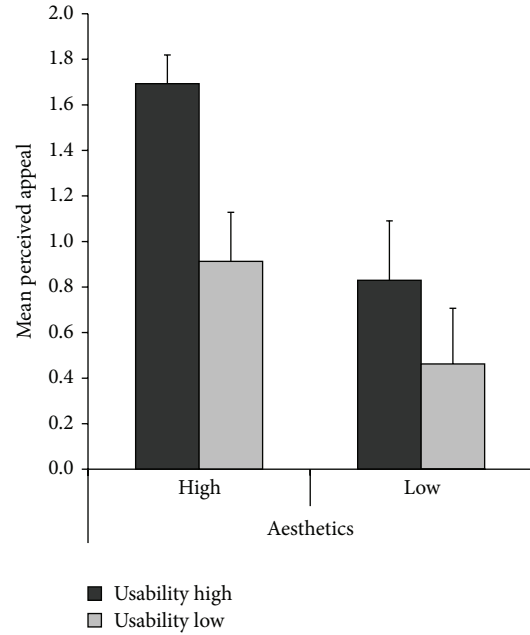

FIGURE 8: Mean rating of the perceived appeal of the mobile depending on the mobile's manipulated usability (high versus low) and the manipulated aesthetics (high versus low). Note that the scale ranges from −3 to +3. Vertical lines indicate the standard error of the mean.

aesthetics for HQI ($F(1, 84) = .691$; $P = .408$; $\eta_p^2 = .008$) nor for HQS ($F(1, 84) = 2.088$; $P = .152$; $\eta_p^2 = .024$) was revealed. Hence, the effect of manipulated aesthetics on HQI was not moderated by manipulated usability. The main effect of usability on HQI and HQS is presented below in "H2.2".

H2.2: Manipulated Usability Affects Hedonic Quality Identity/Stimulation. Besides the main effect of manipulated aesthetics on HQI and HQS, the ANOVA also showed a small-to medium-sized effect of the manipulated usability on HQI ($F(1, 84) = 3.868$; $P = .053$; $\eta_p^2 = .044$), with higher ratings for the high usable mobile phone. Unlike HQI, the analysis of variance revealed no main effect of manipulated usability on HQS ($F(1, 84) = .217$; $P = .643$; $\eta_p^2 = .003$).

Following Tuch et al. [10], we tested which items of HQI were affected by the system's usability and aesthetics with a series of 2 × 2 (manipulated usability × manipulated aesthetics) ANOVAs. In contrast to Tuch et al. [10], most items of the HQI scale of the AttrakDiff2 questionnaire were sensitive to the aesthetics manipulation, while only one item revealed a clear effect for the usability manipulation (non-inclusive-inclusive; see Table 1). None of the items showed an interaction effect (usability ∗ aesthetics).

H2.3: Manipulated Aesthetics Affects Perceived Appeal. To test the effect of manipulated aesthetics on perceived appeal, we computed a 2 × 2 ANOVA with perceived appeal as dependent variable (APPEAL scale, AttrakDiff2 questionnaire). Results of the analysis show a main effect of the manipulated aesthetics on perceived appeal ($F(1, 84) = 9.307$; $P = .003$;

$\eta_p^2 = .100$). The high aesthetic prototype was rated more appealing than its counterpart (Figure 8).

H2.4: Manipulated Usability Affects Perceived Appeal. In addition to the main effect of manipulated aesthetics on perceived appeal, an ANOVA revealed also a main effect of manipulated usability on perceived appeal ($F(1, 84) = 7.100$; $P = .009$; $\eta_p^2 = .078$). The high-usable version was rated as more appealing than the low-usable version independent of the manipulated aesthetics of the prototype (Figure 8). Manipulated usability did not interact with the manipulated aesthetics ($F(1, 84) = .917$; $P = .341$; $\eta_p^2 = .011$). It is important to note that the effects of manipulated usability and aesthetics on perceived appeal were similar in size. Consequently, the mobile's usability as well as its aesthetics contributed to the mobile phone's perceived appeal to more or less the same extent.

