

Please cite this article as:

Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68(10), 689-709. doi:10.1016/j.ijhcs.2010.05.006

Facets of visual aesthetics

Morten Moshagen

University of Mannheim and University of Duesseldorf, Germany

Meinald T. Thielsch

University of Muenster, Germany

Correspondence should be addressed to:

Morten Moshagen

Lehrstuhl Psychologie III

University of Mannheim

Schloss, EO 254

68133 Mannheim

Germany

Phone: +49-621-1812124

Fax: +49-621-1813997

E-Mail: moshagen@uni-mannheim.de

Abstract

Visual aesthetics has been shown to critically affect a variety of constructs such as perceived usability, satisfaction, and pleasure. Given the importance of visual aesthetics in human-computer interaction, it is vital that it is adequately assessed. The present research aimed at providing a precise operational definition and to develop a new measure of perceived visual aesthetics of websites. Construction of the Visual Aesthetics of Website Inventory (VisAWI) was based on a comprehensive and broad definition of visual aesthetics so that the resulting instrument would completely describe the domain of interest. Four interrelated facets of perceived visual aesthetics of websites were identified and validated in a series of seven studies. Simplicity and Diversity have repeatedly been treated as formal parameters of aesthetic objects throughout the history of empirical aesthetics, Colors are a critical property of aesthetic objects, and Craftsmanship addresses the skillful and coherent integration of the relevant design dimensions. These four facets jointly represent perceived visual aesthetics, but are still distinguishable from each other and carry unique meaning. The subscales contained in the VisAWI demonstrate good internal consistencies. Evidence for the convergent, divergent, discriminative, and concurrent validity of the VisAWI is provided. Overall, the present research suggests that the VisAWI appears to be a sound measure of visual aesthetics of websites comprising facets of both practical and theoretical interest.

Keywords: aesthetics; assessment; beauty; design; measurement; website

1. Introduction

The question of what constitutes beauty has been given a variety of answers over the past centuries (e.g., Feagin and Maynard, 1997; Fenner, 1996; Osborne and Balakian, 1968). Many theorists conceived beauty as a property of an object that produces a pleasurable experience in any perceiver. In contrast to this objectivist view, the subjectivist view proposes that anything could be beautiful as long it pleases the senses. Beauty is regarded as a mere function of idiosyncratic qualities of the perceiver, rather than being directly determined from attributes of an object. Most modern philosophical analyses, however, reject the objective versus subjective distinction and adopt an interactionist perspective: Beauty is seen as a function of both, properties of an object and characteristics of the perceiver, that is, beauty emerges from patterns in the way perceivers and objects relate. In line with this interactionist viewpoint, the philosopher George Santayana (1955) describes three defining features of beauty. Beauty is value positive, intrinsic, and objectified. Beauty is value positive, because it provides pleasure. Beauty is intrinsic, because an object is perceived without any reasoning about expected utility. This feature of beauty implies that aesthetic responses occur immediately at first sight, rather than being the result of a long lasting cognitive analysis. Finally, beauty is objectified, because people experience beauty as something that lies in an object, rather than exclusively being the result of a positive sensation of the body. This is not to be confused with an objectivist viewpoint on beauty. Beauty is not objective, but directed toward an object.

We adopt the interactionist perspective by regarding beauty as an immediate pleasurable subjective experience that is directed toward an object and not mediated by intervening reasoning. Because the definition of beauty as being value positive, intrinsic, and objectified closely resembles the definition of aesthetic experience used in empirical aesthetics (Arnheim, 1974; Kawabata and Zeki, 2004; Leder et al., 2004; Martindale and Moore, 1988; Reber et al., 2004; Solso, 2003), we use the words “beauty” and “aesthetics” in the remainder of this paper interchangeably. Finally, it is important to distinguish between aesthetic appraisal and aesthetic judgment. Aesthetic appraisals originate from positively valued immediate subjective experiences as elaborated above. Aesthetic judgments entail references to normative criteria of what is aesthetic and is therefore concerned with the question of what should be considered being aesthetic in a more artistic sense (Leder et al., 2004). Aesthetic appraisal and aesthetic judgment are not necessarily related (Leder et al., 2004; Reber et al., 2004). For example, a perceiver may think that a certain object is an important piece of modern art (aesthetic judgment), but may not positively value its perception (aesthetic appraisal). The present work is only concerned with aesthetic appraisal.

1.1. Empirical studies of visual aesthetics

Early attempts to identify attributes of objects that may critically influence aesthetics were inspired by the objectivist perspective on aesthetics. Accordingly, research was guided by the idea that beauty is a function of certain characteristics of an object. Among those identified features were balance and proportion (Arnheim, 1974; Birkhoff, 1933; Fechner, 1876; Gombrich, 1995), novelty and prototypicality (Hekkert and van Wieringen, 1990; Hekkert et al., 2003), as well as contrast and clarity (Gombrich, 1995; Solso, 2003). Birkhoff (1933), von Ehrenfels (1890), and Eysenck (1941) suggested that aesthetics of an object depends on two factors, namely simplicity and complexity, but disagreed regarding the relationship between these factors. According to Birkhoff, complexity is a deterrent to the aesthetics of an object. Contrastingly, von Ehrenfels and Eysenck argued that simplicity and complexity act multiplicatively to determine the aesthetic appraisal. This latter notion can already be found in

early work by Fechner (1876) according to whom the most central principle of aesthetics is “the right combination of diversity in unity” (p. 39).

These contrasting viewpoints on the relationship between simplicity and complexity can be resolved by considering Berlyne’s (1971) psychobiological theory on aesthetics. Berlyne proposed that aesthetic appraisal is determined by the arousal potential of an object. The psychobiological theory posits that objects with a moderate arousal potential result in the most favorable response in a perceiver, whereas stimuli with either a very low or a very high arousal potential are perceived as being unpleasant. In Berlyne’s theory, arousal is seen as both, an objective property and an idiosyncratic reaction. The objective part stems from the fact that the physical properties of an object are identical for all perceivers. Nevertheless, the same physical characteristic may result in varying arousals in different perceivers. Berlyne, however, was mostly concerned with investigating properties of an object, rather than characteristics of the perceiver. Berlyne argued that the arousal potential of an object is determined by so-called collative variables. Collative variables define the structural properties of the stimulus such as complexity, ambiguity, novelty, and familiarity. Thus, depending on the initial arousal potential of an object, increasing complexity can both increase (in line with Fechner, von Ehrenfels, and Eysenck) and decrease (in line with Birkhoff) aesthetic appraisal.

