

# Expected usability is not a valid indicator of experienced usability

Meinald T. Thielsch<sup>1</sup>, Ronja Engel<sup>1</sup> and Gerrit Hirschfeld<sup>2</sup>

<sup>1</sup> Department of Psychology, University of Münster, Münster, Germany

<sup>2</sup> Faculty of Business Management and Social Sciences, University of Applied Sciences Osnabrück, Osnabrück, Germany

## ABSTRACT

Usability is a core construct of website evaluation and inherently defined as interactive. Yet, when analysing first impressions of websites, expected usability, i.e., before use, is of interest. Here we investigate to what extent ratings of expected usability are related to (a) experienced usability, i.e., ratings after use, and (b) objective usability measures, i.e., task performance. Furthermore, we try to elucidate how ratings of expected usability are correlated to aesthetic judgments. In an experiment, 57 participants submitted expected usability ratings after the presentation of website screenshots in three viewing-time conditions (50, 500, and 10,000 ms) and after an interactive task (experienced usability). Additionally, objective usability measures (task completion and duration) and subjective aesthetics evaluations were recorded for each website. The results at both the group and individual level show that expected usability ratings are not significantly related either to experienced usability or objective usability measures. Instead, they are highly correlated with aesthetics ratings. Taken together, our results highlight the need for interaction in empirical website usability testing, even when exploring very early usability impressions. In our study, user ratings of expected usability were no valid proxy neither for objective usability nor for experienced website usability.

**Subjects** Human–Computer Interaction, World Wide Web and Web Science

**Keywords** Usability, Aesthetics, First impression, Website evaluation, Website

Submitted 22 April 2015  
Accepted 10 August 2015  
Published 2 September 2015

Corresponding author  
Meinald T. Thielsch,  
thielsch@uni-muenster.de

Academic editor  
Helen Petrie

Additional Information and  
Declarations can be found on  
page 16

DOI 10.7717/peerj-cs.19

© Copyright  
2015 Thielsch et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

## INTRODUCTION

The importance of the World Wide Web, especially in the context of business matters, raises the question of how users perceive and evaluate websites. A core construct for the evaluation of websites and other interactive products is usability (*Hornbæk, 2006*). The definition of usability focuses strongly on interactivity (see below), but in recent research expected usability, i.e., the usability based on first impressions, as well as experienced usability, i.e., usability evaluations made after interaction, became of interest (cf. *Lee & Koubek, 2012*). Several studies have to some extent implicitly relied on the idea that subjects are able to evaluate the usability of a website based on visual cues on a static website screenshot before interacting with it. Thus far, only few studies focussed on such effects of expected usability (cf. *Lee & Koubek, 2012*): some studies found that expected usability does have an influence on the perceived overall quality of a website (*Van Schaik & Ling, 2008*), overall impressions (*Kim & Fesenmaier, 2008*) and user preferences for a web

design (Lee & Koubek, 2010). Additionally, Lindgaard et al. (2011) found a high stability of very early ratings of perceived usability given after the presentation of screenshots for 50 ms ( $r = .92, p < .01$ ).

However, several studies have shown that expected usability of a system (Ben-Bassat, Meyer & Tractinsky, 2006; Tractinsky, Katz & Ikar, 2000) or a website (Lee & Koubek, 2010; Katz, 2010; Lindgaard et al., 2011) is systematically connected to ratings of visual aesthetics. In particular, when evaluating a website via a brief presentation time such as 50 or 500 ms (as established in a study by Lindgaard et al., 2006), it may be asked whether a valid measurement of usability is feasible at all. The current study therefore aims to examine the connection between early and very early ratings of expected usability and experienced usability as well as objective usability measures after use. Based on the prior research on early aesthetic perceptions, we expect to encounter major problems in validly assessing usability before use, especially when analysing early first impressions. However, before we discuss what kind of aesthetics bias might influence such first impression measurements, we will examine the definition and measurement of website usability and the current research on expected usability based on first impressions.

### Definition and evaluation of website usability

Several definitions of usability have been proposed in the field of human–computer interaction, most of which stress the importance of interaction and the context of the evaluation (Hornbæk, 2006, p. 79). ISO 9241-11 is cited frequently and defines the usability of a human–system interaction as the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO, 1998, p. 2). This definition is primarily based on user performance and often is used as a definition of website usability (e.g., Flavián, Guinalíu & Gurrea, 2006; Green & Pearson, 2006; Thielsch, Blotenberg & Jaron, 2014). Other aspects of website usability are sometimes referred to as the perceived “ease of use” and “usefulness,” with the latter term being connected to information richness and other aspects closely associated with content perception (cf., Flavián, Guinalíu & Gurrea, 2006). Thus, the core aspects of website usability are the website’s simplicity of use, especially in the initial use phases, and the ease of understanding its structure and navigation.

Furthermore, it is crucial to distinguish between objective and subjective measures of usability (Hornbæk, 2006). To a developer, a website might appear very usable according to objective criteria (e.g., a flat link depth, fast loading and well-functioning search functions), but it still can be experienced as unusable from a user’s subjective point of view (e.g., due to misunderstood navigation, problems with the wording of links or distracting website elements). To grasp these two different sides of usability, Kurosu & Kashimura (1995) established the terms “inherent” (= objective) and “apparent” (= subjective) usability. Thus, subjective usability results from users’ perception and attitudes towards a website, while objective usability refers to website aspects not dependent on users’ perception (Hornbæk, 2006). For example, such aspects can be error rates, task completion time and speed of performance or retention over time, but

objective usability can also be derived from an expert's assessment (*Hornbæk, 2006*). This differentiation between objective and subjective usability has been utilized subsequently by several authors (e.g., *Hornbæk, 2006*; *Lee & Koubek, 2012*; *Sonderegger et al., 2012*) and will be used in the current paper as well. Additionally, subjective usability ratings can be split in those ratings given before an interaction with a website (pre-use) and those given after an interaction (post-use, cf. *Lee & Koubek, 2012*). In this paper, we will refer to subjective pre-use ratings of usability as “expected usability.” Subjective post-use ratings of usability given after an interaction will be referred to as “experienced usability.”

The remaining question is to what extent users are able to judge the usability of a website without actually using it and how this can be utilized in first impression testing. Therefore, we will give an overview of findings in respect to expected usability based on first impressions in the next section.

### **Based on first impressions: expected usability**

Thus far, only few studies focussed on the effects of expected usability in terms of a usability prediction by users before interacting with a system (cf. *Lee & Koubek, 2012*)—and of those few studies only some examined the usability of websites in particular. For example, *Van Schaik & Ling (2008)* expected a connection of pre-use perceptions and post-use evaluations. Thus, they proposed a model based on system characteristics, in which pre-use perceptions influence evaluations after use. *Van Schaik & Ling (2008)* tested four experimentally varied versions of an intranet web site with  $N = 111$  students and showed that pragmatic quality was a good predictor of evaluations of goodness (the overall quality of a website) before and after use. *Kim & Fesenmaier (2008)* asked  $N = 65$  students to rate screenshots of 50 official state tourism web sites in the United States and found that the first impression was a significant predictor of overall impression using a regression analysis.

