Visual Aesthetics and Performance: A First Meta-Analysis

Meinald T. Thielsch
Department of Psychology
University of Münster
Münster, Germany
thielsch@uni-muenster.de

Jana Scharfen
Department of Psychology
University of Münster
Münster, Germany
jana.scharfen@uni-muenster.de

Ehsan Masoudi
Department of Psychology
University of Münster
Münster, Germany
ehsan.masoudi@uni-muenster.de

Meike Reuter
Department of Psychology
University of Münster
Münster, Germany
meike.reuter@uni-muenster.de

ABSTRACT

Aesthetics has become a central construct in Human-Computer Interaction (HCI), and it has clear beneficial effects on users' perceptions and attitudes. Yet, do attractive interfaces actually enhance user performance? In the light of the importance of the construct and the mixed findings in various studies on the matter a systematic approach is urgently needed. Thus, the present meta-analysis examines in detail the effects of visual interface aesthetics of websites, software and other interactive systems on objective user performance. A systematic literature search yielded 25 eligible studies with 101 observed effects and a total of 3,025 participants. The conducted meta-analysis revealed a small significant overall effect of interface aesthetics on user performance (g = 0.12), while a high heterogeneity of effects was observed. As potential moderators, we tested the type of used interaction medium, task, goal orientation, measure of performance, measure of aesthetics, and aesthetics manipulation. None showed a significant moderating influence. Thus, aesthetics can be considered to have a small but heterogeneous influence on user performance that so far cannot be further resolved by moderating variables reported in eligible studies. Therefore, the discussion sketches avenues for future research and encloses a call to action for the HCI community.

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI) → Empirical studies in HCI

MuC '19, September 8–11, 2019, Hamburg, Germany
 ACM ISBN 978-1-4503-7198-8/19/09…$15.00
https://doi.org/10.1145/3340764.3340794

KEYWORDS

aesthetics, beauty, appeal, attractiveness, user performance

ACM Reference Format:

Meinald T. Thielsch, Jana Scharfen, Ehsan Masoudi and Meike Reuter. 2019. Visual Aesthetics and Performance: A First Meta-Analysis. In *Mensch und Computer 2019 (MuC'19), September 8—11, 2019, Hamburg, Germany.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3340764.3340794

1 Introduction

Before the launch of personal computing in the late 1970s, the only people who interacted with computers were professionals in the information technology industry. Today, computers of various types (laptops, tablets, smartphones etc.) have become an integral part of all private and professional life. Moreover, trends such as the "Internet of Things" and "Industry 4.0" indicate that the digitalization is far from finished. In light of this evolution, it comes as no surprise that the interest in human-computer interaction (HCI) and its improvement has continued to grow. In particular, as the quality of HCI at workplaces strongly influences employees' performance and thus the economic success of a company (e.g., [27], [47]), the topic is critically important, especially for organizations.

Early HCI research stressed that workplace computer systems were meant to enhance efficiency, not to give pleasure. Some authors even argued that aesthetic designs interfered with work goals (e.g., [26], [51]), and they objected to favoring aesthetics over usability (e.g., [69]). Continuing research has altered that perspective. In the past years, aesthetics has gained more and more attention, especially as it has been shown to positively affect users' subjective perceptions (for overviews, see [67] and [92]). In several studies, participants rated interactive systems (e.g., websites, browser-based applications, ATM software) that were aesthetically appealing as being more usable than aesthetically unappealing systems – even if these systems did not differ in functionality (e.g., [8], [29], [94]). Consequently, the following question seems to be integral to this vein of research: Do users only *feel* that they perform better when working with

an aesthetically pleasing system, or do they actually *perform* better? If the latter is the case, designing aesthetically pleasing user interfaces would be an opportunity to enhance employee's performance and thus increase the efficiency of an organization.

Even though research on aesthetics in HCI began in the mid-1990s, the number of studies directly investigating the effects of visual aesthetics on objective user performance is limited. In addition, the existing studies offer a wide range of findings: from positive effects to negative effects to even no effects at all (see [88]). As this topic is highly relevant for various applications, the contradictions are in urgent need of clarification. In order to achieve this, the following meta-analysis examines the effects of interface aesthetics on user performance. In this paper, the term *performance* refers to an objectively measurable outcome of a user's interplay with a website, software or other interactive system (e.g., task duration).

1.1 Definition of Aesthetics in HCI

In recent years aesthetics has gained attention and is nowadays a highly prevalent construct in the area of HCI (e.g., [50], [89], [92]). Since experimental aesthetics is one of the oldest subjects in psychological research, many scientists have tried to define this construct throughout history, starting in the late nineteenth century (e.g., [44], [99]). Today, aesthetics is known as a multi-dimensional construct [57], [59], [67], but it lacks a consistent definition.

The present paper explicitly targets aesthetics in the context of HCI, which is why Noam Tractinsky [91] must be named as one of the first researchers who framed this term. He defined aesthetics as "beauty" but specified the construct later by distinguishing between classical and expressive aesthetics [57]. Classical aesthetics includes attributes such as clean, pleasant and symmetrical, whereas expressive aesthetics is meant to be creative, original and sophisticated. Extending this framework, Moshagen and Thielsch [67] introduced four aesthetic dimensions: simplicity, diversity, colorfulness and craftsmanship. The first two dimensions highly correlate with Lavie and Tractinsky's [57] classical and expressive aesthetics. Based on prior research, Moshagen and Thielsch [67] specified aesthetics to be an "immediate pleasurable subjective experience that is directed toward an object and not mediated by intervening reasoning" ([67], p. 690). We draw on this interactionistic viewpoint for the remainder of the paper.

Despite these descriptions, HCI research uses some slightly divergent phrases to describe aesthetics, such as "visual appeal" or "attractiveness" (e.g., [63], [14]). Thus, for the present meta-analysis, we follow the approach of the systematic literature review performed by Thielsch and Niesenhaus [88] in that we use the term aesthetics when referring to the appeal, beauty, attractiveness, or pleasure a user experiences when visiting a website or using an interactive system. Aesthetics is part of the general user experience, which is defined in ISO 9241-210 [54] as "person's perceptions and responses that result from the use and/or anticipated use of a product, system or service".

1.2 Importance of Aesthetics in HCI

Although there is no firmly fixed definition, the relevance of aesthetics is well acknowledged in the HCI community (e.g., [50], [67], [92]). Thanks to promising research over the last two decades, aesthetics has developed from a "nice-to-have" into a "must-have" factor in HCI design [89]. In Chapter 19 of The Encyclopedia of Human-Computer Interaction, Tractinsky [92] illustrates that visual aesthetics is important to HCI from a psychological perspective by formulating three main aspects:

First, aesthetics in general has intrinsic value; it satisfies basic human needs and contributes to pleasure and well-being (e.g., [74], [81]). This also applies to aesthetics of interactive systems: subjects interacting with visually pleasing designs were shown to experience increased pleasure, engagement, trust, and satisfaction with the system and were more likely to recommend and reuse it (e.g., [22], [57], [62], [67], [73], [94]).

The second point is that aesthetics in HCI is a major way for people to express themselves. By modifying desktop backgrounds, lock screens, cell phone covers etc., users seek to express who they are and underline their individuality [49]. This stems from the process of self-extension, which posits that "we learn, define, and remind ourselves who we are by our possessions" ([28], p. 160). Especially because many interactive products (computers, smartphones etc.) are used in social contexts, it is plausible that people utilize the flexibility of technological design to differentiate from each other.

Third, aesthetic impressions play a decisive role in how users evaluate an interactive system. It sounds rather trivial that users prefer attractive systems or products over unattractive ones, that they are more satisfied while using them, and that they trust these systems more (see [67], for an overview), but – remarkably – users also perceive attractive systems to be more *usable* than unattractive ones, even if they do not differ in functionality (e.g., [8], [11], [29], [94]). Moreover, users process aesthetics very quickly: they respond to aesthetics immediately at first sight, and the evaluation process takes only just over half a second [32]. Thus, visual aesthetics is probably one of the first facets of an interactive system users perceive and, consequently, has a strong impact on first impressions (e.g., [63], [89], [93], [96]). According to the primacy of first impressions on attitudes – well known from social science research (e.g., [39]) – it is therefore not surprising that visual aesthetics can affect how users evaluate several other system attributes, such as perceived usability (e.g., [90], [92]).