The Gestalt approach (e.g., Arnheim, 1974; Koffka, 1935) rejected the idea that it suffices to describe elemental parts of an object to describe the whole figure. Instead, it was suggested that Gestalten are perceived first and their constituent parts second. The primary focus of the Gestalt theory, however, lies on perceptual organization rather than on aesthetics. Gestalt theory attempts to explain how stimulus elements are perceived and mentally organized into groups. Various grouping principles (e.g., similarity, closure, and proximity) have been identified that were subsumed under the principle of Prägnanz (Wertheimer, 1923). This general law describes the tendency to perceive the simplest and most stable figure of all possible perceptual alternatives. Stimuli are organized such that regular and simple figures emerge. Moreover, it is assumed that the perception of stimuli following the principle of Prägnanz provides pleasure, whereas the perception of stimuli violating this principle is negatively valued (Hüppe, 1984; Palmer, 1982). Therefore, stimuli in accord with the principle of Prägnanz are considered as good Gestalten that are perceived as being aesthetically pleasant. However, Gestalt theory has been criticized for its rather vague definitions of central concepts such as the notion of good Gestalt (e.g., Palmer, 1982). In an attempt to formalize and quantify figural goodness, it has been suggested in the information theoretic approach to aesthetics (Hochberg and McAlister, 1953; van der Helm and Leeuwenberg, 1996) that good Gestalten are characterized by a limited amount of information (low redundancy). Within this approach, figural goodness is defined as invariance against transformations such as rotations and flipping. It should be noted, however, that this information theoretic approach focuses on physical properties of an object, whereas Gestalt theory was concerned with its psychological, perceptual representation. Identical physical properties of an object may result in different psychological representations and vice versa.

Apart from the question what constitutes figural goodness, it is widely recognized that stimuli that follow the principle of Prägnanz promote efficient and fast perception, whereas stimuli that violate Gestalt principles take longer to perceive and are regarded as being ambiguous (e.g., Hoeger, 1997; Sharps and Nunes, 2002). These fundamental principles of perceptual organization gain importance for psychological aesthetics when considering the processing fluency account (Reber et al., 2004). The processing fluency approach posits that the aesthetic experience is tied to the processing demands of an object. Anything can be beautiful, as long as a given perceiver finds it easy to process. This approach fits particularly well into the interactionist viewpoint, because characteristics of an object and of the perceiver interact in determining aesthetic pleasure. An object by itself has certain characteristics that

can facilitate or impede processing fluency. Those physical properties of an object are governed by Gestalt principles such as amount of information, symmetry, and clarity. The influence of objective properties on aesthetic appraisal, however, is mediated by processing fluency, which also depends on certain characteristics of the perceiver such as its learning history. For example, it has repeatedly been demonstrated that prototypical objects are generally preferred over nonprototypical objects (e.g., Martindale and Moore, 1988; Winkielman et al., 2006) and that repeated exposure to a stimulus results in more favorable evaluations (e.g., Zajonc, 1968). Thus, aesthetic pleasure emerges from the specific processing experience, which itself is a function of the properties of the object and characteristics of the perceiver: The more fluently a perceiver can process an object, the more positive is his or her aesthetic response.

1.2. Visual aesthetics in human computer interaction

Research in human-computer interaction has traditionally focused on effectiveness and efficiency of interactions. In the past decade, it has been increasingly recognized that user's needs go beyond usability and utility, leading to a shift in focus toward a more general experiential perspective (e.g., Hassenzahl and Tractinsky, 2006; International Organization for Standardization, 2009). This perspective takes the whole user experience into account, including emotions and visual aesthetics.

This shift in focus can be dated back to the seminal work by Tractinsky, Katz, and Ikar (2000), who replicated and extended previous results (Kurosu and Kashimura, 1995; Tractinsky, 1997) by showing a reliable association between perceived visual aesthetics and subjective evaluations of usability. Tractinsky et al. (2000) provoked with the conclusion that beautiful designs are usable, which in turn generated much research on the relationship between visual aesthetics and perceived usability (Ben-Bassat et al., 2006; Hartmann et al., 2007; Hassenzahl, 2004; Parizotto-Ribeiro and Hammond, 2005; Roast et al., 2002; Sutcliffe and de Angeli, 2005; Thüring and Mahlke, 2007) with recent results indicating that visual aesthetics may actually be capable of improving performance (Moshagen et al., 2009; Sonderegger and Sauer, 2010). There is also evidence that visual aesthetics can influence perceptions of a system apart from perceived usability, such as evaluations of a website's content (Aladwani and Palvia, 2002; De Angeli et al., 2006; Palmer, 2002; Sutcliffe and de Angeli, 2005; Thielsch, 2008), credibility (Fogg et al., 2003), and trustworthiness (Cyr et al., 2008; Karvonen, 2000; Kim and Moon, 1998). Moreover, aesthetically pleasant websites seem to affect a variety of constructs related to its capability to elicit positive affect (Norman, 2004; Zhang and Li, 2005; Zhou and Fu, 2007), such as satisfaction (Cyr et al., 2008; Lindgaard and Dudek, 2003; Zhang and von Dran, 2000), preference (Cober et al., 2003; Finn and Kayande, 2003; Roberts et al., 2003; Schenkman and Jönsson, 2000; Tarasewich et al., 2001), urge to buy impulsively (Parboteeah et al., 2009), customer loyalty (Chang and Chen, 2009), intention to revisit (Mahlke, 2002; Mundorf et al., 1993; Yoo and Donthu, 2001), fun, joy, and pleasure (Creusen and Snelders, 2002; Jordan, 1998; Mathwick et al., 2001). Finally, evidence suggests beneficial effects of visual aesthetics in the context of learning environments, for example by increasing ease and effectiveness of learning (Grabinger, 1993; Pomales-Garcia et al., 2005; Szabo and Kanuka, 1999), enhancing transfer of information (Aspillaga, 1991), and increasing motivation (Zain et al., 2007).

In light of the manifold pieces of evidence concerning the role of visual aesthetics in human-computer interaction, various models have been developed that attempt to integrate visual aesthetics in more general theoretical frameworks (e.g., Cox and Dale, 2002; Hassenzahl, 2004; Hartmann et al., 2008; Lindgaard and Whitfield, 2004; Liu, 2003a, 2003b; Sutcliffe et al., 2006; Thüring and Mahlke, 2007; Tractinsky and Lowengart, 2007; van der

Heijden, 2003; Zhang and von Dran, 2000) and visual aesthetics has become a research topic in human-computer interaction on its own right. For example, the stability of the evaluation of visual aesthetics over time has been investigated repeatedly (Lindgaard et al., 2006; Tractinsky et al., 2006; van Schaik and Ling, 2009) and there are various experimental studies targeted at identifying characteristics of user-interfaces that influence perceived visual aesthetics. This line of research examined various factors coinciding with characteristics of aesthetic objects identified in studies on empirical aesthetics as reviewed above. Specifically, properties influencing aesthetic appraisal of user-interfaces were shown to include colors (Cyr et al., 2010; Hall and Hanna, 2004; Ling and van Schaik, 2002; Shieh and Lin, 2000; Simon, 2001; Thielsch, 2008), balance, grouping, structure, and order (Bauerly and Liu, 2006, 2008; Brady and Phillips, 2003; Ling and van Schaik, 2002, 2004; Schenkman and Jönsson, 2000; Schmidt et al., 2009; Schrepp et al., 2006; Zheng et al., 2009), simplicity and density (Ngo et al., 2000, 2003), novelty and prototypicality (Haig and Whitfield, 2001) as well as complexity (Pandir and Knight, 2006; Tuch et al., 2009). A summary of the factors that were shown to critically affect interface aesthetics is shown in Table 1.