*Lee & Koubek (2010)* requested  $N = 10$  students to evaluate nine online bookstore web sites. They found small correlations between pre-use usability ratings and task completion time ( $.21 \leq r \leq .28$ ), but high correlations between pre-use usability and user preferences for a web design ( $.48 \leq r \leq .61$ ). *Katz (2010)* gave a sample of  $N = 60$  students the task to evaluate a search website varied on different levels of aesthetics in a between-subject-design. She found small, non-significant correlations between pre-use usability ratings and post-evaluations of usefulness ( $-.10 \leq r \leq .18$ ), but a significant correlation between pre-use usability and pre-use aesthetics ( $r = .40$ ).

If expected usability is estimated before use and aspects like first impressions are evaluated, researchers tend to use static website screenshots due to technical reasons (e.g., *Kim & Fesenmaier, 2008*; *Lindgaard et al., 2011*; *Schenkman & Jönsson, 2000*; *Tuch et al., 2012b*). This approach is based on the idea that subjects are able to evaluate the expected usability before actual use based on visual cues on a static website screenshot. Such cues may be easy-to-find search boxes, precisely labelled navigation elements, or a clear and structured overall design. But this procedure contradicts to some extent the definition of usability which focuses strongly on interactivity (see above). And if asked, most professionals in the field would agree that interactive tasks are necessary to measure

website usability in depth, based on the impossibility of measuring user performance without asking the user to interact with the website under evaluation. Additionally, several studies have shown that pre-use evaluations of usability and aesthetics of a system (e.g., *Ben-Bassat, Meyer & Tractinsky, 2006*; *Tractinsky, Katz & Ikar, 2000*) or a website (e.g., *Lee & Koubek, 2010*; *Lindgaard et al., 2011*) are systematically connected. This leads to the question how expectations of usability in the formation of first impressions are connected to aesthetics. We will discuss this potential bias in the following section.

### Aesthetics and its connection to usability

Aesthetics has become a core construct in the understanding of website perception (e.g., *Moshagen & Thielsch, 2010*; *Thielsch, Blotenberg & Jaron, 2014*; *Tuch et al., 2012a*). Website aesthetics can be defined “as an immediate pleasurable subjective experience that is directed toward an object and not mediated by intervening reasoning” (*Moshagen & Thielsch, 2010*, p. 690). It is important to note the high speed of aesthetic perceptions; such impressions occur at first sight, i.e., within milliseconds (*Leder et al., 2004*). The perception of websites is no exception: aesthetic first impressions are made quickly, probably within less than 500 ms after viewing a webpage for the first time (cf. *Lindgaard et al., 2006*; *Lindgaard et al., 2011*; *Tractinsky et al., 2006*; *Thielsch & Hirschfeld, 2012*; *Tuch et al., 2012a*).

Furthermore, aesthetic perceptions appear to have a strong impact on subjective usability evaluations (for overviews, see *Hassenzahl & Monk, 2010*; *Lee & Koubek, 2012*; *Tuch et al., 2012b*). The analysis of the connection between subjective usability and aesthetics extends back to the study conducted by *Kurosu & Kashimura (1995)* and its replication by *Tractinsky (1997)*. Additionally, with the research started by *Tractinsky, Katz & Ikar (2000)*, we have learned that even post-use evaluations of experienced usability may be influenced by aesthetics perceptions. *Lee & Koubek (2012)* summarize how pre-use evaluations of a system demonstrate a strong connection between expected usability and aesthetics and how this connection persists even after use. Most studies examining this connection report high correlations between the constructs. Many different explanations for these correlations have been proposed: halo effects (e.g., *De Angeli, Sutcliffe & Hartmann, 2006*; *Hartmann, Sutcliffe & De Angeli, 2008*), mediation by other constructs (e.g., *Hassenzahl & Monk, 2010*), mediation by mood or affective experiences (e.g., *Moshagen, Musch & Göritz, 2009*; *Tuch et al., 2012b*), common design features (e.g., *Lavie & Tractinsky, 2004*; *Tarasewich, Daniel & Griffin, 2001*), common method bias<sup>1</sup> because in some experiments, usability and aesthetics were treated without affecting each other (*Thielsch, 2008*), or measurement issues that reduce scale validity (*Tuch et al., 2012b*).

Moreover, the temporal dimension of website perception is particularly interesting. Thus far, we know that aesthetic perceptions occur very quickly, as described above. *Thielsch, Blotenberg & Jaron (2014)* demonstrated that aesthetics had the greatest impact on deliberate first impressions while usability and content are also relevant to first and overall impressions. Content has the largest impact on the user’s intention to recommend or revisit a website, while aesthetics exerts only a small effect and usability exhibits no

<sup>1</sup> For further information about common method bias effects, see *Podsakoff et al. (2003)*.

significant influence (*Thielsch, Blotenberg & Jaron, 2014*). Additionally, several studies deal with pre- and post-use differences in user evaluations; an excellent overview can be found in *Lee & Koubek (2012)*, who proposed a model in which usability and aesthetics are strongly connected in pre-use evaluations and relatively weakly connected after use. Additionally, in their model, aesthetics influence preference more strongly than usability before use; after use, both constructs affect web users' preferences equally. Such a waning effect of aesthetics on perceived usability over time has been found in recent studies and different domains (e.g., for mobile phones, see *Sonderegger et al., 2012*; for websites, see *Moshagen & Thielsch, 2010*). Nevertheless, some authors question the decrease in correlation between usability and aesthetics after use (for an overview, see *Lee & Koubek, 2012*). However, because many of the existing studies in the field do not focus on websites (examining instead other IT products, such as ATMs, mobile phones or software products), differences in the use scenarios may be influencing the results. *Lee & Koubek (2012, p. 107 et seq.)* discuss several possible reasons for conflicting findings and possible moderating variables, such as measurement issues, task characteristics or participants' experience. Still, although we do not understand the exact mechanism and timeline behind the effects, it is probable that website aesthetics have a strong effect on expected usability evaluations of websites.

### **Aim of our study**

To summarize our introduction: website usability is generally defined as a construct largely based on user interaction with a given website. This interaction can be subjectively perceived by a user or analysed objectively. Subjective usability evaluations can be further classified in ratings given before and after an interaction, in this paper referred as expected usability (= pre-use) and experienced usability (= post-use). Furthermore, recent studies suggest that there is a strong connection between usability and aesthetics, especially in pre-use scenarios. Because aesthetic perceptions occur very quickly—within milliseconds of a website's visit—it seems likely that they affect subjective usability evaluations, especially in pre-use scenarios when users are asked for the expected usability of a website based on their first impressions.