H3.1: The Effect of Perceived Beauty on Perceived Usability Is Mediated by Goodness. The mediation model was tested using regression analysis according to Baron and Kenny [36]. Both operationalizations of perceived beauty were positively correlated with perceived usability (PQ, AttrakDiff2) (single-item scale: $r = .407$, $P < .001$; APPEAL item: $r = .410$, $P < .001$). Moreover, both items correlated positively with perceived goodness (single-item scale: $r = .676$, $P < .001$; APPEAL item: $r = .720$, $P < .001$). In the third step (done separately for both operationalizations of perceived beauty), perceived beauty and perceived goodness were entered as predictors while PQ served as criterion in the multiple

TABLE 1: Effects of manipulated beauty and manipulated usability on HQI items.

HQI items	Manipulated aesthetics			Manipulated usability			Usability ∗ aesthetics		
	$F(1, 84)$	P	η_p^2	$F(1, 84)$	P	η_p^2	$F(1, 84)$	P	η_p^2
Takes me distant from people-brings me closer to people	0.949	.333	.011	1.110	.295	.013	0.198	.657	.002
gaudy-classy	38.397	<.001	.314	2.105	.151	.024	<0.001	.992	<.001
Cheap-valuable	16.316	<.001	.163	2.146	.147	.025	0.241	.625	.003
Noninclusive-inclusive	4.420	.039	.050	8.452	.005	.091	0.632	.429	.007
Isolating-integrating	3.836	.055	.044	3.127	.081	.036	0.109	.742	.001
Amateurish-professional	0.641	.426	.008	0.927	.338	.011	1.939	.167	.023
Unpresentable-presentable	27.081	<.001	.244	2.985	.088	.034	0.024	.878	<.001

Inverted items were reversed so that higher values always indicate a higher HQI value.

TABLE 2: Results for the mediation model specified in H3.1. The model was tested with regression analysis predicting perceived usability (PQ-scale AttrakDiff2 questionnaire) on the basis of perceived beauty and perceived goodness. Perceived beauty was measured with a 7-point rating scale (Model 1) and by a single-item of the APPEAL scale of the AttrakDiff2 questionnaire (Model 2).

Variable	Coefficient			R^2	Model statistics		
	β	t	P		adj. R^2	F	P
Model 1				.434	.421	33.033	<.001
Beauty	−0.069	−0.629	.531				
Goodness	0.704	6.394	<.001				
Model 2				.440	.427	33.802	<.001
Beauty	−0.131	−1.127	.263				
Goodness	0.752	6.460	<.001				

regression equation. The result of the analysis revealed that perceived goodness completely mediated the relationship between perceived beauty and PQ (Table 2), independent of the operationalization of perceived beauty. The effect of perceived beauty on PQ, controlled for perceived goodness, was nonsignificant, that is, zero. The full mediation by goodness was independent of the manipulated aesthetic condition. When computing the mediation analysis separately for the beautiful and ugly prototype, respectively, the results did not change from those reported in Table 2.

H3.2: The Effect of Manipulated Aesthetics on Perceived Usability Is Mediated by Goodness. We also hypothesized an effect of manipulated aesthetics on perceived usability, and, if so, that this link was mediated by perceived goodness. However, no correlation between manipulated aesthetics (dummy-coded) and perceived usability (PQ) was found ($r = .069$, $P = .520$) while manipulated aesthetics and perceived goodness showed a significantly positive correlation ($r = .287$, $P = .007$). Hence, no support for hypothesis H3.2 was found.

H3.3: The Effect of Manipulated Usability on Perceived Usability Was Moderated by Manipulated Aesthetics. As expected, results of the manipulation check and on hypothesis H1.1 revealed a significant effect of manipulated usability on

perceived usability (PQ), but this effect was not moderated by manipulated aesthetics. Moreover, manipulated aesthetics did not show a significant impact on perceived usability. Therefore, H3.3 is not supported by the results, as the null hypothesis cannot be rejected.

5. Discussion

The present paper dealt with the interplay between usability and aesthetics of a mobile phone prototype in the area of usability evaluation. The first research question addressed the effect of the system's aesthetics and usability on perceived beauty and perceived usability, respectively. Results revealed no effect of manipulated aesthetics on perceived usability but, in accordance with Ben-Bassat et al. [14], Lee and Koubek [15], and Tuch et al. [10], an effect of manipulated usability on perceived beauty, such that the high usable prototype was rated as more beautiful than the low usable prototype. It is important to note that in the present study the manipulation of the prototype's aesthetics as well as its usability could be realized to a broadly similar degree, as indicated by the nearly similar effect sizes for both factors. Therefore, we can rule out that an inconsistent manipulation of the two factors may have

eliminated the "what is beautiful is usable" effect, as discussed by Tuch et al. [10, page 21].