1.3. Assessment of visual website aesthetics

Given the importance of visual aesthetics in human-computer interaction, it is vital that it is adequately assessed. Although, in principle, there are several ways to measure visual aesthetics (including physiological measures, Kawabata and Zeki, 2004, Thüring and Mahlke, 2007, formalized mathematical assessment, Ngo et al., 2000, 2003, Papachristos et al., 2006, and paired-comparison data, Schenkman and Jönsson, 2000), existing studies mostly used one of the following self-report measures: Single-item measures, ad-hoc developed scales, single scales purportedly measuring aesthetics taken from more general instruments for the evaluation of websites on various dimensions, or instruments specifically developed to measure visual website aesthetics.

Single-item measures are commonly used to assess visual aesthetics (e.g., Hassenzahl, 2004; Sonderegger and Sauer, 2010; Tractinsky et al., 2000). However, the use of single-item measures is generally discouraged for various reasons (Anastasi and Urbana, 1997; Kline, 2000; Nunnally and Bernstein, 1994). First, single-item measures are not as reliable, since the measurement error is not averaged out as in multiple-item scales (Schmidt and Hunter, 1996; Spector, 1992). Second, single-item measures are inadequate to capture complexities of a construct (Baumgartner and Homburg, 1996). Third, single-item measures are not able to finely discriminate between the degrees of visual aesthetics, but only allow for a rather coarse assessment (Dollinger and Malmquist, 2009; Spector, 1992). Finally, single-item measures are subject to acquiescence bias that cannot be controlled by mixing positively and negatively keyed items (Motl and DiStefano, 2002). In addition to these more general problems, there are further problems of single-item measures specifically in the domain of visual website aesthetics. Single-item measures assume the measured construct being one-dimensional, but this assumption has rarely been empirically tested. Available evidence, however, indicates that visual aesthetics seems to be a multidimensional construct (Lavie and Tractinsky, 2004). Moreover, the exact item wording used in different studies was rarely the same. For example, some researchers asked participants to rate attractiveness, others to state whether a layout was beautiful (vs. ugly), still others to rate the appeal, and yet others to state whether the layout was aesthetically pleasing. The use of different item wordings (presented in different languages) makes it impossible to judge whether these items in fact tap the same construct, in turn, scrutinizing comparability of results across different studies. For all of these reasons, it is clear that standardized multiple-item scales for the assessment of visual aesthetics are to be preferred.

Considering multiple-item measures of visual aesthetics, various studies used scales that were developed on an ad-hoc basis (e.g., Cyr et al., 2010; De Wulf et al., 2006; Hartmann et al., 2007; Hong and Kim, 2004; Moshagen et al., 2009; Thüning and Mahlke, 2007; van der Heijden, 2003) or single scales from more general instruments aimed at evaluating website as a whole (e.g., Cober et al., 2003; Kirakowski and Claridge, 2001; Lindgaard and Dudek, 2003; Finn and Kayande, 2003; Loiacono et al., 2002; Mathwick et al., 2001; van Schaik and Ling, 2005). Both approaches are subject to various drawbacks. First, it is again assumed (but rarely validated) that visual aesthetics is a one-dimensional construct. Second, the psychometric properties of ad-hoc scales are largely unknown. Due to the lack of standardization, ad-hoc scales compromise comparability of the results across different studies. Finally, neither of these scales has been properly validated. For example, the attractiveness scale of the Web Analysis and Measurement Inventory (Kirakowski and Claridge, 2001; Lindgaard and Dudek, 2003) is clearly confounded with usability and content. Similarly, a scale used by van Schaik and Ling (2005) contains the items “comprehension”, “legibility”, and “meaningfulness”, which again rather measure usability than visual aesthetics. Some instruments (Finn and Kayande, 2003; Mathwick et al., 2001) contain scales that attempt to measure aesthetics by few globally formulated items. For example, the aesthetics scale of the WebQual (Loiacono et al., 2002, 2007) is comprised of the items “visually pleasing“ and „visually appealing“. Scales such as these are essentially single-item measures that introduce systematic biases because the same question is asked more than once (Robins et al., 2001; Rossiter, 2002). Still other instruments are based on descriptive rather than evaluative items such as “The site used color” (Cober et al., 2003). However, the use of color could be both, aesthetically pleasant and very ugly. These examples show that validation is a vital ingredient in scale development, but has often been neglected in studies relying on both, ad-hoc scales and single scales from more general instruments.

Given these deficiencies, it is desirable to have instruments available which were developed specifically targeting at visual aesthetics and have been undergone proper validation procedures. Currently, there is only one instrument available that satisfies these requirements (Lavie and Tractinsky, 2004). Starting with a set of 44 items, Lavie and Tractinsky (2004) identified two dimensions of visual website aesthetics in a series of four studies. These dimensions were labeled classic and expressive aesthetics. *Classic aesthetics* refers to orderliness in design and comprises the items “aesthetic”, “pleasant”, “symmetric”, “clear”, and “clean”, whereas *expressive aesthetics* reflects the perception of the designers’ creativity and originality and comprises the items “creative”, “fascinating”, “original”, “sophisticated”, and “uses special effects”. The two-factor structure was obtained and validated through a series of exploratory and confirmatory factor analyses. Convergent and divergent validity was established by demonstrating high correlations to a measure of pleasure and moderate correlations to measures of usability and service quality, respectively.

Although the measure of classic and expressive aesthetics is a good first step towards operationalizing visual aesthetics and providing a sound assessment instrument, several issues regarding psychometric properties, utility, and validity remain. First, the expressive aesthetics scale contains the purely descriptive item “uses special effects” and the classic aesthetics scale contains the item “symmetrical”. However, neither special effects nor symmetry are necessarily aesthetically per se. Second, it is difficult to explain why the item “aesthetic” only loads on the classic aesthetic scale, as both scales should represent dimensions of visual aesthetics (see also, Lindgaard et al., 2006). Third, from an applied, design-oriented perspective, the items and scales are too abstract to allow for optimizing the design of a website. For example, if a particular website received low ratings on the items “pleasant” and “aesthetic”, it is unclear for a designer how to change the layout of the website to enhance classic aesthetics. Finally, it is questionable whether the scales represent all relevant aspects

of visual website aesthetics. Specifically, not a single item refers to colors, although the importance of colors is well recognized in empirical aesthetics in general (Arnheim, 1974; Kawabata and Zeki, 2004; Martindale and Moore, 1988; Ramachandran and Hirstein, 1999) and in human-computer interaction in particular (e.g., Cyr et al., 2010; Hall and Hanna, 2004; Hoadley, 1990; Moshagen et al., 2009; Mundorf et al., 1993; Thielsch, 2008).