This leads to our research approach: First, in the light of the interactive definition of usability, is the measurement of website usability based on the expected usability build on first impressions impossible? Or is it possible to validly assess usability using shortly presented, static screenshots with evaluations based only on visual usability cues like easy-to-find search boxes or precisely labelled navigation elements? Our approach to further investigate this matter combines subjective ratings of expected usability before actual use with measures of objective and experienced usability after use.

Secondly, it seems clear that early first impressions of websites are mainly driven by aesthetics, and *Lindgaard et al. (2011)* found a strong impact of appeal on first impression usability ratings. Several studies have analysed the stability of aesthetics made after very short presentations of screenshots (e.g., *Lindgaard et al., 2006*; *Thielsch & Hirschfeld, 2012*; *Tractinsky et al., 2006*; for an overview see *Tuch et al., 2012a*), but thus far only *Lindgaard*

*et al. (2011)* ran the same analyses on usability ratings given after the presentation of screenshots for 50 ms—and found a high stability of usability ratings ( $r = .92, p < .01$ ). But to our knowledge a systematic comparison of such expected usability ratings given after short (500 ms) and very short (50 ms) time presentations of website screenshots and post-use usability ratings is still missing. Besides the study of *Lindgaard et al. (2011)* analysing very early first impression ratings of usability after 50 ms presentations of screenshots this was done mainly with a focus on early perceptions of aesthetics (see *Tuch et al., 2012a*). First impressions often influence mid- and long-term human behaviour, being a crucial moment for capturing the users interests (cf. *Tuch et al., 2012a*)—and thus are part of many web usability tests. Additionally, stimulated by the current research shorter and shorter time frames are tested (even 17 or 33 ms) and very early first impressions are becoming more and more relevant. Thus, the question arises, whether our current knowledge of the impact of aesthetics on first impression means that it is completely impossible to validly measure expected usability and very early first impressions of usability before website use, or if such usability ratings are only partly influenced by aesthetics but still feasible for usability measurements.

Summing up, those two aspects lead to two research questions we would like to examine in our current study:

1. To what extend are ratings of expected usability connected to experienced usability after use and objective usability measures?
2. To what extend are ratings of expected usability evaluations connected to the perceived aesthetics of a website?

To address these questions, we will utilize correlation analyses on both aggregated and individual levels, using data from usability tests of several websites. Based on previous studies revealing a strong connection between pre-use evaluations of usability and aesthetics (cf. *Lee & Koubek, 2012*) and the rapid processing of aesthetic stimuli (cf. *Tuch et al., 2012a*), we assume that ratings of expected usability will show only weak connections to objective and experienced usability after use. This accounts for early (presentation of a screenshot for 500 ms) and very early (presentation of a screenshot for 50 ms) first impressions of expected usability, which are probably biased by the user's aesthetic perceptions.

## METHOD

The aim of our study is to investigate to what extend ratings of expected usability are related to (a) experienced usability ratings after use, and (b) objective usability measures (i.e., task performance). Additionally, we examine how ratings of expected usability are correlated to aesthetic judgments. Thus, we performed an experiment with 57 participants who submitted expected usability ratings after the presentation of website screenshots in three viewing-time conditions (50, 500, and 10,000 ms) and after an interactive task (experienced usability). Furthermore, objective usability measures (task completion and duration) and subjective aesthetics evaluations were recorded for each website. A detailed description of our methods can be found within the following subsections.



## Participants

A total of 57 students ( $n = 40$  female) volunteered to participate in this study on an anonymous basis and received course credits. The participants' ages ranged from 19 to 32 years ( $M = 23.09$ ;  $SD = 2.79$ ), and the mean length of their experience using the Internet was 9.61 years ( $SD = 2.45$ ). The amount of time participants surfed the Internet per day varied between 10 min and 4 h ( $M = 1.41$  h;  $SD = 0.90$ ). Of the participants, 13.1% had a part-time job in the area of web design, user experience or usability, but none worked full-time in these occupational fields.

## Stimulus material

To represent a wide range of corporate and institutional websites in Germany, we selected 10 websites (see [Appendix S1](#)) from the following 10 different content domains: download & software, e-commerce, entertainment, e-learning, e-recruiting & e-assessment, information sites, web portals, social software & weblogs, corporate sites and search engines. In the first three conditions, screenshots of these websites showed the index page and were scaled to  $1280 \times 800$  pixels. They were sandwich-masked with a scrambled version of themselves. Masks are used to inhibit the formation of after-images (e.g., [Breitmeyer, 2007](#); [Enns & Di Lollo, 2000](#)); these were created by decomposing the targets into  $2 \times 2$ -pixel segments that were randomly rearranged using Matlab (Version 7.8).

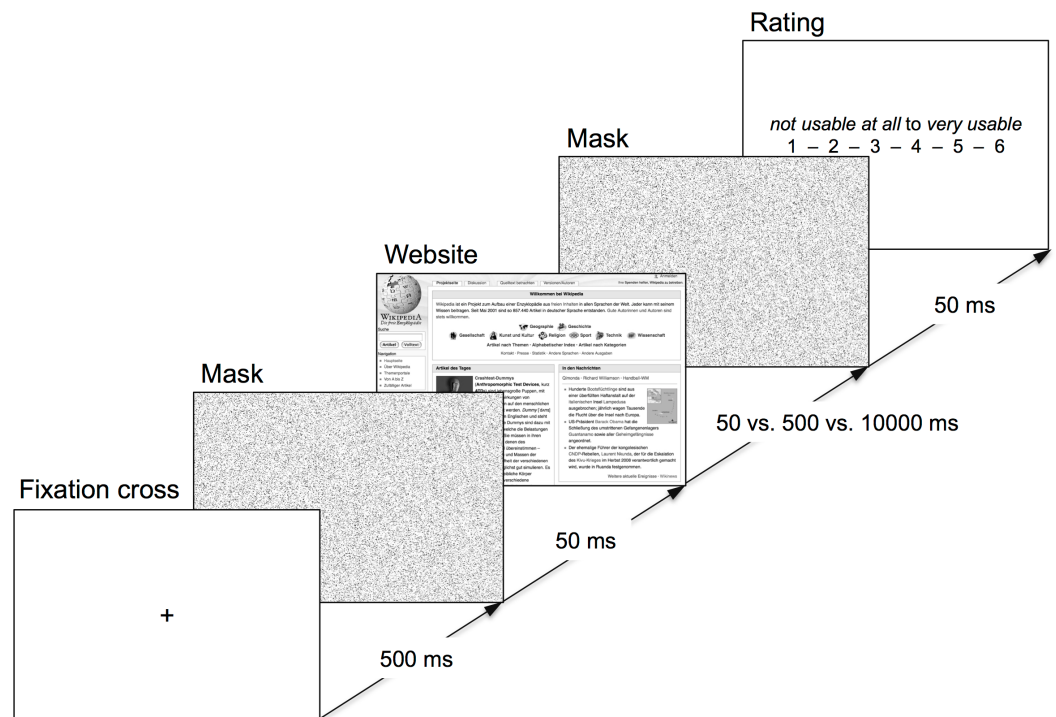
## Procedure and measures

We used a within-subject design in which participants rated usability of websites after viewing a screenshot for 50, 500, or 10,000 ms (expected usability), and after interacting with the website (experienced usability). Furthermore, two aspects of task-performance—time to complete these tasks as well as their accuracy—were recorded as objective markers of usability. Finally, participants rated the perceived aesthetics of the websites.