In summary, aesthetic design represents a way to positively influence users' emotional as well as cognitive processes [71], [60]. With regard to HCI, an aesthetic system is therefore able to improve its users' experience, their appraisal and their attitudes towards it (e.g., [22], [48], [67]). Furthermore, aesthetics can

serve as a central distinguishing feature in crowded marketplaces (see [31]), especially if interactive products or websites have other comparable features, such as usability or content. In line with these findings, Norman [70] formulated the daring thesis that "attractive things work better." However, it remains to be shown if this is true. Does a good initial impression based on attractive design mean that users really work better with the system?

1.3 Current State of Research on Aesthetics and Performance

The present research situation is contradictory. While various studies postulate that interface aesthetics positively or at least partially positively affects user performance (e.g., [1], [3], [10], [13], [16]), several studies conducted in similar ways showed the opposite results, revealing that aesthetics lowers user performance (e.g., [15], [18], [25]). One explanation for the negative effects of aesthetics is the "prolongation of joyful experience" hypothesis [16], indicating that users interact longer with aesthetically appealing products to lengthen the pleasant experience (instead of giving priority to directly solving tasks), which leads to higher response times and thus can impair performance in terms of speed. The positive effects, by contrast, are often explained by one of the three following theories:

1. **Affect mediation:** Norman [70], [71] proposed that positive emotions are induced by aesthetic design, leading to a better working cognitive system and therefore better performance (the "positive affect mediation" hypothesis). According to this theory, aesthetics can improve performance, particularly in creative and problem-solving tasks. Even more, aesthetics may compensate for usability problems, as users who feel good about the system will overlook design flaws. Several researchers follow Norman's theory (e.g., [12], [14], [78]). In addition, emotional design has revealed positive effects in several studies in learning contexts (e.g., [72], [97]). Yet, the emotions evoked from aesthetic design might be relatively weak or short-lived (e.g., [2], [7]). Furthermore, until now positive effects of attractive interface in creative tasks (as expected by Normans theory) have not been found in experimental research [20].

2. **Attentional and cognitive effects:** Szabo and Kanuka [86] were among the first to argue that aesthetic interface designs simply require less cognitive effort and can be processed more easily, as less complex and more coherent designs are thought to allow users to detect visual objects faster. More specifically, they thought that good designs promote automatic processing, whereas bad designs provoke a less efficient manual processing. In addition, through the halo effect, good designs may drive users to perceive the content as being of higher quality, leading users to pay more attention to content that is presented with a good design [86]. Building on those ideas, other researchers discussed how aesthetics induces attentional effects (e.g., [79]) or other cognitive perception aspects such as mental models, bottom-up perception processes, visual complexity and prototypicality (e.g., [3], [95]). Tractinsky and colleagues [94] introduced the "what is beautiful is usable" viewpoint and discussed that aesthetics induced positive effects potentially based on halo effects, stereotype perception, and affective responses. Several studies agreed with this viewpoint, but others argued that aesthetic pleasure is the result not the cause of processing dynamics, and "what is beautiful is usable" may actually be reversed under certain conditions (for an overview, see [96]). Along similar lines, in the processing fluency theory, Reber and colleagues [76] assumed that the more fluently a person can process an object, the more positive his or her aesthetic response will be. Yet, to date, there is mixed evidence on the applicability of the processing fluency theory on aesthetics in HCI ([32], [66], [87]).

3. **Motivational effects:** The "increased motivation" hypothesis by Sonderegger and colleagues [16] states that aesthetically appealing technologies increase users' motivation and thus enhance their performance. Aesthetically appealing interfaces may put the user at ease (see [61]) or "in flow" (see [36]), a state in which they perceive their skills to be congruent with challenges in engrossing activities, leading to fast task completion. To the best of our knowledge, until now there has been no direct experimental investigation of this idea in the context of interface aesthetics.

In sum, systematic investigations of these theories are mostly lacking, and it remains unclear which model best explains the observed findings. Moreover, numerous studies have reported that aesthetics neither positively nor negatively affect user performance (e.g., [2], [6], [7], [11], [21], [29], [34], [52], [94]).
In order to gain an overview of the current state of research on aesthetic user experience and performance, Thielsch and Niesenhaus [88] conducted an initial literature review. They found 46 studies on the matter and grouped them in terms of research on learning performance (8 studies), task performance (23 studies), and performance in testing and under stress (15 studies). Based on reviewed literature, Thielsch and Niesenhaus [88] concluded that users benefit from pleasant and joyful interactions with technical systems, and their positive experience can enhance performance in Internet-based working contexts under certain circumstances. They suggest that, especially in learning contexts, good user experiences relate positively to learning outcomes ([88], p. 94). However, such a systematic review is not capable of deeper analyses that take into account, for example, varying sample sizes, moderating variables or bias factors such as publication bias.

1.4 Chosen Methodology and Research Question

Given the series of studies postulating that visual aesthetics affect user performance in various ways, a meta-analytical approach to assess these claims is urgently needed. By using statistical methods to summarize the results of different studies, a meta-analysis can provide more precise estimates of effects than individual studies [45]. The present paper is therefore an effort to more accurately understand the true effect of aesthetics on performance by combining all relevant study outcomes in one main effect. By providing clarification, the results of this meta-analysis are expected to be pioneering for future research as well as for practitioners, as the results of the meta-analysis will represent a reference for organizations to help them decide whether or not to invest in aesthetic design, particularly if they seek to increase user performance. The leading research question is:

To what extent does the aesthetic design of an interactive product, a software or a website affect the objective performance of its users?

As this is the first meta-analytical approach to explore the relationship between interface aesthetics on user performance, we aim to include all relevant studies. This leads to a range of papers examining the topic at hand but varying in their methodical implementation; for example, they vary in terms of aesthetics measurement, aesthetics manipulation or performance measurement. With regard to variables that potentially moderate the effect of aesthetics on performance, we further seek to answer six explorative research questions:

1. How does the interaction medium influence the effects of aesthetics on performance?

Browser-based applications ([2], [9]), company websites [3], search engines [7], online shops [5], [76], intranet sites [25], software [77], [80], ATM applications [94], mobile phones [14], [15], [16], [18], [46], portable digital audio players [22], and electronic map displays in driving simulators [58] are only some of the interactive systems that have been investigated in previous studies. So far, the moderating influence that type of interaction medium has on how aesthetics affects performance has only been discussed cursorily in research (e.g., [14], [11], [10]), which is why the present meta-analysis will examine this as a potential moderating variable. In doing so, type of interaction medium is grouped into three categories: web, software, and mobile applications.

2. How does the task influence the effects of aesthetics on performance?

Existing studies have applied a variety of tasks, in that participants have often been asked to use a specific feature of a given interface. For example, tasks have involved adding new entries into a virtual phonebook [29], sending a prepared text message to another phone user [15] or withdrawing cash at an ATM [94]. Other common tasks are search-and-find tasks (e.g., [8], [9], [14], [20]) and "free viewing" tasks, where participants first get a certain amount of time to look around the system, and afterwards they are required to complete a questionnaire (e.g., [1], [6], [25]). Having these different types of tasks in mind, we assume that the kind of task moderates the effect of aesthetics on performance.

3. How does the goal orientation influence the effects of aesthetics on performance?

Researchers distinguish between the different goals associated with learning performance and task performance ([42], [65], [82]): When users aim to achieve learning goals they are concentrated on enhancing their competence and knowledge, which is why they are expected to take their time to gain a deeper understanding, and making mistakes is an incremental part of learning. In contrast, when they are focused on achieving task performance goals, they need to find the best solution and, therefore, concentrate more on outcomes and favorable judgements. This resembles typical work situations; thus, we will refer to this type of goals as "working goals". Furthermore, some tasks involve quick action and competition, such that participants must focus on high scores and accuracy, thus resembling test situations. Additionally, some learning scenarios might also result in stressful test situations, thus changing the character of the learning goal and resultant user behavior. This is why in the present research we investigated three modes of goal orientation: learning goals, working goals, and testing goals. Thielsch and Niesenhaus [88] mention that aesthetics is presumed to enhance performance especially in learning scenarios because, in their systematic review, many studies embedded in learning settings reveal at least partial positive effects (e.g., [13], [85], [72]). However, Hall and Hanna [6] did not find any effects of aesthetics on performance in a learning scenario. To gain clarity about this issue, goal orientation will also be examined as a moderating variable.