1.4. The present research

On the basis of an apparent need for a well-designed measure and a precise operational definition, the present research aimed at developing a measure of perceived visual aesthetics of websites. The remainder of this paper describes a series of seven studies directed at the development and validation of the Visual Aesthetics of Website Inventory (VisAWI). To construct the VisAWI, we relied on interviews of experts and end-users, but also considered factors identified in the growing body of literature on attributes of interfaces that critically affect perceived visual aesthetics. We made an effort to be as inclusive as possible in defining the domain of visual website aesthetics, so that the questionnaire would be an appropriate tool for both, research and applied settings. In what follows, we first describe four studies on the construction of the VisAWI and then three studies that aim at demonstrating various types of validity.

2. Scale Construction

This section comprises four studies on the construction of the VisAWI. Study 1 presents the definition of the domain and describes the construction of an initial item pool. Study 2 identifies the factor structure using exploratory factor analysis. In Study 3, the items are refined and further selected to purify the scales. Study 4 provides the factor structure and the basic psychometric properties of the final version.

2.1 Study 1: Defining the Domain and Item Generation

The purpose of this study was to define the domain of the questionnaire, to identify core aspects, and to generate items.

2.1.1. Defining the domain

The first stage of test development was to define the domain of visual website aesthetics and identify the core dimensions of this construct. The theoretical and empirical literature on visual aesthetics and design of websites was reviewed to understand how aesthetics has been conceptualized and operationalized across disciplines. The aim was to sample all relevant aspects of visual website aesthetics, ideally without reference to other dimensions such as content, trust, and usability. This comprehensive review of the literature culminated in the identification of 34 aspects describing visual aesthetics of websites.

2.1.2 Empirical evaluation of the identified aspects

The aspects obtained in the literature review were empirically evaluated to be as inclusive as possible to comprehensively define the construct of visual aesthetics of websites. This was done by questioning web users and experts in the domain of information technology.

First, the aspects were rated by a convenience sample of 256 web users in terms of perceived importance to visual aesthetics of websites on a five-point Likert scale ranging from 1 ('not important at all') to 5 ('very important'). Participants were also asked to generate additional aspects of visual website aesthetics; however, participants did not state aspects that were not already included in the pool. Results indicated that simplicity ($M = 4.72$), colors ($M = 4.19$), and proportion ($M = 4.01$) were judged to be most important, whereas depth ($M = 2.37$), symmetry ($M = 2.39$), and movement ($M = 2.48$) were regarded least important. Thus, whereas the mean importance ratings showed that no aspect was judged to be truly unimportant, there was also evidence for a hierarchy between the aspects. However, this study is subject to some drawbacks. Ordinary users may not be capable to evaluate the aspects according to the same definitions. Also, it might be difficult to distinguish between ratings of importance and ratings of liking. For example, the low importance ratings for movement may rather reflect that movement is negatively valued. Thus, expert interviews were conducted to more comprehensively evaluate the identified aspects.

Additional data were obtained from interviews with web designers, artists, and psychologists with a strong background in human-computer interaction. Those 10 experts were individually asked to rank a set of 20 websites according to aesthetic preference and to group the websites by similarity. The think-aloud technique was used to gain insights into relevant decision criteria for judgments of similarity and preference, respectively. After performing these tasks, the experts took part in an interview to test the content validity of the identified aspects. The preference sorting task revealed that the websites that were judged by the experts to be most aesthetically pleasant were characterized by color harmony, creativity, overview, simplicity, unity, and variety. Aesthetically distorted websites were not only characterized by the absence of those positive attributes, but also by lack of professionalism and style, poor grouping and structure, and lack of homogeneity. The sorting tasks as well as the interviews suggested including eight additional aspects in the pool, leading to an intermediate set of 42 aspects describing visual aesthetics of websites.

Closer examination of these 42 aspects showed that two aspects ('sequence' and 'repetition') specifically refer to the experience of browsing multiple individual sites. Given that the present work aimed at providing an assessment instrument that can be used for individual sites, we opted to discard those aspects. The remaining 40 aspects referring to individual sites are shown in Table 1. Several aspects that showed a great deal of content overlap were collapsed for the purpose of obtaining abstract content domains suited for item generation. For example, homogeneity, regularity, uniformity, and unity were merged into one category. We finally arrived at twelve broad content domains describing visual aesthetics of websites, which are shown in Table 1.

2.1.3 Item generation

The next step in scale development consisted of developing items that represent the identified aspects underlying the aesthetic quality of websites. Although the aim was to generate a sufficient number of items for each aspect, the importance ratings obtained from the expert interviews and the users' ratings were used to prioritize certain aspects. For example, more items were developed reflecting the color aspect, given that both user's ratings and previous research suggest that color is one of the most important properties of websites. Following recommendations in the literature (Anastasi and Urbana, 1997; Kline, 2000; Spector, 1992), items were designed to be concise, unambiguous, unipolar, easily understandable, and avoid colloquialisms and professional jargon. Moreover, items were required to be evaluative rather than descriptive and to be clearly related to only one aesthetic aspect while minimizing

content overlap with other aspects and other dimensions apart from visual aesthetics, respectively. Initially, 96 items reflecting the twelve content domains were generated.

This initial item set was independently rated for quality by four experts in the domain of human-computer interaction and three laypeople. Raters were instructed to evaluate each item for clarity and relevance for the intended aspect. Thirteen items that received low quality ratings were discarded due to lack of clarity and irrelevance. Several items that captured core aspects of visual aesthetics (i.e., received high relevance ratings), but were too complex or poorly worded (i.e., received low clarity ratings), were revised. Two items were dropped due to redundancy. Three additional items were developed such that all aspects were well represented in the item pool. The final item pool comprised 84 items in total. Each content domain was represented by at least five items. The exact number of items reflecting each content domain is shown in the final column of Table 1. In order to counter the tendency to agree with all questions (acquiescence bias), approximately half of the 84 items were negatively keyed (e.g. Motl and DiStefano, 2002).

2.2. Study 2: Explorative Factor Analysis

The second stage of scale development was to determine the factor structure of the item pool and to construct initial scales. Exploratory factor analysis was used to identify the underlying dimensionality of the item pool and to determine which items to retain or discard on the basis of their factor loadings.

2.2.1. Methods

2.2.1.1. Participants

A total of 300 volunteers (56% male) participated in the study. Age ranged from 15 to 64 years ($M = 27.80$; $SD = 9.69$). Participants were excluded from data analysis if they reported knowledge of the Finnish language (see below). Participants completed the study on an anonymous basis and did not receive any compensation for completing the study.

2.2.1.2. Procedures

After providing demographic background information, participants entered a screen that was split into two parts. The website to be evaluated was displayed in the lower panel. The upper panel randomly showed one item at a time. Thus, participants simultaneously viewed both, the website and the items. Completing the study took about 20 minutes.