The experiment took place in a computer lab at the University of Münster in groups of up to 10 participants. It consisted of five parts; (1) information and demographics, (2) expected usability, (3) task performance, (4) experienced usability, and (5) visual aesthetics.

At first, participants were informed about anonymity, procedure and the use of data for scientific purposes and completed a short sociodemographic survey (e.g., age, gender, occupation, experience using the Internet).

Second, participants rated the expected usability. For this, stimuli were presented at a resolution of  $1,280 \times 1,224$  pixels on 19-inch LCD displays connected to IBM PCs (2.16 GHz clock speed, 2 GB RAM) running Inquisit (Version 3). Each participant completed three blocks of stimuli, in which the presentation time was set to either 50, 500 or 10,000 ms. According to prior research, presentation times of 50 ms and 500 ms reflect very early and early first impression phases, respectively ([Lindgaard et al., 2006](#); [Thielsch & Hirschfeld, 2012](#); [Tractinsky et al., 2006](#)); the 10,000 ms condition reflect normal viewing without any interaction. The order of the time conditions was pseudo-randomly distributed across the participants following a Latin Square. Each block consisted of 60 websites: a starting series of 10 randomised warm-ups, followed by the 10 experimental



**Figure 1** Procedure used in the first part of the experiment (illustration of a single trial).

websites (named in [Appendix S1](#)) mixed with 40 other website screenshots (not pertinent to the current study), all presented in randomised order. Subsequent evaluations of a website might be influenced by the first presentation because repetition improves the processing of the repeated stimuli, an effect called *repetition priming*. This effect manifests itself in various ways (see [Grill-Spector, Henson & Martin, 2006](#) for an overview) and as well in website evaluations (see [Thielsch & Hirschfeld, 2012](#)). Thus, to reduce possible repetition bias effects we added those 40 websites mentioned above to the first three website presentation trials. The warm-ups and additional screenshots were not analysed further. Each trial began with a black fixation cross presented in the middle of the white screen for 500 ms, followed by the presentation of the individual mask for 50 ms. Directly after the offset of the mask, the screenshots were presented for either 50, 500 or 10,000 ms. After each screenshot, the mask appeared again for 50 ms. As soon as this final backward mask disappeared, the participants were asked to indicate their usability evaluations on a six-point Likert scale ranging from “not usable at all” to “not usable,” “rather not usable,” “rather usable,” “usable” and “very usable” by clicking on a label using the computer mouse (see [Fig. 1](#)). After the rating, there was a short break of 300 ms before the next trial began with the fixation cross.

Third, task-performance was assessed after participants interacted with the website. At the beginning of this part participants received brief instructions indicating that they would perform several tasks based on different websites. We created tasks with a medium difficulty and five answers in a multiple-choice format (see [Appendix S1](#)). For each website,



one alternative was correct, and three functioned as distractors. These first four alternatives were presented in randomised order; the fifth alternative was always “I have tried but could not find an answer.” After one warm-up task 10 tasks followed that were analysed. To avoid possible sequence effects, we presented the 10 tasks in inverse order to half of the participants. For this part of the study, we used EFS Survey (Version 8). At the beginning of each trial, we presented the task and then the corresponding fully functioning website. While the user worked with the website, the task was still visible in a small frame at the top of the screen. The participants had unlimited time to complete the given task. When the task was complete, they could click a button marked “next”; they were then presented with a question to answer by choosing between five alternatives.

Fourth, experienced usability was assessed. Participants were asked to rate the website’s usability on the familiar six-point Likert scale used in the first three trials and to response to an additional questionnaire adapted to German from *Flavián, Guinalú & Gurrea (2006)*. The translation and adaption of this questionnaire is discussed in *Thielsch (2008)*. In comparison with the original the German version was shortened by one item and adjusted in wording. Like the original, this adapted version showed objectivity, high reliability (Cronbachs  $\alpha = .95$ ) and high validity (see *Thielsch, 2008*). The adapted questionnaire consisted of the following seven statements about perceived subjective usability, to be rated on a seven-point Likert scale (ranging from “totally disagree” to “totally agree”):

1. I think the use of this website is easy to understand.
2. This website is simple to use, even when using it for the first time.
3. It is easy for me to find the sought information.
4. I can easily understand the structure of this website.
5. It is easy to navigate within this website.
6. Contents are organized in a way that I know where I am at any time.
7. I am able to find the required information quickly.

Fifth, visual aesthetics was assessed. For this the original screenshots of the 10 websites were shown again in randomised order and without a time limit. Participants were asked to indicate their ratings of the websites’ aesthetics on six-point Likert scales ranging from “not aesthetic at all” to “not aesthetic,” “rather not aesthetic,” “rather aesthetic,” “aesthetic” and “very aesthetic.” Additionally, their familiarity with each website was assessed on Likert scales ranging from “I have never used the website before” to “less than once per month,” “about once per month,” “several times per month,” “several times per week” and “daily.” At the end of the experiment, participants were thanked and asked for their informed consent. They had the opportunity to exclude their data from the subsequent analysis and to comment on the study. The full experiment took approximately 40–50 min.

## RESULTS

### Checks for bias effects

Before starting the main analyses, we checked our data for possible bias effects. To this end, we noted the average usability ratings in each condition. These means were as follows:

- $M = 3.76$  ( $SD = 0.34$ ) in the 50 ms screenshot presentation condition,
- $M = 3.81$  ( $SD = 0.41$ ) in the 500 ms screenshot presentation condition,
- $M = 3.80$  ( $SD = 0.47$ ) in the 10,000 ms screenshot presentation condition, and
- $M = 3.82$  ( $SD = 0.64$ ) in the interaction condition.

We used the ratings aggregated across participants and calculated a repeated-measures ANOVA with the rating as the dependent variable and the condition as the independent variable (with four levels). Because there are no significant differences between the usability ratings of the four conditions ( $F(2, 24) = 0.04$ ,  $p = .96$ ), a selective, conditions-based answer bias is quite improbable (means per website can be found in [Appendix S2](#)).

Additionally, we checked for possible bias effects within the interaction condition due to task difficulty. The difficulties of all the tasks were comparable within a range from .65 to .95, and there were no outlier or extreme values (see [Appendix S1](#)). Furthermore, we found no significant correlation between familiarity with a website presented as a screenshot and usability ratings ( $.03 \leq r \leq .21$ , all  $ps > .05$ ).