So far, results support the "What is usable is beautiful" in favor of the "What is beautiful is usable" hypothesis. This finding conforms to the viewpoint of functionalism in architecture and industrial design from the early 1920s of the last century stating that the aesthetic properties of an object depend on its functionality [37]. However, in accordance with the thesis of aesthetic dualism [37], assuming that the aesthetic assessment of a system may depend on its practical function and its appearance as well [38], we found an effect of manipulated aesthetics on perceived beauty too. Therefore, we should avoid a reductionist viewpoint when looking at factors influencing the aesthetic assessment of a system, even if it is supposed, in line with our and Tuch et al.'s [10] results, that satisfaction of functional requirements in most cases contributes positively to aesthetic value [37].

The second research question addressed the effect of both manipulated usability and manipulated aesthetics on variables of noninstrumental quality, namely, hedonic quality and appeal. Concerning the former dependent variable, results showed a large-sized main effect of the mobile phone's manipulated aesthetics and a small- to medium-sized effect of manipulated usability on hedonic quality "identity" (HQI). These two effects indicate that the more aesthetic and the more usable a system is designed, the higher it is rated with respect to HQI or, in other words, the better it supports communicating a desirable identity to others. These findings conform to the results of Hassenzahl [11] and Tuch et al. [10].

In accordance with Tuch et al. [10], we found that the usability manipulation affects more different HQI-Items than the aesthetics manipulation. However, the resulting patterns concerning this issue were different from those of Tuch et al. [10]. While Tuch et al. [10, page 22] found, contrary to their assumption, HQI to be mainly associated with the usability manipulation, the results of the present study support that HQI is indeed more strongly affected by a system's aesthetics.

Concerning the global evaluation of the mobile phone prototype, the present results revealed significant effects both of manipulated usability and manipulated aesthetics on perceived appeal, with comparable medium to large effect sizes (usability: $\eta_p^2 = .078$; aesthetics: $\eta_p^2 = .100$). Both factors explained a similar amount of variance in perceived appeal. Therefore, our results suggest that perceived appeal, serving for a global evaluation of noninstrumental quality [21], depends on aesthetics and usability more or less to the same extent. This finding corresponds to a result of Lee and Koubek [15], who found that manipulated aesthetics and usability both significantly affect user preference, which in turn might be determined by perceived appeal.

Regarding the "indirect" effect of aesthetics on perceived usability (research questions 3.1 and 3.2), we found that goodness completely mediates the relationship between *perceived* beauty and *perceived* usability (H3.1) but not between *manipulated* aesthetics and *perceived* usability (Hypothesis 3.2). Consequently, neither *perceived* beauty nor *manipulated* aesthetics seem to be directly linked to *perceived* usability. However, the former finding (H3.1) corresponds to the inference rule of evaluative consistency, which implies an indirect

effect of (perceived) beauty on pragmatic quality, as stated by Hassenzahl and Monk [4] as well as by van Schaik et al. [13, page 17]. Following van Schaik et al. [13], this kind of inference results in a highly beauty-driven overall evaluation ("What is beautiful is good" and "I like it, it must be good on all attributes"), which may include the judgment of usability. Concerning the argument of van Schaik et al. [13], it is noteworthy to mention that the relation between *perceived* beauty and *perceived* usability is based only on variance due to subjects. Consequently, we may assume that the mediation effect solely depends on an individual evaluative process and therefore might be better characterized as "What is *perceived* as beautiful is *perceived* as good" and "What is *perceived* as good is *perceived* as usable," independent of whether a system is really ugly or beautiful (in the sense of manipulated aesthetics). On the contrary, in the current study, the mediation analysis for *manipulated* aesthetics, goodness, and perceived usability (Hypothesis 3.2) did not reveal an effect, because manipulated aesthetics and perceived usability were not correlated. Regarding this finding, we may conclude that there is neither a direct nor an indirect link between aesthetics and perceived usability, or, in other words: people judge the usability of a system independently of whether it is ugly or beautiful (in the sense of manipulated aesthetics).

Finally we found no support for the hypothesis that manipulated aesthetics moderates the impact of manipulated usability on perceived usability (H3.3). That is, the perceived usability of a prototype with a given "objective" usability seems not to be influenced by its aesthetics; a system providing limited "objective" usability is not perceived as more usable if it is designed aesthetically.