2.2.1.3. Materials

In order to minimize the effect of other variables (e.g., usability and content) on evaluations of visual aesthetics, websites in Finnish language (thus, incomprehensible for German participants) were used. Moreover, hyperlinks on each site were deactivated such that they were still displayed as hyperlinks and could be selected, but did not lead to another site. These measures precluded confounds of content and usability. An initial pool of 100 Finnish websites was prepared based on the following criteria: (1) fast loading times, (2) (largely) browser invariant, (3) no hints to website content. Each website from this initial pool was rated by four experts in the domain of information technology on eight descriptive criteria (on an eight-point bipolar scale ranging from -4 to +4): Lots of pictures – no pictures, lots of text – no text, lots of elements – few elements, empty – crowded, professional – amateurish, monochrome – colorful, static – dynamic. A cluster-analysis on the expert's ratings yielded five clusters. Websites for the present study were selected such that each cluster was represented by at least

one website. Two websites were selected each from the largest clusters and one website was selected from each of the three smaller clusters. Hence, the stimulus materials for the present study consisted of seven websites in Finnish language.

2.2.1.4. Measures

The initial version of the VisAWI comprised 84 items from the item generation phase. Participants were asked to indicate their level of agreement to each item on a 7-point Likert scale ranging from 1 ('strongly disagree') to 7 ('strongly agree'). This format was chosen in order to capture response variability and because of its quasi-interval nature. Moreover, in this first empirical evaluation of the items, an additional response option was provided to indicate that an item was considered problematic by participants (e.g., difficult to understand, ambiguous, not applicable to a website, etc.).

2.2.2. Results and Discussion

Prior to subjecting the items to exploratory factor analysis, seven items were dropped that exhibited excessive skew or kurtosis, very high ($p_s > .85$) or very low ($p_s < .15$) item difficulties, or were considered being problematic by more than 5% of participants. In the context of Likert-type scales, item difficulties are defined as the sum of the observed ratings in relation to the maximum possible rating (e.g., Goldstein and Hersen, 2000). Items with very high or very low difficulties have little power in discriminating between visual aesthetics and were therefore removed.

The remaining 77 Items were subjected to a maximum likelihood factor analysis with oblimin rotation using the computer program Factor (Lorenzo-Silva and Ferrando, 2006). We deliberately opted for an oblique rotation procedure as it was expected that the factors represent interrelated facets that jointly reflect the domain of visual website aesthetics. Fourteen factors with an eigenvalue greater than one emerged. Given that applying the Kaiser-Guttman criterion to decide on the number of factors to retain often results in overextraction of factors (Cliff, 1988; Gorsuch, 1997), we also considered the Scree plot and results from parallel analysis (Horn, 1965). Parallel analysis in 1,000 random data sets suggested a five-factor solution. At root six, the actual eigenvalue (1.78) fell slightly below the mean random data eigenvalue (1.82). However, the difference between the random and the actual eigenvalue at root six was relatively small. Because it is generally recommended to rather extract too many than too few factors (Fava and Velicer, 1992), we comparatively evaluated four-, five-, six- and seven-factor solutions. The emerging factors from four to six factors lead to increasingly cleaner and more interpretable solutions. The addition of a seventh factor did not enhance the overall solution beyond six factors. Thus, the six factor solution was considered the best representation of the items on both theoretical and statistical grounds. The six factors jointly explained 53% of the variance. Factor intercorrelations ranged from .14 - .46. The highest correlations were evident between factor I and VI ($r = .46$) and between factor IV and V ($r = .45$). Table 2 shows a summary of the analysis.

Whereas factors II and III represented items related to Craftsmanship and Colorfulness, respectively, interpretation of the remaining factors was more challenging. Two factors tapped items related to balance, density, consistency, order, and homogeneity, with factor I placing a stronger emphasis on clarity and factor VI on unity. Factors IV and V comprised items related to creativity, diversity, movement, and elements of surprise. We tentatively concluded that factor IV rather reflected novelty, whereas the factor V was more likely to reflect variety.

In order to improve fidelity with the constructs and to purify and shorten the scales, the total number of items was reduced from 77 to 37 by dropping items on the basis of their

primary and secondary factor loadings. An item was selected for further consideration if its primary loading exceeded .4 and there were no substantial ($\lambda \geq .2$) secondary loadings (cf., Floyd & Widaman, 1995). Moreover, four items were revised and two new items were added, leading to an intermediate set of 39 items.

2.3. Study 3: Scale Refinement

The aim of the third study was to replicate the factorial structure and to further refine and shorten the scales using confirmatory factor analysis methodology.

2.3.1 Methods

2.3.1.1. Participants and procedures

This study is based on a heterogeneous convenience sample of 506 volunteers (52.6% male) who completed the study on an anonymous basis. Age ranged from 13 to 63 years with a mean of 27.56 (SD = 8.11) years. Participants did not receive any compensation for completing the study. The procedures were identical to study 2. Participation in the study took approximately 10 minutes.

2.3.1.2. Materials

Similar to the previous study, websites in Finnish language that were not navigable were used to preclude confounds of content and usability. Twenty websites were selected as stimulus materials for the present study from the pool of 100 websites also used in study 2. Selection of websites was based on the results from the cluster analysis described in study 2, with the restriction that no website from the previous study was used again.

2.3.1.3. Measures

The VisAWI version administered in this study consisted of 39 items in total. A seven point Likert scale was given to indicate agreement to each item. In contrast to the previous study, however, there was no response option available to indicate whether an item was considered in some way problematic.

2.3.2. Results and Discussion

Confirmatory factor analysis was conducted on the item covariance matrix using Mplus 4.20 (Muthén and Muthén, 2008) to determine whether exploratory results could be replicated. Factors were allowed to correlate and each item was constrained to load on only one factor. Since cross-loadings were not permitted, a perfect simple structure was imposed on the data. Each factor was assigned a scale by fixing one unstandardized loading to one. Because there were substantial departures from multivariate normality, a robust maximum likelihood estimation procedure (Satorra and Bentler, 2001) was employed for parameter estimation of the factor model. Since the genuine log-likelihood ratio test of model fit has excessive power when the sample is large and rejects models on the basis of trivial misspecifications (e.g., Bollen, 1989; Kline, 2005), model fit was evaluated by means of descriptive indices of model fit (CFI, RMSEA, and SRMR). For the CFI, values greater than 0.95 constitute good fit and values above 0.90 are seen as indicative of acceptable fit to the data (Bentler and Bonett, 1980; Bollen, 1989; Hu and Bentler, 1999). A RMSEA close to 0.05 or below suggest a good fit to the data and values up to 0.08 indicate a reasonable error of approximation (Browne and

Cudeck, 1993; MacCallum et al., 1996; Steiger, 1990). A SRMR up to 0.05 indicates a good fit and between 0.05 and 0.10 an acceptable fit to the data (Hu and Bentler, 1999).