Finally, we checked for the correlation between the single item that we used in the usability evaluation and the usability scale adapted from [Flavián, Guinalíu & Gurrea \(2006\)](#): the correlations between the scale and the single-item usability ratings after 50 ms ( $r = .16$ ;  $p = .65$ ), 500 ms ( $r = .19$ ;  $p = .60$ ), and 10,000 ms ( $r = .25$ ;  $p = .48$ ) were all non-significant, while the post-use single-item rating given after the interaction with a website correlated highly with the usability scale ( $r = .98$ ;  $p < .001$ ).

### Correlations at the group level

Our main research questions concerned the correlations between expected usability and experienced usability, objective usability (i.e., task performance), and aesthetics. As a first step, the correlations between subjective ratings and objective usability were calculated based on the average rating across all participants, which is the most important way to summarise these data for professionals designing websites ([Monk, 2004](#)) and common practice in this area of research (e.g., [Lindgaard et al., 2006](#); [Tractinsky et al., 2006](#); [Thielsch & Hirschfeld, 2012](#)). The results of this analysis revealed a very clear pattern, which is presented in [Table 1](#) (for a full correlation table see [Appendix S3](#)). First, we found low, and mostly non-significant, correlations between expected usability (made after viewing a screenshot for 50, 500 or 10,000 ms) and both measures of experienced usability. Second, we found low and non-significant correlations between expected usability and task-performance. Third, after interaction with the website, experienced usability was significantly correlated with task-performance, i.e., task completion and duration. Websites on which participants showed good performance (more correct and faster responses) had higher experienced usability ratings. Fourth, very high correlations were found between the expected usability ratings and the aesthetic ratings. At the same time, the experienced usability ratings showed a non-significant correlation to aesthetic ratings, indicating that this is not a general effect of aesthetics.

In summary, on the level of aggregated data participants' expected usability ratings seem quite unrelated to experienced usability and objective measures of task performance. At

**Table 1** Correlations between expected usability ratings in different conditions with experienced usability and objective usability measures as well as with subjective aesthetics ratings.

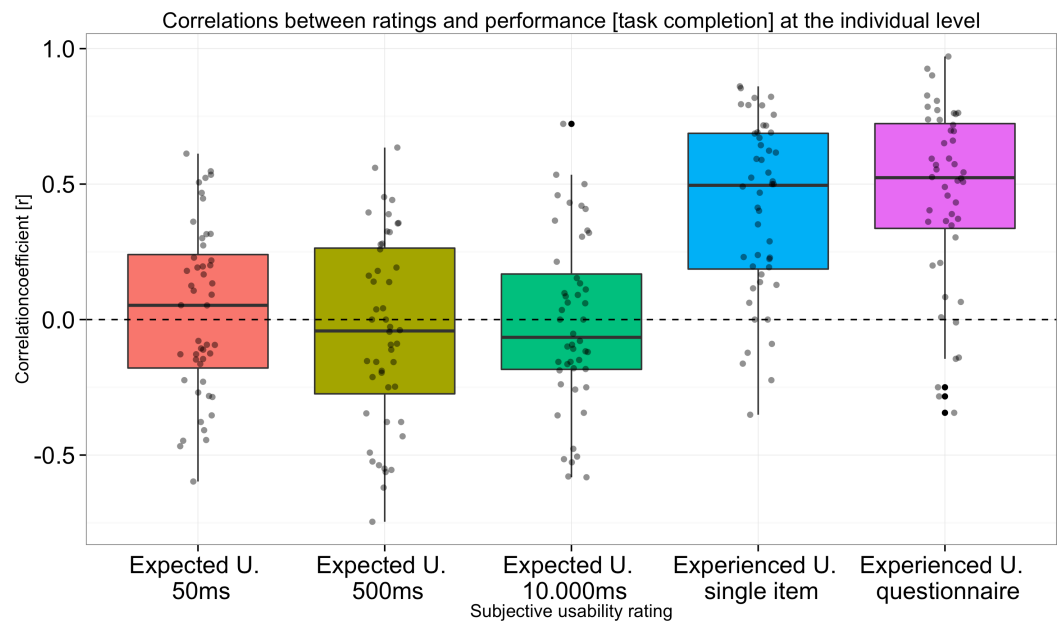
	Experienced usability (measured via single item)	Experienced usability (measured via usability questionnaire)	Task completion	Task duration <sup>a</sup>	Aesthetics rating
Expected usability (rating given after 50 ms presentation)	.13	.16	.19	.08	.75 <sup>*</sup>
Expected usability (rating given after 500 ms presentation)	.17	.19	.07	-.11	.75 <sup>*</sup>
Expected usability (rating given after 10 s presentation)	.19	.25	-.07	-.19	.84 <sup>**</sup>
Experienced usability (measured via single item)	–	.98 <sup>**</sup>	.62 <sup>*</sup>	-.86 <sup>**</sup>	.05
Experienced usability (measured via usability questionnaire)	.98 <sup>**</sup>	–	.59	-.94 <sup>**</sup>	.06

**Notes.**\*  $p < .05$ .\*\*  $p < .01$ .<sup>a</sup> Task duration was calculated only with correct answers.

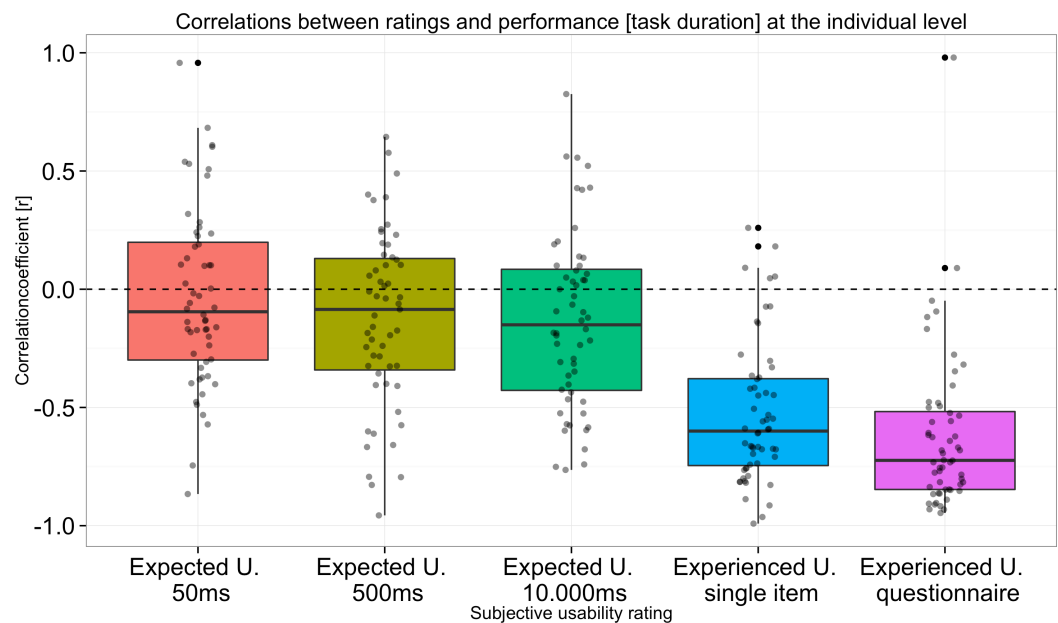
the same time expected usability ratings are strongly correlated to aesthetic ratings. These findings are in line with our assumption that users' ratings of pre-use usability are strongly connected to aesthetics perceptions, especially in first impression testing, preventing a valid measurement of experienced usability only based on usability expectations build before use. As pre-use ratings of expected usability are highly connected to aesthetic judgments but not at all to several other measures of subjective or objective usability, such ratings appear to be inappropriate as a proxy of experienced usability.