5.1. Practical and Theoretical Implications. There are several practical implications in the current study regarding practices of usability engineering. Concerning usability evaluation, we may conclude that aesthetics of a prototype, for example, due to a high level of visual refinement, will not influence judgments about its usability in a postuse evaluation. This conclusion is supported by the present results as we found no evidence for the assumptions that (a) manipulated aesthetics affects perceived usability (H1.1), (b) there is an indirect link between aesthetics and perceived usability via goodness (H3.2), or (c) aesthetics moderates the effect of manipulated usability on perceived usability (H3.3).

With regard to user experience design, two conclusions from the current study may be helpful. First, objective usability seems to be an antecedent for perceived beauty. Second, in combination, both objective usability and aesthetic contribute to a positive perception of the overall appeal of a product and provide the user with the opportunity to express a desirable identity to others (HQI). Further research should take a longitudinal perspective on this issue and examine whether these effects also persist over time (see [17]).

If we take hedonic quality (HQI) and appeal as measures of user experience, the findings presented here lead to an integrative perspective on user experience (UX) determined by instrumental as well as noninstrumental qualities. Consequently, and in line with the ISO 9241-210 [6] definition of user experience, both qualities should be considered in user

experience design. However, the impact of additional design elements like brand [39] on user experience should be taken into account in future research.

5.2. Strength and Limitations. In the current experiment, a thorough and successful manipulation of the key variables "usability" and "aesthetics" could be examined to support the study's internal validity. To obtain ecological validity, we tried to provide a realistic setup for the usability test that served as a cover story for the experiment. The external validity of the present study might be limited by the fact that results are based on the evaluation of a computer-based prototype of just one device, namely, a mobile phone, with a given level of fidelity [20]. Regarding the latter issue, former research has shown that the fidelity, as well as other factors, such as the materials of prototypes and testing conditions, may affect the results of the evaluation of a prototype in a usability test [40, 41]. Hence, generalizability may be limited by the fact that the mobile phone prototypes used in the experiment were characterized by a specific combination of fidelity levels on the five dimensions of the mixed-fidelity model proposed by McCurdy et al. (level of visual refinement, breadth of functionality, depth of functionality, richness of interactivity, richness of data model) [20], including a high level of visual refinement as well as a sufficient level of interactivity and functionality that allowed users to perform a set of tasks and to gain hands-on experience with the mobile phone prototype in the usability test.

However, results concerning the interplay between usability and aesthetics are supported by the studies of, for example, Ben-Bassat et al. [14], Lee and Koubek [15], and Tuch et al. [10] which present partially similar findings concerning the effect of "objective" usability on perceived beauty for different systems than the prototype used here. Despite this, more experimental research on the impact of instrumental qualities like usability on noninstrumental variables regarding a broader range of interactive systems is recommended.

Unlike other studies dealing with the interplay between usability and aesthetics, in the current study the measurement of usability and aesthetics was obtained only postuse. This was done for two reasons: first, it is not common practice in usability tests to ask for a usability rating before using a system, so preuse evaluation of usability could have undermined belief in the cover story. Second, we tried to avoid a testing effect (answering a questionnaire preuse influences answers of the same questionnaire postuse) as a possible threat to internal validity. Besides, previous articles on this subject found that neither interface aesthetics nor interface usability had an effect on perceived usability at the preuse phase (see [10]).

6. Conclusion

The current study provides further significant insights into the interplay between the aesthetics and usability of interactive systems. Findings about the influence of usability on perceived beauty support the "What is usable is beautiful"