Confirmatory factor analysis of the 39 items representing the six scales as identified in the explorative factor analysis did not yield a good fit to the data (Table 3). Model misfit may be due to several reasons. For example, the factor structure imposed on the data may be suboptimal, such that a different number of factors would better describe the data. Furthermore, items may have substantial loadings on other than their designated factor. Finally, some items may have a great deal of content overlap which is not sufficiently captured by a common factor.

Considering the adequacy of the factor structure, we evaluated the patterns of loadings and the factor intercorrelations.¹ As expected, correlations between the factors were moderate to high, reflecting the notion that they represent facets of a common construct (cf., Borsboom et al., 2004; Gorsuch, 1997). However, very high correlations were evident between the factors initially labeled novelty and variety ($r = .97$) as well as between the factors representing clarity and unity ($r = .88$), indicating that these scales substantially overlapped. According to Kline (2005), poor discriminant validity (that is, unreasonable high correlations) may indicate that a model has too many factors. Although these factors were statistically discernable, the high subscale intercorrelations suggest that the degree of practical distinctiveness among the scales may be quite modest. Accordingly, it was decided to collapse these factors. Moreover, given that the remaining four factors still exhibited at least moderate intercorrelations ($.60 < r < .74$), we imposed a general second order factor reflecting the domain of visual website aesthetics. Thus, the final model consisted of a general second order factor comprised of four factors representing different facets of visual aesthetics of websites (Figure 1).

The first lower order factor was labeled *Simplicity*, reflecting aspects that facilitate perception and the processing of a layout, such as clarity, orderliness, homogeneity, grouping, and balance. The second lower order factor was labeled *Diversity*, reflecting visual richness, dynamics, variety, creativity, and novelty. The third lower order factor comprised items related to the evaluation of individual colors and their composition and was therefore labeled *Colorfulness*. The final lower order factor was labeled *Craftsmanship*, reflecting whether the site was designed with skill and care using modern technologies.

Post-hoc modifications were performed in an attempt to purify the scales and to improve model fit. Inspection of the loadings, residuals, and modification indices associated with the modified model indicated that model fit could be substantially improved by dropping several items. Five items were removed due to low primary loadings and nine items were discarded due to substantial secondary loadings. Moreover, seven items showed multiple large standardized residuals and were therefore dropped from the model. The remaining 18 items were again subjected to confirmatory factor analysis. Results indicated that model fit substantially improved as compared to the first CFA (Table 3). In this model, each primary loading exceeded .60 and there were no substantial secondary loadings. The factor loadings of the items on the VisAWI facets are shown in Table 4. Figure 1 shows the second order loadings of the facets on the general factor. Each scale contained at least two negatively keyed items to counter the acquiescence response bias. Given that modification indices did not indicate further areas of localized strain and given that similar goodness-of-fit statistics have been reported in related work (Lavie and Tractinsky, 2004), this model was deemed to be the final measurement model for the VisAWI.

It is important to note that this study used confirmatory factor analysis methodology in an explorative manner. Given that an increased number of model modifications can result in a sample-specific model that is not replicable (capitalizing on chance; MacCallum et al., 1992),

the modified measurement model of the VisAWI was validated in an independent sample of participants evaluating a different set of websites in study 4.

2.4. Study 4: Confirmatory Factor Analysis

The purpose of the fourth study was to replicate the factor structure obtained through explorative analyses in studies 2 and 3 using confirmatory factor analysis in an independent sample. Furthermore, we wanted to go beyond the rather artificial materials used in previous studies by employing a large number of natural occurring websites in a language comprehensible for the participants.

2.4.1. Methods

2.4.1.1. Participants

A total of 512 volunteers participated anonymously in this study. Of the participants, 347 (67.8%) were female. Age ranged from 15 to 82 years ($M = 30.50$; $SD = 10.61$). Participants received no compensation for completing the study.

2.4.1.2. Procedures

Upon entering the website, participants were requested to provide demographic background information. Next, they completed the VisAWI and several questionnaires not pertinent to the present study with respect to a randomly drawn website. Questionnaires included a measure of classic and expressive visual aesthetics (10 items; Lavie & Tractinsky, 2004), two scales of the AttracDiff questionnaire (14 items; Hassenzahl et al., 2003), a scale measuring perceived website usability (7 items; Flavián et al., 2006), a measure of quality of content (9 items; Thielsch, 2008), and a scale assessing the intention to revisit the site (4 items). These measures are analyzed and described in greater detail in study 5. Presentation of the items and the websites were identical to the previous studies. The questionnaires took approximately 15 minutes to complete.

2.4.1.3. Materials

Forty-two websites from nine different content domains served as stimulus materials. Websites were selected to represent a wide range of institutional and corporate websites in Germany, including corporate websites, e-commerce, e-learning, e-recruitment, entertainment, information sites, search engines, social software, and web portals. Readers are referred to Thielsch and Hirschfeld (in press) for a more detailed description of this categorization scheme for websites. Although the goal was to represent each content domain by five typical sites (as indexed by high Google ratings), only two search engines could be included due to technical difficulties in displaying them in the experimental environment.

2.4.1.4. Measures

The remaining 18 items identified in study 3 were used to define the final version of the VisAWI.

2.4.2. Results and Discussion

The procedures to estimate the confirmatory factor model and to evaluate goodness-of-fit were identical to the ones employed in study 3. Confirmatory factor analysis of the 18 items representing the four subscales of the VisAWI and the second order factor yielded an

acceptable fit to the data (Table 3). Thus, the results suggest that the measurement model involving a general second order factor and four facets adequately describes the factorial structure underlying the items of the VisAWI. The item wordings and the loadings on their designated factor are shown in Table 4. Loading magnitudes ranged from .63 to .89 and there were no substantial secondary loadings. Figure 1 shows the structural part of the model along with the loadings of the four facets on the second order general factor. The general factor is reflected equally strong by each facet, indicating that the facets are of similar importance for general visual website aesthetics.

Internal consistency of the scales was examined using coefficient α . As shown in Table 5, the Cronbach's α estimates for the subscales were good with α ranging from .85 to .89. The Cronbach's α estimate for the total score was excellent ($\alpha = .94$). The scale means (averaged across participants and websites) lie close to the neutral point on the response scale indicating that the scales are neither too simple nor too difficult

3. Scale Validation

This section presents three studies demonstrating various types of validity of the VisAWI. Study 5 contains the convergent, divergent, concurrent, and discriminative validity of the VisAWI. Study 6 provides further evidence concerning divergent and construct validity using an experimental approach. Study 7 presents an additional experiment supporting the construct validity of the VisAWI.