### Correlations at the individual level

To test whether the lack of correlation between expected usability ratings and experienced usability and objective usability (i.e., task performance) was due to the data aggregation, we also computed those correlations separately for each participant. This leads to a similar pattern of results: Inspection of the correlations for task completion (Fig. 2) and task duration (Fig. 3) revealed that the pre-use ratings of expected usability (after a presentation of website screenshots for 50, 500, and 10,000 ms) did not predict objective measures of performance (median  $r$  between  $-.09$  and  $.03$ ). In contrast, both the post-use single-item rating and the questionnaire rating of experienced usability provided after interacting with the website were systematically related to performance (median  $r = -.61$ ). Therefore, participants who judged a specific website to be more usable required less time to complete the task on that website.



**Figure 2** Correlations between performance (task completion rate) and usability ratings at the individual level. Usability ratings were given pre-use after the presentation of a screenshot for 50, 500 or 10.000 ms (= expected usability) and post-use after interaction on a single item and on an usability scale (= experienced usability).



**Figure 3** Correlations between performance (task duration time) and usability ratings at the individual level. Usability ratings were given pre-use after the presentation of a screenshot for 50, 500 or 10.000 ms (= expected usability) and post-use after interaction on a single item and on an usability scale (= experienced usability).

In summary, the pattern found in the analysis at the group level also appears in an analysis on the individual level. Thus, especially early and very early ratings of expected usability as well as expected usability evaluations given after 10 s were not significantly connected to objective and experienced usability measures after use.

## DISCUSSION

We observed a very clear result pattern that is in line with our assumptions, on both the group and the individual analysis level: we demonstrated that ratings of expected usability given after the presentation of website screenshots for 50, 500 or 10,000 ms exhibited no significant correlation with objective or experienced usability measures after use, but are highly correlated with aesthetics ratings. Thus, we found evidence that in website first impression testing there is a systematic connection between expected usability and aesthetics in a pre-use scenario (cf. [Lee & Koubek, 2012](#)). On the contrary, there is no significant correlation between ratings of expected usability and users' post-use subjective or objective usability evaluations of a website. The correlations found between expected usability ratings and objective measures as task completion time are even smaller than the ones reported by [Lee & Koubek \(2010\)](#), but comparable to the results of [Stojmenovic, Pilgrim & Lindgaard \(2014\)](#). The slightly higher correlation in [Lee & Koubek \(2010\)](#) might be caused by their homogeneous stimulus set of nine book store websites and performance gains due to the similarity of stimuli. In the current study websites from different content domains were used, thus prohibiting such effects.

The very high correlations between expected usability ratings and aesthetics ratings found in our experiment are in line with prior research (cf. [Lee & Koubek, 2012](#)). This finding is consistent with the idea that the formation of a first impression of usability based on a static website screenshot may be driven by a halo effect of aesthetics (cf. [De Angeli, Sutcliffe & Hartmann, 2006](#); [Hartmann, Sutcliffe & De Angeli, 2008](#); [Lindgaard et al., 2011](#)). As several studies stress the speed of first aesthetic impressions in human–computer interaction (for an overview, see [Tuch et al., 2012a](#)), we suppose that aesthetic website cues might be processed much more quickly than usability cues (cf. [Thielsch & Hirschfeld, 2010](#)). Hence, the idea that subjects are able to evaluate the pre-use expected usability based on visual usability cues on a static website screenshot might not to hold true—at least for early or very early first impression evaluations. A possible alternative interpretation would be that users rely on the same cues to judge expected usability that they use to judge aesthetics. But in sum, our results echo the critique advanced by [Bargas-Avila & Hornbæk \(2011\)](#), who demanded better efforts to study the complex interaction processes of user experience.

Additionally, in our experiment the correlation between usability and aesthetics ratings was not significant after use. This waning of the connection of aesthetics and usability evaluations after an interaction was also observed in prior studies (e.g., [Moshagen & Thielsch, 2010](#); [Sonderegger et al., 2012](#)). The strong but not full correlation between subjective and objective measures of usability after use is consistent with prior results observed by [Kurosu & Kashimura \(1995\)](#) in their study of ATMs. In website research these constructs generally do not appear to be perfectly correlated, as user perceptions can be

affected by other factors (see [Lee & Koubek, 2012](#)). For example, [Moshagen, Musch & Göritz \(2009\)](#) found a compensating effect of aesthetic colours on task completion time under the condition of poor objective usability.

### Practical implications

The main implication is that expected usability seems not to be a good proxy for measures of task performance or experienced usability. We assume that the absence of interaction is a main issue in usability first impression measurement. Our paper is a warning to practitioners (and researchers as well) who might assume that, at the very least, first impression ratings of usability could be gathered based only on static screenshots due to their visual usability cues. Appropriate methods for evaluating subjective and objective usability are needed, as [Hornbæk \(2006\)](#) illustrates. It therefore appears necessary to ask users to complete short tasks instead of relying on screenshot evaluations alone, even when evaluating usability first impressions (for an example of such a short task, see [Lee & Koubek, 2010](#)).

If expected usability is focussed and screenshots or other non-interactive websites must be used (for example, in prototyping), one must be aware of the possible effects of aesthetics and control for them. Due to the strong impact of aesthetics on the early stages of perception, aesthetics must be treated with special care and should always be evaluated in website first impression studies. This could be accomplished using a single item, as in the present study, or with standardised instruments, such as that used by [Lavie & Tractinsky \(2004\)](#), the VisAWI ([Moshagen & Thielsch, 2010](#)) or its four-item short form VisAWI-S ([Moshagen & Thielsch, 2013](#)). Controlling for aesthetics in this way may be helpful in several usability test scenarios because in some cases aesthetics can even influence the perception of objective usability and performance measures (e.g., [Moshagen, Musch & Göritz, 2009](#); [Sonderegger & Sauer, 2010](#)). It is important to mention that our study was designed to test user perceptions of websites. Focusing on first impressions—screenshot approaches might still work very well in web usability evaluations in other contexts: for example, in heuristic evaluations with experts (e.g., [Allen et al., 2006](#)), in Wizard of Oz studies when prototyping (cf. [Dahlbäck, Jönsson & Ahrenberg, 1993](#)), or in the evaluation of special services (e.g., [Haklay & Zafiri, 2008](#)).