4. How does the performance measure influence the effects of aesthetics on performance?

In the present research field, the operationalization of performance varies; it can be measured by task duration (e.g., [12], [34]), number of correct answers (e.g., [3], [25]), number of errors (e.g., [78], [80]), number of commands needed (e.g., [16], [18], [77]), amount of additional information needed to fulfil a task (e.g., [17], [19], [23]) as well as comprehension and transfer [72], [97]. For example, Sonderegger and colleagues [18] examined the effects of website aesthetics on user performance by taking two performance measures: the efficiency of task completion (number of clicks) and the task completion time (seconds). Participants working with the highly aesthetic website needed significantly fewer clicks than participants who interacted with the less aesthetic website. In contrast, the time it took for participants to complete the task was not significantly different between the aesthetics conditions. To further explore such inconsistencies, the type of performance measure will be analyzed as a potential moderator.

5. *How does the aesthetics measurement influence the effects of aesthetics on performance?*
There is a lack of a common standard in measuring aesthetics. Even though scientifically validated measures exist (e.g., scales of [57] or the VisAWI, [67]), research has relied heavily on unstandardized and/or holistic aesthetic scales, where such scales vary widely in item numbers (between a single item up to 14 items), in bipolar items used to measure aesthetics (e.g., attractive to unattractive, beautiful to ugly, eye-catching to plain etc.), and in their scale metrics (from five-point to ten-point Likert scales). These measuring instruments vary in their psychometric quality and thus in their ability to measure the construct in a valid way. Furthermore, some studies define the degree of aesthetics in experimental conditions based on the results of the applied aesthetics measurement. Taking into account the variety of aesthetics measurements, participants from different studies are likely to be confronted with different degrees of aesthetics and consequently may perform differently in different experiments. Thus, we assume that the way aesthetics is measured and accordingly defined in studies the experimental conditions might impact resulting relationships between aesthetics and performance.

6. *How does the aesthetics manipulation influence the effects of aesthetics on performance?*
As there is no universal definition of aesthetics, it is not surprising that the term aesthetics encompasses a broad range of different facets. In studies examining interface aesthetics as the independent variable and user performance as the dependent variable, aesthetics has been manipulated in various ways, such as color (e.g., [1], [12]), typography (e.g., [7], [9]), layout (e.g., [3], [8]), texture (e.g., [10], [22]), shape (e.g., [16], [96]) and complexity (e.g., [4], [58]). As, for example, colorfulness is found to be an essential characteristic of system interfaces (e.g., [37], [61]), changing the whole color scheme may produce different effects on performance than when changing only the background texture of a system. Thus, the type of aesthetics manipulation will be examined as a potential moderating variable as well.

2 Method

2.1 Literature Search

In order to identify all existing studies that provide results regarding the relationship between interface aesthetics and users' objective performance, a detailed literature search was conducted. Therefore, we developed the search term *("aesthetic*" OR "aesthetic* interface" OR "aesthetic* design" OR "visual aesthetic*") AND ("performance" OR "user performance" OR "user behavio*" OR "usability")* and used the network-based online libraries PsychINFO and ACM Digital Library (DL) to perform the literature search. The operator asterisk (*) means that different spellings were taken into account.

The online libraries revealed 1,446 studies spanning 1990 to July 2017. As a first step, all studies were screened by title and abstract, and we subsequently excluded those that clearly did not deal with the research question of this analysis. The remaining studies were then screened in full, which left 13 studies. In addition, studies reviewed by Thielsch and Niesenhaus [88] were assessed in detail for eligibility. After removing duplicates between Thielsch and Niesenhaus [88] and the database search, nine further studies were identified as eligible. In addition, three studies were included from unpublished work requests (sent out to all corresponding authors of the so-far identified studies). In total, 25 studies were included in this meta-analysis.

2.2 Study Selection

Studies included in the meta-analysis had to fulfil a number of criteria. First, they needed to be written in English or German and to report quantitative data. Second, the studies had to examine perceived interface aesthetics of interactive systems such as websites, software or mobile products. Therefore, the study participants must have rated or assessed (either in a preliminary study or in the original study) the aesthetics of user interfaces. If the queried construct did not correspond to our working definition of aesthetics or if the interface aesthetics was estimated by only theory-based design concepts (e.g., particular screen ratios or colors set as highly aesthetic vs. not very aesthetic with no regard to users' actual perceptions), the study was excluded. The same applied for study material: If the participants were not confronted with an interactive user interface, but for example with abstract, unattached graphics, the study was excluded.

As this paper aims to investigate whether and in what way aesthetics affects users' objective performance, to be included in the meta-analysis, studies needed to collect objectively measured performance data. If participants were only asked about their subjectively perceived usability, the study was excluded. To ensure that differences in performance were exclusively caused by different aesthetic designs, studies also had to control for influencing factors such as functionality. If they did not, the study was excluded.

Moreover, studies needed to report enough data to calculate the relevant effect sizes. If studies were missing essential data, the authors were contacted via e-mail and asked to share the missing information. In cases of unavailable information or no response, the studies were excluded. To minimalize a possible publication bias, we included published as well as unpublished work. We also integrated studies regardless of their sample size, as we weighted them in the subsequent analysis according to their number of participants. The search and selection process is depicted in Figure 1.

2.3 Coding Procedure

In accordance with the defined research questions, the following study characteristics were coded: type of interaction medium (mobile, software, website), task (free viewing, search task, use function), goal orientation (learning, testing, working), performance measure (accuracy, amount of additional information needed, completion time), aesthetics measure (self-constructed, scale from [57], VisAWI [67]) and aesthetics manipulation (color, typography, layout, texture, shape, complexity). Additionally, study information (publication date and country) as well as sample information (age and gender distribution) were coded.

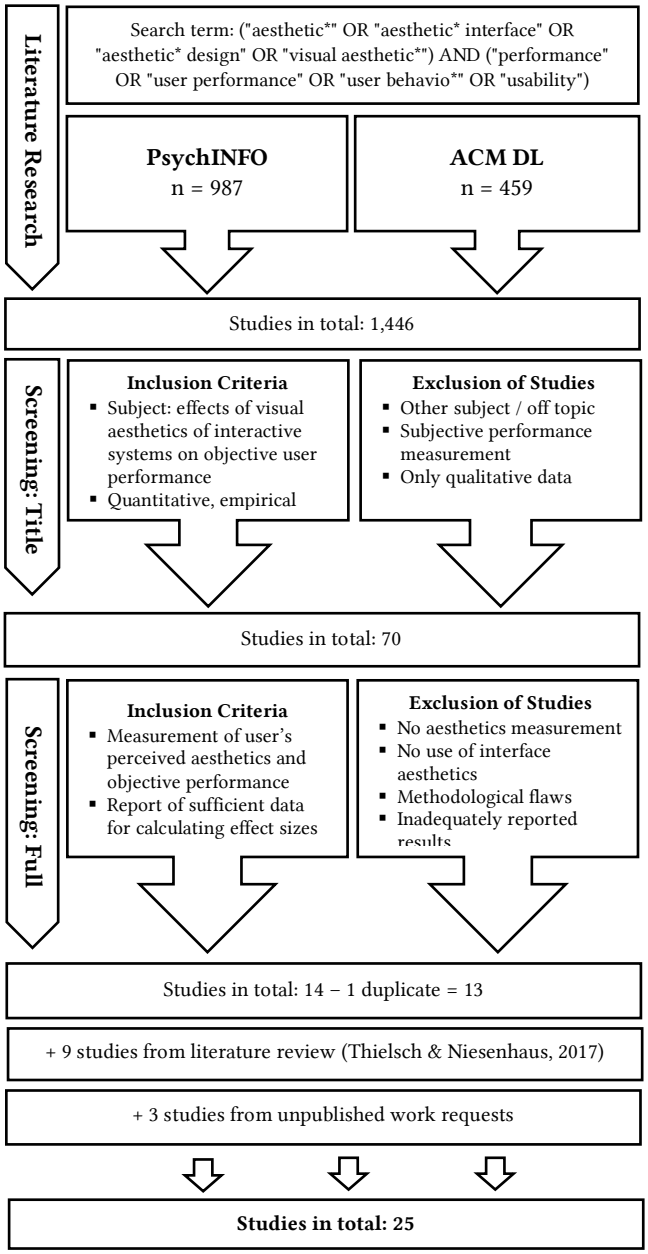

Figure 1: Flow chart of study selection process

2.4 Extraction of Effect Sizes

Participants using a low aesthetic user interface and those using a high aesthetic user interface were compared in their performance outcomes. The standardized mean differences (Cohen's *d*) between these two groups were calculated either based on means (*M*) and standard deviations (*SD*) or based on correlations (*r*), depending on the reported data of each study (*m* = 20 studies reported *M* and *SD*, *m* = 5 studies reported *r*). The pooled standard deviation was used for standardization of the mean differences in determining Cohen's *d*. The formula from Borenstein, Hedges, Higgins, and Rothstein ([33], p. 48) was used in order to transform *r* into *d*.