notion stated by Tuch et al. [10], while neither a direct nor an indirect effect of aesthetics on perceived usability could be supported. Thus, we do not expect that the aesthetic appearance of a prototype in a usability test will distort evaluation of its usability. Concerning user experience design, results indicate that both usability and aesthetics contribute to a positive noninstrumental valuation of a system in terms of the ability of a system to communicate a desirable identity to others (HQI) and perceived appeal.
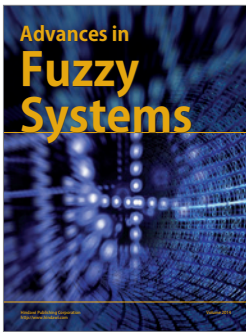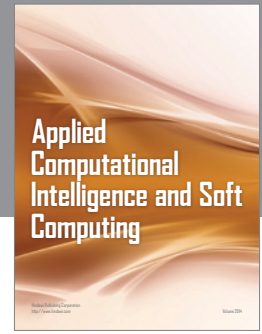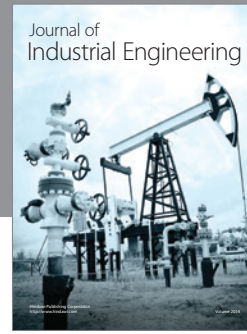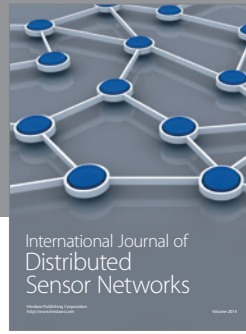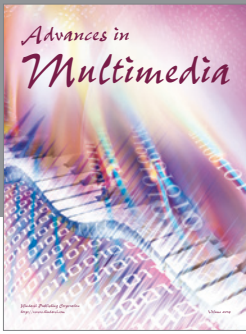
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] J. Nielsen, *Usability Engineering*, AP Professional, Boston, Mass, USA, 1993.

[2] J. A. Bargas-Avila and K. Hornbæk, "Old wine in new bottles or novel challenges? A critical analysis of empirical studies of User Experience," in *Proceedings of the 29th Annual CHI Conference on Human Factors in Computing Systems (CHI '11)*, pp. 2689–2698, ACM, New York, NY, USA, May 2011.

[3] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design*, John Wiley & Sons, Chichester, UK, 3rd edition, 2011.

[4] M. Hassenzahl and A. Monk, "The inference of perceived usability from beauty," *Human-Computer Interaction*, vol. 25, no. 3, pp. 235–260, 2010.

[5] M. Hassenzahl and N. Tractinsky, "User experience—a research agenda," *Behaviour and Information Technology*, vol. 25, no. 2, pp. 91–97, 2006.

[6] ISO 9241-210, *Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems*, ISO, Geneve, Switzerland, 2010.

[7] T. Lavie and N. Tractinsky, "Assessing dimensions of perceived visual aesthetics of web sites," *International Journal of Human Computer Studies*, vol. 60, no. 3, pp. 269–298, 2004.

[8] N. Tractinsky, A. S. Katz, and D. Ikar, "What is beautiful is usable," *Interacting with Computers*, vol. 13, no. 2, pp. 127–145, 2000.

[9] S. Diefenbach and M. Hassenzahl, "The "beauty dilemma": beauty is valued but discounted in product choice," in *Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems (CHI '09)*, pp. 1419–1426, Boston, Mass, USA, April 2009.

[10] A. N. Tuch, S. P. Roth, K. Hornbæk, K. Opwis, and J. A. Bargas-Avila, "Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI," *Computers in Human Behavior*, vol. 28, no. 5, pp. 1596–1607, 2012.

[11] M. Hassenzahl, "The interplay of beauty, goodness, and usability in interactive products," *Human-Computer Interaction*, vol. 19, no. 4, pp. 319–349, 2004.

[12] P. van Schaik and J. Ling, "Modelling user experience with web sites: usability, hedonic value, beauty and goodness," *Interacting with Computers*, vol. 20, no. 3, pp. 419–432, 2008.

[13] P. van Schaik, M. Hassenzahl, and J. Ling, "User-experience from an inference perspective," *ACM Transactions on Computer-Human Interaction*, vol. 19, no. 2, article 11, 2012.

[14] T. Ben-Bassat, J. Meyer, and N. Tractinsky, "Economic and subjective measures of the perceived value of aesthetics and usability," *ACM Transactions on Computer-Human Interaction*, vol. 13, no. 2, pp. 210–234, 2006.

[15] S. Lee and R. J. Koubek, "Understanding user preferences based on usability and aesthetics before and after actual use," *Interacting with Computers*, vol. 22, no. 6, pp. 530–543, 2010.

[16] A. Sonderegger and J. Sauer, "The influence of design aesthetics in usability testing: effects on user performance and perceived usability," *Applied Ergonomics*, vol. 41, no. 3, pp. 403–410, 2010.

[17] A. Sonderegger, G. Zbinden, A. Uebelbacher, and J. Sauer, "The influence of product aesthetics and usability over the course of time: a longitudinal field experiment," *Ergonomics*, vol. 55, no. 7, pp. 713–730, 2012.