### Limitations and future research

Certain limitations should be considered when interpreting the results of our present study. In this section, we will discuss these concerns and some potential approaches for future research.

First, it might be argued that 10,000 ms is too short to assess the subjective usability of a website. Yet from our point of view, a 10-second time span is quite long in terms of website use. For example, in a study conducted by [Robins & Holmes \(2008\)](#), participants took an average of only approximately 3.5 s to assess their first impressions of a website's credibility; no more than five seconds were taken, on average, to make such an assessment for any website in their stimulus set. It might be interesting to test a long viewing condition of more than 10 s in an experiment such as ours, but we would expect the same result as found in our experiment.



Second, we did not control for repetition priming effects (see *Grill-Spector, Henson & Martin, 2006*), which were found to affect aesthetics ratings in website evaluations (*Thielsch & Hirschfeld, 2012*). In our study's design, the participants viewed each stimulus four times, the first three times in different presentation time conditions. This type of repetition could lead to higher correlations and the overestimation of correlative effects (*Thielsch & Hirschfeld, 2012*). We tried to reduce those possible effects by presenting additional 40 website stimuli within the first three trials (but not during website tasks and performance measures to avoid exhaustion caused by a large number of stimuli). Repetition effects might be driven by mere exposure effects or by the participants' attempts to remain consistent in their evaluations. If such effects had been present in our usability measures, they should have increased the correlations. However, because we found no significant correlations between the usability ratings of screenshots and the other usability measures, we conclude that repetition effects are not relevant to the usability measurements in our study.

Third, we used a single item to evaluate subjective website aesthetics. In doing so we attempted to assess a general aesthetics factor. Such single-item measures have also been adopted in prior research on website aesthetics and are common practice in user experience research (e.g., *Hassenzahl, 2004; Sonderegger & Sauer, 2010; Tractinsky, Katz & Ikar, 2000*). Nevertheless, single-item measurement is rightly criticised for its reliability problems, and concerns regarding adequate construct assessment may be raised (*Moshagen & Thielsch, 2010*, p. 692). The overall clear pattern of our results across different time conditions argues that our experimental procedure was appropriate. Still, an experimental validation of our results or their replication using a standardised assessment instrument for aesthetics (such as the instrument of *Lavie & Tractinsky, 2004*, or the VisAWI, *Moshagen & Thielsch, 2010*) would be welcome.

Fourth, all of the tested participants and the stimuli shared a cultural background. There is some empirical evidence for the influence of cultural and ethnic background on perceptions of website aspects such as colour and images (*Cyr et al., 2009; Cyr, Head & Larios, 2010*) or compositional elements (*Bi, Fan & Liu, 2011*). The extent to which our findings are prone to cultural differences should be analysed according to a cross-cultural approach.

## CONCLUSION

To answer our main questions: early and very early ratings of expected usability are not significantly connected to post-use usability measures but highly correlated to perceived aesthetics. Additionally, ratings of expected usability given after presentation of a screenshot for 10 s showed the same result pattern. Hence, such ratings of expected usability are no valid proxy neither for objective usability nor for experienced usability after use.

Some might conclude that our data support the “what is beautiful is usable” perspective on human–computer interaction (cf., *Tractinsky, Katz & Ikar, 2000*). We would like to provide a different viewpoint: first, prior research (*Tuch et al., 2012b*) has shown that

this adage can also be expressed the other way around as “what is usable is beautiful.” Second, even if our approach to measuring subjective usability was biased by aesthetics, we observed no significant correlation to our objective measures of usability. Thus, to obtain a complete perspective on the usability of a given website, a test with interactive tasks on the fully functioning website and goal-oriented methods (both subjective and objective evaluations) with valid measurements is necessary. Short tasks are recommended, even in usability first impression evaluations (e.g., those conducted by *Lee & Koubek, 2010*). If non-interactive screenshots are used in usability tests, there is a likely chance that the results might be driven by website aesthetics rather than usability.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The authors declare there was no funding for this work.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Meinald T. Thielsch conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, performed the computation work, reviewed drafts of the paper.
- Ronja Engel conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, performed the computation work, reviewed drafts of the paper.
- Gerrit Hirschfeld conceived and designed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, performed the computation work, reviewed drafts of the paper.

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.19#supplemental-information>.

## REFERENCES

- Allen M, Currie LM, Bakken S, Patel VL, Cimino JJ. 2006.** Heuristic evaluation of paper-based Web pages: a simplified inspection usability methodology. *Journal of Biomedical Informatics* 39(4):412–423 DOI 10.1016/j.jbi.2005.10.004.
- Bargas-Avila JA, Hornbæk K. 2011.** Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: *Proceedings of the 2011 annual conference on Human factors in computing systems*. New York: ACM, 2689–2698.
- Ben-Bassat T, Meyer J, Tractinsky N. 2006.** Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Transactions on Computer-Human Interaction* 13(2):210–234 DOI 10.1145/1165734.1165737.