As Cohen's d tends to overestimate effect sizes for small samples, we applied Hedges' factor J to correct for the bias: $J = 1 - \frac{3}{4df-1}$ (where df for two independent groups is $n_1 + n_2 - 2$). Consequently, this meta-analysis used the corrected standardized mean difference Hedges' g ($g = d \times J$) and its corrected sampling variance ($\mathrm{Var}(g) = \mathrm{Var}(d) \times J^2$). Following the conventions by Cohen [35], g = 0.20 is interpreted as a small effect, g = 0.50 as a medium effect and g = 0.80 as a large effect.

2.5 Statistical Analyses

The meta-analysis was conducted using the package *metafor* ([98]; Version 2.0.0) in the program *R* ([75], Version 3.4.1). As most included studies reported more than one outcome variable, we had to control for this dependency between outcomes from the same study. We used a multilevel model to control for dependencies between the study outcomes. Two-sided *p*-values and an alpha level of alpha = .05 was applied for all analyses.

To examine potential effects of the defined moderator and control variables, moderator analyses were conducted by including them into the model. For each moderator, differences between subgroups were tested via linear hypotheses. As there are no clear empirical nor theoretical assumptions apparent in which direction these variables could moderate the effects of aesthetics on user performance, explorative moderator analyses were performed with non-directional hypotheses. Associations (Cramer's *V*) between moderators were analyzed to identify possible confounding between moderators.

Studies reporting statistically significant results are more likely to be published than studies without statistically significant results [38]. It can be assumed that studies using small sample sizes and yielding no or only small effects are the least likely to be published. This bias might mean that the meta-analysis would contain an unrepresentative sample of the actually existing studies, such that the results of the meta-analysis would show an over- or underestimated effect size and thus limited overall validity [84]. To examine the presence and degree of a possible publication bias, we inspected funnel plots for asymmetry. In funnel plots, effect sizes are plotted against their standard errors. These plots can hint at publication bias if studies with smaller sample sizes and thus higher standard errors show larger effects

more often than small effects. In the case of severe asymmetry, this might indicate that publication bias is an issue. Further numerical methods, such as Egger regression tests [41] or the trim and fill method [40] have not yet been developed for multilevel meta-analyses.

3 Results

This meta-analysis includes 25 eligible studies and 101 outcomes. Overall, the number of subjects was N = 3,025 with a mean age of 30.5 (SD = 7.7). 55.5% of the participants were female.

3.1 Overall Effects

The multilevel meta-analysis (k = 25 studies, o = 101 observations) showed a significant overall effect size of g = 0.12 with a 95% confidence interval (CI) of [0.03, 0.21] and a p-value < .01. According to Cohen's conventions, this effect size indicates a small, positive effect of interface aesthetics on user performance. The standard deviation of the true effect size was estimated as τ = 0.20, which indicates a high heterogeneity of results.

3.2 Moderator Analyses

Detailed results from moderator analyses can be found in Table 1. We conducted several explorative moderator analyses to examine whether the following possible moderators had an influence on the size of the effect of aesthetics on user performance: type of interaction medium (mobile, software, website), task (free viewing, search task, use function), goal orientation (learning, testing, working), performance measure (accuracy, amount of additional information needed, time), aesthetics measure (unstandardized and holistic measures, VisAWI [67], scale by Lavie and Tractinsky [57]) and aesthetics manipulation (color, typography, layout, texture, shape, complexity). No significant group differences between the specific subgroups of a moderator were found (p > .05), which means that the size of the effect is not significantly determined by these factors. Hence, the moderators investigated in this meta-analysis cannot explain the high heterogeneity of the overall effect.

Interestingly, some of the subgroups within the moderating variables did show effects that were significantly larger than zero, whereas others did not. Still, they did not differ significantly. Specifically, aesthetics seems to have a significant influence on user performance when users interact with mobile devices (g = 0.21, p = .028) and software applications (g = 0.35, p = .019); on websites, the effect is positive but not significantly different from zero (g = 0.06, p = .271). Yet, the effect sizes of these three subgroups do not differ significantly between each other. This becomes apparent when inspecting the p-values comparing subgroups' effects with each other. Thus, we cannot conclude that aesthetics plays a bigger role in the use of software and mobile devices than it does for websites. Yet, one conclusion we can draw is that for mobile devices and software, aesthetics can be considered to have a significant influence on performance, whereas on websites, it might not. Similar patterns were observed for task, goal orientation, performance measures, and aesthetics measures: Aesthetics might influence user performance in free viewing tasks (g = 0.10, p = .042), in a working goal context (g = 0.13, p = .015) and when accuracy is used as a performance measure (g = 0.17, p = .001). Furthermore, the application of unstandardized and more holistic measures[1] is potentially associated with larger observed effects (g = 0.19, p = .006).

Year and country of publication as well as gender and age distribution were further analyzed as control variables, but none of them had a significant influence of the size of the effect (p > .05). The associations between moderators emerged as low to moderate Cramer's Vs.

3.3 Publication Bias

The funnel plot (see Figure 2) is slightly asymmetric. Studies with small sample sizes and thus low standard errors show large effects more often than they show small effects. This is might indicate that small studies with small effects might have not been published, as larger studies did find smaller effects. Publication bias might thus be an issue for the current meta-analysis, and the effects reported above might be overestimating the influence of aesthetics on user performance. Unfortunately, given that no methods have yet been developed that can generate sizable estimates of this overestimation, we cannot give an exact effect size corrected for publication bias.

When inspecting the funnel plot in detail, it becomes apparent that only four effect sizes, of which three stem from the study by Pomales-Garcia, Liu, and Mendes [13], lead to the observed asymmetry. As only a few outcomes from two studies lead to the asymmetry and no overall pattern can be detected, a publication bias might thus not be too critical. Nonetheless, a slight overestimation of the found positive effect of aesthetics on performance is possible.

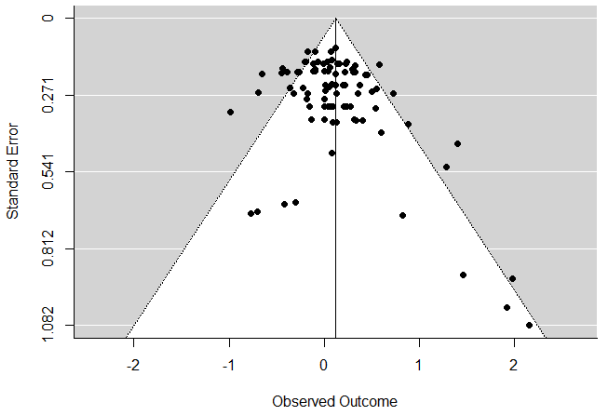

Figure 2: Funnel plot of effect sizes (observed outcomes) plotted against their standard errors

[1] Of the 14 studies in this category, eight studies used single and thus more holistic items (directly asking for beauty or attractiveness), three studies relied on expert ratings, and an additional three studies relied on short aesthetic scales with three to seven items.