[18] M. Thüring and S. Mahlke, "Usability, aesthetics and emotions in human-technology interaction," *International Journal of Psychology*, vol. 42, no. 4, pp. 253–264, 2007.

[19] M. Hassenzahl, M. Burmester, and F. Koller, "Attrak Diff: ein Fragebogen zur Messung wahrgenommrner hedonischer und pragmatischer Qualität," in *Mensch & Computer 2003: Interaktion in Bewegung*, J. Ziegler and G. Szwillus, Eds., pp. 187–196, Teubner, Stuttgart, Germany, 2003.

[20] M. McCurdy, C. Connors, G. Pyrzak, B. Kanefsky, and A. Vera, "Breaking the fidelity barrier: an examination of our current characterization of prototypes and an example of a mixed-fidelity success," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, pp. 1233–1242, Montreal, Canada, April 2006.

[21] M. Hassenzahl, "The effect of perceived hedonic quality on product appealingness," *International Journal of Human-Computer Interaction*, vol. 13, no. 4, pp. 481–499, 2001.

[22] T. Gilovich, D. Keltner, and R. E. Nisbett, *Social Psychology*, W. W. Norton & Company, New York, NY, USA, 2006.

[23] R. R. Hall, "Prototyping for usability of new technology," *International Journal of Human Computer Studies*, vol. 55, no. 4, pp. 485–501, 2001.

[24] N. P. Archer and Y. Yuan, "Comparing telephone-computer interface designs: are software simulations as good as hardware prototypes?" *International Journal of Human-Computer Studies*, vol. 42, no. 2, pp. 169–184, 1995.

[25] K. O. Götz and K. Götz, "Color preferences of art students: surface colors: II," *Perceptual and Motor Skills*, vol. 41, no. 1, pp. 271–278, 1975.

[26] J. Wandmacher, *Software Ergonomie*, De Gruyter, Berlin, Germany, 1993.

[27] ISO 9241-11, *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) Part 11: Guidance on Usability*, ISO, Geneve, Switzerland, 1998.

[28] T. Tullis and B. Albert, *Measuring the User Experience*, Morgan Kaufmann, Amsterdam, The Netherlands, 2008.

[29] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., pp. 139–183, North Holland Press, Amsterdam, The Netherlands, 1988.

[30] C. Pfendler, "Zur Messung der mentalen Beanspruchung mit dem NASA-Task Load Index," *Zeitschrift für Arbeitswissenschaft*, vol. 44, pp. 158–163, 1990.

[31] G. V. Glass and J. C. Stanley, *Statistical Methods in Education and Psychology*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1970.

[32] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, Pearson, Boston, Mass, USA, 2007.

[33] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, L. Erlbaum, Hillsdale, NJ, USA, 2nd edition, 1988.

[34] J. T. E. Richardson, "Eta squared and partial eta squared as measures of effect size in educational research," *Educational Research Review*, vol. 6, no. 2, pp. 135–147, 2011.

[35] K. McNeil, I. Newman, and F. J. Kelly, *Testing Research Hypotheses with the General Linear Model*, SIU Press, Illinois, Ill, USA, 1996.

[36] R. M. Baron and D. A. Kenny, "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," *Journal of Personality and Social Psychology*, vol. 51, no. 6, pp. 1173–1182, 1986.

[37] S. O. Hansson, "Aesthetic functionalism," *Contemporary Aesthetics*, vol. 3, 2005, http://hdl.handle.net/2027/spo.7523862.0003.008.

[38] E. Alvarado, "Form and function in the social perception and appreciation of web sites," *Florida Atlantic Comparative Studies Journal*, vol. 10, no. 1, pp. 151–170, 2007-2008.

[39] A. de Angeli, J. Hartmann, and A. Sutcliffe, "The effect of brand on the evaluation of websites," in *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II*, pp. 638–651, Springer, Uppsala, Sweden, 2009.

[40] J. Sauer and A. Sonderegger, "The influence of prototype fidelity and aesthetics of design in usability tests: effects on user behaviour, subjective evaluation and emotion," *Applied Ergonomics*, vol. 40, no. 4, pp. 670–677, 2009.

[41] Y.-K. Lim, A. Pangam, S. Periyasami, and S. Aneja, "Comparative analysis of high- and low-fidelity prototypes for more valid usability evaluations of mobile devices," in *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, pp. 291–300, Oslo, Norway, October 2006.