3.1. Study 5: Construct Validation

The purpose of this study was to establish the convergent, divergent, discriminative, and concurrent validity of the VisAWI. Convergent validity is the extent of agreement of various measures of theoretically highly related constructs, whereas divergent validity (also called discriminant validity) refers to the degree of disagreement of theoretically unrelated (or theoretically less related) constructs (e.g., Nunnally and Bernstein, 1994). Evidence for convergent and divergent validity is demonstrated by high correlations with convergent constructs and lower correlations with divergent constructs, respectively. Convergent validity was examined by exploring the relationship of the VisAWI with a measure of visual aesthetics and a measure of overall appeal. Divergent validity was explored by correlating the VisAWI to measures of perceived usability, pragmatic quality, and quality of the content of a website. Discriminative validity refers to the ability of an assessment instrument to discriminate between groups or individuals (e.g., Haynes and O'Brien, 2000) and is often considered an aspect of criterion-related validity (e.g., Messick, 1980). Concerning the VisAWI, discriminative validity refers to its utility in distinguishing between different websites. Ideally, websites perceived to be least and most aesthetic should receive very different scores on the VisAWI. Conversely, if the scores on the VisAWI for aesthetically pleasant and aesthetically unpleasant websites lie close to each other, this would be evidence for poor discriminative validity. Finally, concurrent validity refers to the ability of a measure to predict a concurrently assessed criterion (e.g., Anastasi and Urbana, 1997). Concurrent validity was examined by correlating the VisAWI to participants' intention to revisit the website.

3.1.1. Methods

3.1.1.1. Participants and Materials

This study is based on the same sample as described in study 4 that evaluated the same set of websites as in study 4.

3.1.1.2. Procedures

After providing demographical background information, participants were randomly presented one of the 42 websites along with the questionnaires. The questionnaires were given in randomized order, as were the items within each questionnaire. The items of the VisAWI were presented in a completely randomized order (i.e., randomized across facets). Except for the items of the AttracDiff questionnaire (see below), participants were asked to indicate their level of agreement to each item on a 7-point Likert scale (1 ‘strongly disagree’, 7 ‘strongly agree’).

3.1.1.3. Measures

VisAWI. The final 18-item version of the VisAWI was used.

Classic and expressive visual aesthetics. The instrument by Lavie and Tractinsky (2004) comprises two five-item scales reflecting classic and expressive aesthetics of websites. Classic aesthetics taps items related to clarity and orderliness. Expressive aesthetics is conceived to represent visual richness and diversity of a website. Accordingly, these scales were used as convergent measures.

AttracDiff. The AttracDiff questionnaire (Hassenzahl, 2004; Hassenzahl et al., 2003) is a semantic differential consisting of 28 adjective pairs that aim at measuring pragmatic quality, identification, stimulation, and appeal of interactive products. For the purpose of convergent and divergent validation of the VisAWI, the scales measuring pragmatic quality and appeal were used in the present study. Pragmatic quality is measured through 7 adjective pairs and refers to features of websites allowing for efficient and effective goal achievement. Thus, pragmatic quality can be used to assess perceived usability and utility as divergent constructs. Another seven adjective pairs are used to measure appeal. Appeal is considered to represent the overall attractiveness of a website in a holistic way and, therefore, should be positively correlated with the VisAWI.

Perceived website usability. The scale measuring perceived website usability (Flavián et al., 2006) is a seven-item measure assessing perceived ease of use, ease of understanding and speed of information retrieval.

Quality of content. This questionnaire (Thielsch, 2008) consists of nine items on three subscales representing a general factor ‘quality of content’. The items assess interestingness, comprehensibility, and perceived usefulness of content.

Intention to revisit. Four items were used to assess participants’ intention to revisit the website being evaluated: (1) “I will visit the website again”, (2) “I will visit the website on a regular basis”, (3) “I would recommend the website to my friends”, (4) “If I had interest in the content of the website in future, I would consider visiting the website”. Responses to these four items were averaged to form a composite index of participants’ intention to revisit the website. Intention to revisit was used as a criterion to assess concurrent validity.

3.1.2. Results

The correlation coefficients between the VisAWI and the convergent and divergent constructs are shown in Table 6. The correlations to the convergent constructs (classic aesthetics,

expressive aesthetics, and overall appeal) were generally high, indicating good convergent validity. Moreover, classic aesthetics was highly correlated with the Simplicity facet, but showed less strong relationships to the remaining facets. In contrast to this, expressive aesthetics was highly related to the Diversity facet, but not to the other facets. Hence, convergence validity could be successfully demonstrated. The same is only partly true considering divergent validity, however. Although the correlations to pragmatic quality, perceived usability, and quality of content were lower as compared to the convergent constructs, the VisAWI facets exhibited at least moderate relationships with these constructs. Finally, positive correlations between the VisAWI and intention to revisit the website emerged, providing evidence for the concurrent validity of the VisAWI.

The utility of the VisAWI in distinguishing between different websites (discriminative validity) was examined in two ways. First, a MANOVA with the evaluated website as independent variable and the VisAWI facets and the VisAWI total score as dependent variables was computed to examine whether the VisAWI scores differ as a function of website. Thus, the independent variable (website) had 42 levels and there were 5 dependent variables (the sum scores of the 4 VisAWI facets and the VisAWI total score). Sum scores for the VisAWI facets were obtained by summing participants' responses to the items of the respective facet. The VisAWI general factor was determined by computing the total score of all items of the VisAWI. The MANOVA was highly significant [$F(205, 2322) = 2.19; p < .01$] indicating that the websites received different ratings on the VisAWI. Post-hoc univariate ANOVAs with website as independent variable and each VisAWI facet and the VisAWI total score as dependent variables were conducted to qualify the multivariate effect. Results indicated that there were significant differences between the websites on each facet and also with respect to the total score (see final column of Table 5) suggesting that each VisAWI facet as well as the total score reflect differences in visual website aesthetics.

One might object that a statistically significant difference does not necessarily need to be meaningful in practice. Statistical significance is a function of the alpha error, the sample size, and the magnitude of an effect (Cohen, 1988). As the sample size grows large, very small effects that are meaningless in practice may become statistically significant (complementary, large differences may fail to reach statistical significance when the sample size is small). Thus, it is generally recommended to quantify the difference between two groups by effect sizes (e.g., Wilkinson et al, 1999). Accordingly, we also computed the standardized mean difference (Cohen's d) for the websites evaluated least and most aesthetic on each facet and the total score, respectively (see Table 5). According to guidelines provided by Cohen (1988), standardized mean differences of 0.2, 0.5, and 0.8 are considered small, medium, and large effects, respectively. The standardized mean differences of the websites evaluated least and most aesthetic were close to or larger than $d = 2.0$, showing that the least and most aesthetic websites differ by about two standard deviations on each facet and also with respect to the total score, which can be considered as a very large effect. Thus, it can be concluded that the VisAWI performs very well in discriminating between different websites.