Table 1: Results from moderator analyses

	k	*o*	*N*	*g*	*SE*	95% CI	*p*	Group differences
Interaction medium								
mobile$_1$	6	22	1320	0.21	0.10	[0.02, 0.40]	0.028	1 vs. 2: $p = 0.434$
software$_2$	4	13	606	0.35	0.15	[0.06, 0.64]	0.019	2 vs. 3: $p = 0.071$
website$_3$	15	66	10474	0.06	0.06	[-0.05, 0.17]	0.271	1 vs. 3: $p = 0.180$
Task								
free viewing$_1$	7	35	3473	0.19	0.10	[0.01, 0.38]	0.042	1 vs. 2: $p = 0.278$
search task$_2$	8	34	7093	0.06	0.08	[-0.09, 0.21]	0.422	2 vs. 3: $p = 0.390$
use function$_3$	10	32	1834	0.16	0.08	[-0.01, 0.32]	0.057	1 vs. 3: $p = 0.778$
Goal orientation								
learning$_1$	6	14	2381	0.11	0.09	[-0.07, 0.29]	0.255	1 vs. 2: $p = 0.973$
testing$_2$	4	25	2761	0.11	0.12	[-0.11, 0.38]	0.328	2 vs. 3: $p = 0.883$
working$_3$	17	62	7258	0.13	0.05	[0.03, 0.24]	0.015	1 vs. 3: $p = 0.802$
Performance. measure								
accuracy$_1$	22	60	7512	0.17	0.05	[0.07, 0.28]	0.001	1 vs. 2: $p = 0.249$
information$_2$	5	11	556	0.01	0.13	[-0.24, 0.27]	0.911	2 vs. 3: $p = 0.769$
time$_3$	15	30	4332	0.06	0.06	[-0.07, 0.18]	0.389	1 vs. 3: $p = 0.079$
Aesthetics measure								
unstandardized/holistic$_1$	14	39	3683	0.19	0.07	[0.05, 0.33]	0.006	1 vs. 2: $p = 0.213$
Lavie & Tractinsky [57]$_2$	6	39	3175	0.06	0.09	[-0.11, 0.22]	0.509	2 vs. 3: $p = 0.807$
VisAWI [67]$_3$	5	23	5542	0.09	0.09	[-0.09, 0.27]	0.346	1 vs. 3: $p = 0.356$
Aesthetics manipulation								
color$_1$	21	82	11528	0.15	0.10	[-0.04, 0.33]	0.131	1 vs. 2: $p = 0.828$
typography$_2$	4	26	4407	0.19	0.20	[-0.19, 0.57]	0.326	1 vs. 3: $p = 0.916$
layout$_3$	9	40	7316	0.13	0.09	[-0.05, 0.31]	0.157	1 vs. 4: $p = 0.677$
texture$_4$	4	15	858	0.05	0.21	[-0.37, 0.47]	0.824	1 vs. 5: $p = 0.616$
shape$_5$	5	19	1054	0.07	0.11	[-0.14, 0.29]	0.515	1 vs. 6: $p = 0.336$
complexity$_6$	1	3	888	-0.07	0.20	[-0.46, 0.32]	0.732	2 vs. 3: $p = 0.777$
								2 vs. 4: $p = 0.617$
								2 vs. 5: $p = 0.591$
								2 vs. 6: $p = 0.352$
								3 vs. 4: $p = 0.720$
								3 vs. 5: $p = 0.681$
								3 vs. 6: $p = 0.366$
								4 vs. 5: $p = 0.920$
								4 vs. 6: $p = 0.692$
								5 vs. 6: $p = 0.539$

Note. The performance measure "information" refers to the amount of additional information needed to fulfill a task; *k* = number of effect sizes for observed outcomes (extracted from a total of 25 included studies); *o* = amount of observations; *N* = total sample size; *g* = corrected standardized mean difference Hedges' g; *SE* = standard error; CI = confidence interval.

4 Discussion

The purpose of this meta-analysis was to investigate the relationship between interface aesthetics of interactive systems and the objective performance of users. We therefore conducted a systematic literature review yielding 25 studies (including 101 observed outcomes) that fulfilled the defined inclusion criteria. The calculated meta-analysis revealed a small effect size of g = 0.12 indicating a slightly positive impact of aesthetics on performance according to the definition by Cohen [35]. Yet, one has to keep in mind the high heterogeneity of enclosed studies as well as a slight potential publication bias. Still, considering the huge amount of interface users worldwide, even small positive effects can lead to vast consequences. Thus, further investigations on the performance-enhancing capability of aesthetics and the mechanisms driving it will be invaluable.

To gain deeper insight into the found effect, we took into account six important potential moderatos: Type of interaction medium, task, goal orientation, performance measure, aesthetics measure, and aesthetics manipulation. Yet, none of the moderator variables significantly influenced the found effects of aesthetics on performance. When interpreting these findings,

one should keep in mind that only few studies were available for the built categories in the moderator analysis. Moreover, some of the subgroups within a moderating variable did show significant effects (see section 3.2). For example, considerable effects of aesthetics were found for mobile and software applications as well as in working goal contexts. Furthermore, aesthetics might have bigger influence in free viewing than in search tasks. In addition, when investigating the positive effects of aesthetics, researchers should pay special attention to accuracy measures when operationalizing performance.

Importantly, moderators other than the ones investigated might be at work: A study by Moshagen and colleagues [12] – also included in this meta-analysis – found that aesthetics only enhanced performance under conditions of poor usability. When usability was held constant at a high level, attractive design did not lead to increased user performance [12]. These findings suggest a possible interaction between aesthetics and usability. While the level of usability (high vs. poor) could have been included in this meta-analysis as a potential moderator, only this one study [12] explicitly took usability two-tiered into account while varying aesthetics at the same time. Thus, coding the level of usability for each study included in the meta-analysis was not possible. As a result, according to the findings of Moshagen and colleagues [12] it should be further investigated whether the effects of aesthetics on user performance depend on different levels of usability.

4.1 Limitations

To our knowledge, the present paper is the first meta-analytical approach that explores the relationship between interface aesthetics and user performance. As such, some limitations must be mentioned and should be taken into account when interpreting the results.

One of the major challenges we faced was bringing the previous research to a comparable level and combining all aspects in one usable coding system. As HCI is a multidisciplinary research field combining different areas, such as psychology, computer science, web design etc., we had to deal with diverse and unstandardized study reports. Incomplete or totally missing information regarding study design, sample, methodical procedure and study results occurred frequently. Most of the requests we made to study authors were not addressed: either they did not answer our emails, or they admitted to not properly storing the data (and thus not being able to provide missing data). We therefore had to exclude several studies that actually did examine the relevant issue. As such, we consider our present work as a first-attempt meta-analysis on the topic, in hopes that forthcoming research studies are better documented such that future meta-analyses will be able to include more primary data and analyze moderators in more detail.

A second limitation concerns our independent variable: aesthetics. Even though it is one of the core constructs in the area of HCI (e.g., [50], [67], [92]), the term aesthetics does not carry a standard definition as, for example, the terms content or usability [53], [54]. Therefore, we aimed to integrate an aesthetics definition in our coding system in hopes of categorizing different understandings of the construct. In the end, this was not feasible as most papers did not give any precise definition. When a study's understanding of aesthetics obviously differed from our definition and inclusion criteria (aesthetics in terms of "appeal, beauty, attractiveness or pleasure experienced with a website or an interactive system", Thielsch and Niesenhaus [88], we excluded the paper. However, in a more general sense, all studies that did not report a concrete description of aesthetics run the hypothetical risk of not coinciding with our understanding of the construct.

Another problem, which is common in the HCI research community, is the issue of small sample sizes. Examining the included studies more closely, we can find two studies with very small sample sizes (n = 8; n = 10) but high effect sizes (g = 1.5 - 2.2, [13]; g = -0.77 – 2.0, [19]). Even though smaller N's are statistically recognized in the proceeding of meta-analyses, it is possible that effects of aesthetics on performance were not systematically observed because studies with small sample sizes lack statistical power. To avoid this issue, HCI researches should apply tools such as G*Power [43]. In doing so, they can determine sample requirements of a planned study beforehand, ensuring they conduct it with sufficient statistical power. This would allow true effects to be better determined.