- Bi L, Fan X, Liu Y. 2011.** Effects of symmetry and number of compositional elements on chinese users' aesthetic ratings of interfaces: experimental and modeling investigations. *International Journal of Human-Computer Interaction* 27(3):245–259 DOI 10.1080/10447318.2011.537208.
- Breitmeyer BG. 2007.** Visual masking: past accomplishments, present status, future developments. *Advances in Cognitive Psychology* 3(1–2):9–20 DOI 10.2478/v10053-008-0010-7.
- Cyr D, Head M, Larios H. 2010.** Colour appeal in website design within and across cultures: a multi-method evaluation. *International Journal of Human-Computer Studies* 68(1):1–21 DOI 10.1016/j.ijhcs.2009.08.005.
- Cyr D, Head M, Larios H, Pan B. 2009.** Exploring human images in website design: a multi-method approach. *MIS Quarterly* 33:539–566.
- Dahlbäck N, Jönsson A, Ahrenberg L. 1993.** Wizard of Oz studies—why and how. *Knowledge-Based Systems* 6(4):258–266 DOI 10.1016/0950-7051(93)90017-N.
- De Angeli A, Sutcliffe A, Hartmann J. 2006.** Interaction, usability and aesthetics: what influences users' preferences. In: *Proceedings of the 6th ACM conference on designing interactive systems*. New York: ACM Press, 271–280.
- Enns J, Di Lollo V. 2000.** What's new in visual masking? *Trends in Cognitive Sciences* 4(9):345–352 DOI 10.1016/S1364-6613(00)01520-5.
- Flavián C, Guinalú M, Gurrea R. 2006.** The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management* 43:1–14 DOI 10.1016/j.im.2005.01.002.
- Green D, Pearson J. 2006.** Development of a web site usability instrument based on ISO 9241-11. *Journal of Computer Information Systems* (40):66–73.
- Grill-Spector K, Henson R, Martin A. 2006.** Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences* 10(1):14–23 DOI 10.1016/j.tics.2005.11.006.
- Haklay M, Zafiri A. 2008.** Usability engineering for GIS: learning from a screenshot. *The Cartographic Journal* 45(2):87–97 DOI 10.1179/174327708X305085.
- Hartmann J, Sutcliffe A, De Angeli A. 2008.** Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction* 15(4): Article 15 DOI 10.1145/1460355.1460357.
- Hassenzahl M. 2004.** The interplay of beauty, goodness, and usability in interactive products. *Human Computer Interaction* 19:319–349 DOI 10.1207/s15327051hci1904\_2.
- Hassenzahl M, Monk A. 2010.** The inference of perceived usability from beauty. *Human-Computer Interaction* 25:235–260 DOI 10.1080/07370024.2010.500139.
- Hornbæk K. 2006.** Current practice in measuring usability: challenges to usability studies and research. *International Journal of Human-Computer Studies* 64:79–102 DOI 10.1016/j.ijhcs.2005.06.002.
- ISO. 1998.** *ISO 9241: ergonomic requirements for office work with visual display terminals, (VDTS)—part 11: guidance on usability*. Geneva: International Organization for Standardisation.
- Katz A. 2010.** Aesthetics, usefulness and performance in user-search engine interaction. *Journal of Applied Quantitative Methods* 5(3):424–446.
- Kim H, Fesenmaier DR. 2008.** Persuasive design of destination web sites: an analysis of first impression. *Journal of Travel Research* 47(1):3–13 DOI 10.1177/0047287507312405.
- Kurosu M, Kashimura K. 1995.** Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. *Conference companion on Human factors in computing systems*. New York: ACM Press, 292–293.

- Lavie T, Tractinsky N. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies* 60:269–298 DOI 10.1016/j.ijhcs.2003.09.002.
- Leder H, Belke B, Oeberst A, Augustin D. 2004. A model of aesthetic appreciation and aesthetic judgements. *British Journal of Psychology* 95:489–508 DOI 10.1348/0007126042369811.
- Lee S, Koubek RJ. 2010. Understanding user preferences based on usability and aesthetics before and after actual use. *Interacting with Computers* 22:530–543 DOI 10.1016/j.intcom.2010.05.002.
- Lee S, Koubek RJ. 2012. Users' perceptions of usability and aesthetics as criteria of pre- and post-use preferences. *European Journal of Industrial Engineering* 6:87–117 DOI 10.1504/EJIE.2012.044812.
- Lindgaard G, Dudek C, Sen D, Sumegi L, Noonan P. 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction* 18(1):1–30 DOI 10.1145/1959022.1959023.
- Lindgaard G, Fernandes G, Dudek C, Brown J. 2006. Attention web designers: you have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25(2):115–126 DOI 10.1080/01449290500330448.
- Monk A. 2004. The product as fixed effect fallacy. *Human-Computer Interaction* 19(4):371–375 DOI 10.1207/s15327051hci1904\_6.
- Moshagen M, Musch J, Göritz AS. 2009. A blessing, not a curse: experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics* 52(10):1311–1320 DOI 10.1080/00140130903061717.
- Moshagen M, Thielsch MT. 2010. Facets of visual aesthetics. *International Journal of Human-Computer Studies* 68:689–709 DOI 10.1016/j.ijhcs.2010.05.006.
- Moshagen M, Thielsch MT. 2013. A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology* 32(12):1305–1311 DOI 10.1080/0144929X.2012.694910.
- Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88(5):879–903 DOI 10.1037/0021-9010.88.5.879.
- Robins D, Holmes J. 2008. Aesthetics and credibility in web site design. *Information Processing & Management* 44(1):386–399 DOI 10.1016/j.ipm.2007.02.003.
- Schenkman BN, Jönsson FU. 2000. Aesthetics and preferences of web pages. *Behaviour & Information Technology* 19(5):367–377 DOI 10.1080/014492900750000063.
- Sonderegger A, Sauer J. 2010. The influence of design aesthetics in usability testing: effects on user performance and perceived usability. *Applied Ergonomics* 41(3):403–410 DOI 10.1016/j.apergo.2009.09.002.
- Sonderegger A, Zbinden G, Uebelbacher A, Sauer J. 2012. The influence of product aesthetics and usability over the course of time: a longitudinal field experiment. *Ergonomics* 55(7):713–730 DOI 10.1080/00140139.2012.672658.
- Stojmenovic M, Pilgrim C, Lindgaard G. 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*. ACM, 316–323.
- Tarasewicz P, Daniel HZ, Griffin HE. 2001. Aesthetics and web site design. *Quarterly Journal of Electronic Commerce* 2:67–81.
- Thielsch MT. 2008. *Ästhetik von Websites [Aesthetics of websites]*. Münster: MV Wissenschaft.
- Thielsch MT, Blotenberg I, Jaron R. 2014. User evaluation of websites: from first impression to recommendation. *Interacting with Computers* 26(1):89–102 DOI 10.1093/iwc/iwt033.

- Thielsch MT, Hirschfeld G. 2010.** High and low spatial frequencies in website evaluations. *Ergonomics* 53(8):972–978 DOI [10.1080/00140139.2010.489970](https://doi.org/10.1080/00140139.2010.489970).
- Thielsch MT, Hirschfeld G. 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](https://doi.org/10.1080/00140139.2012.665496).
- Tractinsky N. 1997.** Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: *CHI 97 conference proceedings, Atlanta, 22–27 March*. New York: ACM, 115–122.
- Tractinsky N, Cokhavi A, Kirschenbaum M, Sharfi T. 2006.** Evaluating the consistency of immediate aesthetic perceptions of web pages. *Journal of Human Computer Studies* 64:1071–1083 DOI [10.1016/j.ijhcs.2006.06.009](https://doi.org/10.1016/j.ijhcs.2006.06.009).
- Tractinsky N, Katz AS, Ikar D. 2000.** What is beautiful is usable. *Interacting with Computers* 13:127–145 DOI [10.1016/S0953-5438\(00\)00031-X](https://doi.org/10.1016/S0953-5438(00)00031-X).
- Tuch AN, Presslauer EE, Stöcklin M, Opwis K, Bargas-Avila JA. 2012a.** 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](https://doi.org/10.1016/j.ijhcs.2012.06.003).
- Tuch AN, Roth SP, Hornbæk K, Opwis K, Bargas-Avila JA. 2012b.** 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](https://doi.org/10.1016/j.chb.2012.03.024).
- Van Schaik P, Ling J. 2008.** Modelling user experience with web sites: usability, hedonic value, beauty and goodness. *Interacting with Computers* 20(3):419–432 DOI [10.1016/j.intcom.2008.03.001](https://doi.org/10.1016/j.intcom.2008.03.001).