4.2 Recommendations for Future Research

First, we would like to make an appeal to the HCI research community. As stated in the limitations section, many studies could not be integrated in this meta-analysis because of deficient report quality. Due to this, research that actually did examine the effects of interface aesthetics on user performance and therefore could have influenced the overall effect size given here had to be excluded. By not reporting all relevant data, researchers limited not only their own work's reach but also the validity of the present and future interpretations of their findings. To avoid this issue and establish a scientifically clean and stable basis for prospective meta-analyses, HCI studies on aesthetics should always report the following items:

- total number of participants (if a between-subject design is used, also the number of subjects in each group),
- gender and age distribution,
- manipulated facets of aesthetics (e.g., color, typography, texture etc.) and in what way these facets were manipulated,
- measurement method of aesthetics as well as its reliability and validity,
- aesthetics ratings of the non-aesthetic and the aesthetic experimental condition (means and standard deviations),
- method of measuring performance as well as its reliability and validity,
- achieved performance scores in the non-aesthetic and the aesthetic experimental condition (means and standard deviations).

In addition, future research should agree on a scientifically substantiated definition of aesthetics acknowledged in the HCI research community. This would create a solid basis for prospective studies with aesthetics as a dependent variable. Furthermore, valid testing measures for interface aesthetics need to be applied. So far, few scientifically validated instruments are available: The scales of Lavie and Tractinsky [16]; the VisAWI (Visual Aesthetics of Websites Inventory, [67]) and its short version VisAWI-S [68]; and the recently published Aesthetic Pleasure in Design Scale (which has been, so far, only partly validated in an HCI context, see [30]). The majority (14 of 25) of the studies included in the meta-analysis used unstandardized, often non-validated scales that varied extremely in number and content of items as well as in their scale metrics. If and in what way these scales represent useful measurement methods remains unanswered. By consequence, such unstandardized scales set the degree of aesthetics in the experimental conditions at their own discretion. Usually, the design rated as most appealing was applied as the aesthetic condition, and the design rated as least appealing was used as the unaesthetic condition. But how aesthetic or unaesthetic these designs actually were and how much they really differed from each other is untraceable. The moderator analysis points to the fact that those unstandardized (and often more holistic) measures were associated with a larger found effect of aesthetics on performance. Based on the existing data, it is hard to conclude if this is caused by a more comprehensive (compared to existing standardized instruments) but valid measurement of aesthetics or a lack of psychometric quality of such measures and thus a measurement error. Researcher should treat this issue with care; for example, they could apply standardized and holistic measures within one study aiming for an optimal measurement of the construct.

Furthermore, future studies should particularly pay attention on two aspects: First, research should address the understanding of causing mechanisms. Several theories have been proposed as to how aesthetics alters user performance (see section 1.3). Future research should use existing theories to direct experimental and comparative tests. Understanding the causing mechanism will be a tremendous step in advancing aesthetics research and in creating a solid base for design decisions when creating apps, software, and websites. Second, future research should aim for high practical relevance. While many previous studies have measured performance in abstract testing environments without practical connections, future studies might ideally be designed to represent more typical work and real-life situations. This could be done by using realistic tasks and performance measures as well as realistic software / products and aesthetics manipulations.

On this basis, future research should also invest in longitudinal studies. So far, most of the existing papers represent cross-sectional studies examining only short-term effects of aesthetics on performance. This approach might be reasonable when only taking into account websites, which are usually only visited for a short time (e.g., because they provide specific information about a specific topic that users look up only once and then leave the website). But when we look at interactive systems such as computer software or programs that users have to work with day after day, the results of short-term studies might not be transferable. This issue is particularly important, as some recent studies have found alterations or vanishing effects of aesthetics over time [11], [56], [83].

4.3 Practical Implications

The present meta-analysis revealed three main implications for practice: First, there is no overall negative effect of interface aesthetics on user performance as feared by early HCI research (e.g., [26], [51]). Thus, aesthetic designs in general do not interfere with work goals nor harm user performance.

Second, in contrast, aesthetics could be applied as a (minor) performance booster when practitioners aim to improve efficiency. The effect size might be small, but it is still of high relevance if many users are affected. Depending on the scale of application, even small effects can lead to huge economic outcomes. If we think of the enormous number of office software users, even a minor increase in aesthetics might lead to a substantial improvement in the total productivity. Yet, practitioners should have realistic expectations for what aesthetics can add to performance – the ergonomics of work-related applications is crucial (e.g., [47], [64]).

Third, even as the moderator analysis was not able to explain the overall effect, it revealed that it might be beneficial to give special attention to aesthetics when mobile and software applications are concerned, especially in a working goal context when accuracy is important. In such situations, investments in aesthetics may pay off quickly.

Finally, the general positive impact of aesthetics on subjective outcomes should be kept in mind: higher user preference, satisfaction, trust, and willingness to reuse and recommend (see [67], p. 691, [92]). This means that if an organization plans to invest in a new interface anyway – whether for employees or clients – it is very much worth it to consider aesthetics.

4.4 Conclusion

According to the outcomes of this meta-analysis, visual aesthetics of websites, software and other interactive systems has a small, positive effect on objective user performance. Consequently, our meta-analysis backs the claim by Norman [70]: Attractive devices work better. Still, further research on the matter is necessary to investigate potential moderators. More and higher quality publications in terms of methodical procedures and reporting quality are in urgent need to create the foundations for reliable guideline works such as meta-analyses. Overall, the present meta-analysis should be considered as a first attempt to provide clarification and indicate a rough direction for future work. As such, we hope that this meta-analysis gives the HCI community an incentive to initiate further, differentiated research regarding effects of interface aesthetics on user performance. Nonetheless, the present state of research according to our meta-analysis is that users not only feel better if they use aesthetically pleasant interfaces, but, on average, they also perform a little better as well.

REFERENCES

*References marked with an asterisk indicate studies included in the meta-analysis.

[1] *N. Bonnardel, A. Piolat, & L. Le Bigot (2011). The Impact of Colour on Website Appeal and Users' Cognitive Processes. Displays, 32(2), 69 – 80.

[2] *M. Douneva, R. Haines, & M. T. Thielsch (2015). Effects of Interface Aesthetics on Team Performance in a Virtual Task. In *Proceedings of the 23rd European Conference on Information Systems* (pp. 1 – 10).

[3] *M. Douneva, R. Jaron, & M. T. Thielsch (2016). Effects of Different Website Designs on First Impressions, Aesthetic Judgments, and Memory Performance after Short Presentation. Interacting with Computers, 28, 552 – 567. doi:10.1093/iwc/iwv033

[4] *S. Gersie (2010). *Wie beeinflussen Ästhetik und Stimmung die Leistung bei der Bearbeitung von Suchaufgaben in einer webbasierten Umgebung?* (Diploma thesis). Westfälische Wilhelms-Universität Müster.

[5] *H. Gu, W. Hou, X. Qin, L. Zhang, & Y. Dai (2016). The Effects of Aesthetics in Usability Testing for B2C E-commerce Websites. In *Proceedings of the Fourth International Symposium on Chinese CHI* (p. 2). ACM.

[6] *R. H. Hall, & P. Hanna (2004). The Impact of Web Page Text-Background Colour Combinations on Readability, Retention, Aesthetics and Behavioural Intention. Behaviour & Information Technology, 23(3), 183–195.

[7] *A. Katz (2010). Aesthetics, Usefulness and Performance in User--Search-Engine Interaction. Journal of Applied Quantitative Methods, 5(3), 424–445.

[8] *S. Lee, & R. J. Koubek (2010). Understanding User Preferences Based on Usability and Aesthetics Before and After Actual Use. Interacting with Computers, *22*(6), 530 – 543. doi:10.1016/j.intcom.2010.05.002

[9] *J. Ling, & P. Van Schaik (2006). The influence of font type and line length on visual search and information retrieval in web pages. International Journal of Human-Computer Studies, 64(5), 395–404.

[10] *C. Miller (2011). Aesthetics and E-Assessment: the Interplay of Emotional Design and Learner Performance. Distance Education, 32(3), 307–337.

[11] *M. Minge & M. Thüring (2018) Hedonic and Pragmatic Halo Effects at Early Stages of User Experience. International Journal of Human-Computer Studies, 109, 13-25. doi:10.1016/j.ijhcs.2017.07.007

[12] *M. Moshagen, J. Musch, & A. S. Göritz (2009). A Blessing, not a Curse: Experimental Evidence for the Beneficial Effects of Visual Aesthetics on Performance. Ergonomics, 52(10), 1311–1320. doi:10.1080/00140130903061717

[13] *C. Pomales-García, Y. Liu, & D. Mendez (2005). Web-Based Distance Learning Technology: Does Appearance Matter?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(7), 755–758. Sage CA: Los Angeles, CA: SAGE Publications.

[14] *J. M. Quinn, & T. Q. Tran (2010, April). Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 353–362). ACM.

[15] *J. Sauer & A. Sonderegger, A. (2011). The Influence of Product Aesthetics and User State in Usability Testing. Behaviour & Information Technology, 30(6), 787–796. doi:10.1080/0144929X.2010.503352

[16] *A. Sonderegger & J. Sauer (2010). The Influence of Design Aesthetics in Usability Testing: Effects on User Performance and Perceived Usability. Applied Ergonomics, 41(3), 403 – 410.

[17] *A. Sonderegger, J. Sauer, & J. Eichenberger (2014). Expressive and classical aesthetics: two distinct concepts with highly similar effect patterns in user–artefact interaction. Behaviour & Information Technology, 33(11), 1180–1191.

[18] *A. Sonderegger, A. Uebelbacher, M. Pugliese, & J. Sauer (2014). The influence of aesthetics in usability testing: the case of dual-domain products. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 21–30). ACM.

[19] *M. Stojmenovic, C. Pilgrim, & G. Lindgaard (2014). Perceived and objective usability and visual appeal in a website domain with a less developed mental model. In *Proceedings of the 26th Australian computer-human interaction conference on designing futures: the future of design* (pp. 316–323). ACM.

[20] *M. T. Thielsch, R. Haines, & L. Flacke (2019). Experimental investigation on the effects of interface aesthetics on user performance in different virtual tasks. PeerJ, 6:e5483. doi:10.7717/peerj.5483

[21] *M. T. Thielsch, D. Nikolaeva, & G. Kebeck (2009). "Beeinflusst Ästhetik die Performanz bei Suchaufgaben im Web?" In H. Wandke, S. Kain, and D. Stroove (Eds), *Mensch & Computer 2009*. München: Oldenbourg, pp. 487–490.

[22] *M. Thüring, & S. Mahlke (2007). Usability, Aesthetics and Emotions in Human–Technology Interaction. International Journal of Psychology, 42(4), 253–264.

[23] *A. N. Tuch, S. P. Roth, K. Hornbæ, K. Opwis, & J. A. Bargas-Avila (2012). Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. Computers in Human Behavior, 28(5), 1596–1607. doi:10.1016/j.chb.2012.03.024

[24] *E. R. Um, H. Song, & J. Plass (2007). The effect of positive emotions on multimedia learning. In EdMedia: *World Conference on Educational Media and Technology* (pp. 4176-4185). Association for the Advancement of Computing in Education (AACE).

[25] *P. Van Schaik, & J. Ling (2009). The role of context in perceptions of the aesthetics of web pages over time. International Journal of Human-Computer Studies, 67(1), 79–89.

[26] A. D. Andre, & C. D. Wickens (1995). When users want what's not best for them. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *3*(4), 10–14.

[27] S. Aral, E. Brynjolfsson, & M. Van Alstyne (2012). Information, technology, and information worker Productivity. Information Systems Research, 23(3 PART 2), 849–867. doi:10.1287/isre.1110.0408

[28] R.W. Belk (1988). Possessions and the extended self. Journal of consumer research, 15(2), 139–168.

[29] T. Ben-Bassat, J. Meyer, & N. Tractinsky (2006). Economic and subjective measures of the perceived value of aesthetics and usability. ACM Transactions on Computer-Human Interaction, 13(2), 210–234. doi:10.1145/1165734.1165737

[30] J. Blijlevens, C. Thurgood, P. Hekkert, L.-L. Chen, H. Leder, & T. W. A. Whitfield (2017). The Aesthetic Pleasure in Design Scale: The development of a scale to measure aesthetic pleasure for designed artifacts. Psychology of Aesthetics, Creativity, and the Arts, 11(1), 86–98. doi:10.1037/aca0000098

[31] P. H. Bloch, F. F. Brunel, & T. J. Arnold (2003). Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement . Journal of Consumer Research, 29(4), 551–565.

[32] J. Bölte, T. Hösker, G. Hirschfeld, & M. T. Thielsch (2017). Electrophysiological correlates of aesthetic processing of webpages: A comparison of experts and laypersons. PeerJ, 5, e3440. doi:10.7717/peerj.3440.

[33] M. Borenstein, L. V. Hedges, J. P. T. Higgins, & H. R. Rothstein (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd.

[34] B. Chawda, B. Craft, P. Cairns, D. Heesch, & S. Rüger (2005). Do "Attractive Things Work Better"? An Exploration of Search Tool Visualisations. In L. McKinnon, O. Bertelsen, & N. Bryan-Kinns (Eds.), *Proceedings of '05 British HCI Group Annual Conference* (pp. 46–49). Swindon, UK: British HCI Group.

[35] J. Cohen (1988). Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, N.J.: Erlbaum.

[36] M. Csikszentmihalyi (1977). *Beyond Boredom and Anxiety*. Jossey-Bass.

[37] D. Cyr, M. Head, & H. Larios (2010). Colour appeal in website design within and across cultures: A multi-method evaluation. International Journal of Human-Computer Studies, 68(1-2), 1–21.

[38] K. Dickersin, Y. L. Min, & C. L. Meinert (1992). Factors influencing publication of research results: Follow-up of applications submitted to two institutional review boards. Jama 267(3), 374–378.

[39] K. Dion, E. Berscheid, E. Walster (1972). What is beautiful is good. Journal of Personality and Social Psychology 24(3), 285–290.

[40] S. Duval, & R. Tweedie (2000). Trim and fill: a simple funnel-plot – based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56(2), 455–463. doi:10.1111/j.0006-341X.2000.00455.x

[41] M. Egger, G. D. Smith, M. Schneider & C. Minder (1997). Bias in meta-analysis detected by a simple, graphical test. Bmj, 315(7109), 629–634.

[42] E. S. Elliott, & C. S. Dweck (1988). Goals: an approach to motivation and achievement. Journal of Personality and Social Psychology, 54(1), 5–12. doi:10.1037/0022-3514.54.1.5

[43] F. Faul, E. Erdfelder, A. G. Lang, & A. Buchner (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175–191.

[44] G. T. Fechner (1871). *Vorschule der Ästhetik*. Hildesheim: Olms.

[45] G. V. Glass (1976). Primary, secondary, and meta-analysis of research. Educational researcher, 5(10), 3–8.

[46] K. C. Hamborg, J. Hülsmann, & K. Kaspar (2014). The interplay between usability and aesthetics: more evidence for the "What is usable is beautiful" notion. Advances in Human-Computer Interaction, 2014, 15.

[47] B. Hanrahan, & J. M. Carroll (2017). Ergonomics of Information Technologies at Work. In G. Hertel, D. Stone, R. Johnson, & J. Passmore (Eds.), *The Wiley Blackwell Handbook of the Psychology of the Internet at Work*, 39-60. Chichester: Wiley-Blackwell.

[48] J. Hartmann, A. Sutcliffe, & A. De Angeli (2007). Investigating attractiveness in web user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 387–396). ACM.

[49] M. Hassenzahl (2003). The thing and I: understanding the relationship between user and product. In *Funology* (pp. 31–42). Springer Netherlands.

[50] M. Hassenzahl, & A. Monk (2010). The Inference of Perceived Usability From Beauty. Human–Computer Interaction, 25(3), 235–260.

[51] E. Hollnagel (2003). Is affective computing an oxymoron? International Journal of Human Computer Studies, 59(1-2), 65–70. doi:10.1016/S1071-5819(03)00053-3

[52] W. Ilmberger, M. Schrepp & T. Held (2008). Cognitive Processes Causing the Relationship between Aesthetics and Usability. In *HCI and Usability for Education and Work* (pp. 43–54). Heidelberg: Springer.
[53] ISO (1998). *ISO 9241: Ergonomic requirements for office work with visual display terminals(VDTs) - Part 11: Guidance on usability*. Geneve: International Organization for Standardisation. *Switzerland*.
[54] ISO (2006). *ISO 9241: Ergonomics of Human-System Interaction – Part 151: Guidance on World Wide Web Interfaces*. Geneva: International Organization for Standardisation. *Switzerland*.
[55] ISO (2009). *ISO 9241: Ergonomics of Human-System Interaction – Part 210: Human-centered design for interactive system*. Geneva: International Organization for Standardisation. *Switzerland*.
[56] G. H. Iten, A. Troendle, & K. Opwis (2018). Aesthetics in Context—The Role of Aesthetics and Usage Mode for a Website's Success. Interacting with Computers, 30(2), 133–149.
[57] T. Lavie, & N. Tractinsky (2004). Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. International Journal of Human-Computer Studies, 60(3), 269 – 298. doi:10.1016/j.ijhcs.2003.09.002
[58] T. Lavie, T. Oron-Gilad, & J. Meyer (2011). Aesthetics and usability of in-vehicle navigation displays. International Journal of Human-Computer Studies, 69(1), 80–99.
[59] H. Leder, & M. Nadal (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode - Developments and challenges in empirical aesthetics. British Journal of Psychology, 105(4), 443–464. doi:10.1111/bjop.12084
[60] H. Leder, B. Belke, A. Oeberst, & D. Augustin (2004). A model of aesthetic appreciation and aesthetic judgements. British Journal of Psychology, 95(4), 489–508.
[61] G. Lindgaard (2007). Aesthetics, visual appeal, usability and user satisfaction: What do the user's eyes tell the user's brain? Australian Journal of Emerging Technologies and Society, 5(1), 1–14.
[62] G. Lindgaard, & C. Dudek (2003). What is this evasive beast we call user Satisfaction? Interacting with Computers, 15(3), 429–452.
[63] G. Lindgaard, G. Fernandes, C. Dudek, & J. Brown (2006). Attention webdesigners: You have 50 milliseconds to make a good first impression!. Behaviour & Information Technology 25(2), 115–126.
[64] Y. Liu (2003). Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. Ergonomics, (13/14), 1273–1292. doi:10.1080/0014013031000161082 9
[65] E. A. Locke, & G. P. Latham (2006). New directions in goal-setting theory. Current Directions in Psychological Science, 15(5), 265–268. doi:10.1111/j.1467-8721.2006.00449.x
[66] S. McDougall, I. Reppa, J. Kulik, & A. Taylor (2016). What makes icons appealing? The role of processing fluency in predicting icon appeal in different task contexts. Applied ergonomics, 55, 156–172.
[67] M. Moshagen, & M. T. Thielsch (2010). Facets of Visual Aesthetics. International Journal of Human-Computer Studies, 68(10), 689–709. doi:10.1016/j.ijhcs.2010.05.006
[68] M. Moshagen & M. T. Thielsch (2013). A short version of the visual aesthetics of websites inventory. Behaviour & Information Technology, 32(12), 1305–1311. doi:10.1080/0144929X.2012.694910
[69] D. A. Norman (1988). *The design of everyday things: Revised and expanded edition*. Basic Books (AZ). New York.
[70] D. A. Norman (2002). Emotion and Design: Attractive Things Work Better. Interactions Magazine, 9(4), 36–42.
[71] D. A. Norman (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books. doi:10.1145/966012.966013
[72] J. L. Plass, S. Heidig, E. O. Hayward, B. D. Homer, & E. Um (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. Learning and Instruction, 29, 128–140. doi:10.1016/j.learninstruc.2013.02.006
[73] T. Porat, & N. Tractinsky (2012). It's a pleasure buying here: The effects of web-store design on consumers' emotions and attitudes. Human–Computer Interaction, 27(3), 235–276.
[74] V. Postrel (2002). The substance of style: How the Rise of Aesthetic Value Is Remaking Commerce, Culture, and Consciousness. *New York*.
[75] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.
[76] R. Reber, N. Schwarz, & P. Winkielman (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? Personality and Social Psychology Review, 8(4), 364–382.
[77] K. Reinecke, & A. Bernstein (2011). Improving Performance, Perceived Usability, and Aesthetics with Culturally Adaptive User Interfaces. ACM Transactions on Computer Human Interaction, 18(2), 8. doi:10.1145/1970378.1970382
[78] I. Reppa, & S. McDougall (2015). When the going gets tough the beautiful get going: aesthetic appeal facilitates task performance. Psychonomic Bulletin & Review, 22(5), 1243–1254. doi:10.3758/s13423-014-0794-z
[79] I. Reppa, D. Playfoot, & S. J. P. McDougall (2008). Visual Aesthetic Appeal Speeds Processing of Complex but not Simple Icons. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52(18), 1155–1159. doi:10.1177/154193120805201801
[80] C. Salimun (2013). *The relationship between visual interface aesthetics, task performance, and preference* (Doctoral dissertation, University of Glasgow).
[81] G. Santayana (1896). *The sense of beauty: Being the outlines of aesthetic theory*. C. Scribner's sons.
[82] G. H. Seijts, & G. P. Latham (2005). Learning versus performance goals: When should each be used? Academy of Management Perspectives, 19(1), 124–131. doi:10.5465/AME.2005.15841964
[83] A. Sonderegger, G. Zbinden, A. Uebelbacher, & J. Sauer (2012). The influence of product aesthetics and usability over the course of time: a longitudinal field experiment. Ergonomics, 55(7), 713–30. doi:10.1080/00140139.2012.672658
[84] T. D. Sterling, W. L. Rosenbaum, & J. J. Weinkam (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. The American Statistician, 49(1), 108–112. doi:10.2307/2684823
[85] R. Strebe (2016). Aesthetics on the web: effects on approach and avoidance behaviour. Behaviour & Information Technology, 35(1), 4–20. doi:10.1080/0144929X.2015.1070202
[86] M. Szabo, & H. Kanuka (1998). Effects of violating screen design principles of balance, unity, and focus on recall learning, study time, and completion rates. Journal of Educational Multimedia and Hypermedia, 8(1), 23–42.
[87] M. T. Thielsch & G. Hirschfeld (2012). Spatial frequencies in aesthetic website evaluations – explaining how ultra-rapid evaluations are formed. Ergonomics, 55(7), 731–742. doi:10.1080/00140139.2012.665496
[88] M. T. Thielsch & J. Niesenhaus (2017). User experience, gamification, and performance. In G. Hertel, D. Stone, R. Johnson, & J. Passmore (Eds.), *The Wiley Blackwell Handbook of the Psychology of the Internet at Work*, 79-102. Chichester: Wiley-Blackwell.
[89] M. T. Thielsch, I. Blotenberg, & R. Jaron (2014). User Evaluation of Websites: From First Impression to Recommendation. Interacting with Computers, 26(1), 89–102. doi:10.1093/iwc/iwt033
[90] M. T. Thielsch, R. Engel, & G. Hirschfeld (2015). Expected usability is not a valid indicator of experienced usability. PeerJ Computer Science, 1, e19. doi:10.7717/peerj-cs.19
[91] N. Tractinsky (1997). Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In Pemberton, S. (Ed.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Atlanta, Georgia, USA, 115–122. ACM Press, New York. doi:10.1145/258549.258626
[92] N. Tractinsky (2013). Visual aesthetics. *The Encyclopedia of Human-Computer Interaction, 2nd Ed, 19*. Retrieved July 16, 2018, from https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/visual-aesthetics
[93] N. Tractinsky, A. Cokhavi, M. Kirschenbaum, & T. Sharfi (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. International Journal of Human - Computer Studies, 64(11), 1071–1083.
[94] N. Tractinsky, A. S. Katz, & D. Ikar (2000). What is Beautiful is Usable. Interacting with Computers, 13(2), 127–145. doi:10.1016/S0953-5438(00)00031-X
[95] A. N. Tuch, J. A. Bargas-Avila, K. Opwis, & F. H. Wilhelm (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. International Journal of Human-Computer Studies, 67(9), 703–715.
[96] A. N. Tuch, E. E. Presslaber, M. Stöcklin, K. Opwis, & J. A. Bargas-Avila (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. International Journal of Human-Computer Studies, 70(11), 794–811. doi:10.1016/j.ijhcs.2012.06.003
[97] E. Um, J. L. Plass, E. O. Hayward, & B. D. Homer (2012). Emotional Design in Multimedia Learning. Journal of Educational Psychology, 104(2), 485–498. doi:10.1037/a0026609
[98] W. Viechtbauer (2010). Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software, 36(3), 1–48.
[99] W. M. Wundt (1874). *Grundzüge der physiologischen Psychologie*. Leipzig: Engelmann.