3.1.3. Discussion

In line with our hypothesis concerning convergent validity, high correlations to classic and expressive aesthetics as well as overall appeal were observed. Convergent validity was further supported by strong correlations of the Simplicity facet with the conceptually related classic aesthetics scale, whereas the Diversity facet was highly correlated with the scale measuring expressive aesthetics. Thus, these results provide strong support for convergent validity of the VisAWI. Moreover, discriminative validity of the VisAWI could be established by large differences of the websites on each facet and on the total score, indicating that the VisAWI

performs very well in discriminating between different websites. Finally, the VisAWI exhibited positive correlations of a moderate size to participants' intention to revisit the website being evaluated. Clearly, visual aesthetics is only one variable that affects the intention to revisit a website (e.g., van der Heijden, 2003) and content is probably the most important factor (Agarwal and Venkatesh, 2003). Thus, this result can be seen as indicative for concurrent validity.

However, the results were rather mixed as far as divergent validity is concerned. Although it is not surprising that the Simplicity facet is related to measures of perceived usability and pragmatic quality (Karvonen, 2000; Lavie and Tractinsky, 2004), the remaining facets also showed at least moderate relationships to perceived usability, pragmatic quality, and quality of content. We believe that there are two reasons that may explain these findings. It is important to note that naturally occurring websites were used as stimulus materials, that is, we had no control regarding the properties of the websites. It is reasonable to assume that web designers strive to optimize not only visual aesthetics, but also content and usability. Good designers will perhaps create aesthetically pleasant and usable website that provide an interesting content, whereas not so good designers may fail on all three aspects. As a consequence, evaluation of aesthetics, content, and usability might be correlated due to incomplete experimental control of the stimulus materials used, rather than due to a conceptual overlap of the constructs. In addition to this, participants did not actually use the website they evaluated, but were merely shown the site. Thus, it is possible that participants' evaluation of divergent constructs was based on visual aesthetics or that visual aesthetics produced a strong halo effect (Tractinsky et al., 2000). An additional study (study 6) was conducted to examine whether these interpretations could account for the rather high correlations of the VisAWI to divergent constructs.

3.2. Study 6: Experimental validation I

The purpose of the sixth study was twofold. The first purpose was to examine whether the rather high correlations of the VisAWI to perceived usability obtained in study 5 are an artifact due to the lack of real interaction with the evaluated website. To this end, participants in the present study had to perform several tasks prior to evaluate perceived usability and visual aesthetics. The second purpose of this study was to further validate the VisAWI using an experimental approach. Albeit the facets of the VisAWI are interrelated and jointly reflect a general aesthetics of websites factor, validity of the facets can be demonstrated by finding an experimental manipulation that selectively affects ratings on one facet while leaving the remaining facets unaffected. In the present study, the color scheme of a website was manipulated by a professional web designer such that the combination of colors was either aesthetically pleasing or aesthetically distorted. It was expected that this manipulation only affects ratings of the Colorfulness facet, whereas no difference between the experimental conditions should emerge with respect to Simplicity, Diversity, or Craftsmanship.

3.2.1. Methods

3.2.1.1. Participants

The study was announced as a study on information retrieval from websites. Seventy-eight volunteers (55.1% male) participated in this study. Participants were randomly assigned to one of two conditions (aesthetically pleasant colors, $n = 39$, vs. aesthetically distorted colors, $n = 39$). Age ranged from 15 to 77 years ($M = 38.82$; $SD = 14.05$). Participants were selected

via a German online panel. Participants received credit points within the panel (value was about 3€) for completing the study.

3.2.1.2. Procedures

After providing demographical background information, participants were randomly presented one of the two websites, and were requested to perform a series of search tasks. Next, they completed the VisAWI, a scale measuring perceived website usability, and several questions not pertinent to the present study. All items were presented in a completely randomized order. Presentation of the items and the websites were identical to the previous studies. Participation in the experiment took approximately 20 minutes.

3.2.1.3. Materials

A mock website portraying a fictitious German foundation was created and built as stimulus material for this study. The website was designed in a way that it was representative of corporate websites. Design of the site was based on other foundation and business websites. The website provided information concerning the foundation itself, its history, working areas, committees, events, and awards. Visual aesthetics of this website was manipulated by applying two different color schemes that were created by a professional web designer. Thereby we obtained two variants of the website that merely differed by the color scheme used. Figure 2 shows screenshots of these two variants. The colors used in both schemes were required to be isoluminant in order to minimize the effect of the color manipulation on other variables apart from visual aesthetics.

3.2.1.4. Measures

The final 18 item version of the VisAWI was used. Perceived usability was measured in the same way as in the previous study (Flavián et al., 2006).

3.2.2. Results

The correlations between the VisAWI and the scale measuring perceived usability are presented in the lower panel of Table 6. Generally, the correlations were much weaker than in the previous study. Only the Simplicity and the Craftsmanship facets as well as the general factor were significantly related to perceived usability.

A MANOVA was performed on the VisAWI scores to examine whether the means differ significantly as a function of the website's color scheme (aesthetically pleasant colors vs. aesthetically distorted colors). The MANOVA was significant [$F(4,73) = 2.69, p < .01$]. Post-hoc univariate ANOVAs were conducted on the single scales to qualify the multivariate effect (Table 7). As hypothesized, only the mean scores on the Colorfulness facet differed significantly by website. Means, standard deviations, and standardized mean differences are shown in Table 7.

3.2.3. Discussion

The first purpose of this study was to examine the possibility that the correlations between the VisAWI and perceived usability obtained in study 5 were inflated due to the lack of real interaction with the website being evaluated. In the present study, participants were required to complete various tasks prior to the evaluation of visual aesthetics and perceived usability. Consistent with the idea that interacting with a website provides more independent evaluations of perceived usability and visual aesthetics, the relationship between the VisAWI and perceived usability was found to be much less pronounced than in study 5. Specifically, perceived usability was moderately correlated with Simplicity, Craftsmanship, and the total

score, whereas no significant correlations emerged regarding Diversity and Colorfulness. However, a relationship between perceived usability and the Simplicity facet is not surprising and may just mirror that attributes enhancing simplicity of a layout seem to be features of both, aesthetically pleasant and usable websites (Karvonen, 2000; Lavie and Tractinsky, 2004). Taken together, this pattern of results provides further evidence for divergent validity of the VisAWI.

The second purpose of this study was to examine the construct validity of the VisAWI using an experimental approach. The facets of the VisAWI jointly reflect visual aesthetics of websites and are therefore expected to be interrelated. However, each facet also carries unique meaning. Hence, it was examined whether altering the color scheme of a website selectively affects ratings on the Colorfulness facet. Consistent with the predictions, manipulation of the color scheme exerted a strong effect on the Colorfulness facet, but had no effect on the remaining facets. Thus, the facets are only sensitive to certain layout features providing evidence for construct validity.

3.3. Study 7: Experimental validation II

The purpose of the final study was to provide further evidence concerning the construct validity of the VisAWI. A professional web designer was instructed to create an aesthetically pleasing and an aesthetically distorted version of the same website. It was expected that these two versions receive significantly different ratings on each facet of the VisAWI.

3.3.1. Methods

3.3.1.1. Participants and procedures

A total of 375 web users volunteered to participate anonymously in this study. Participants did not receive compensation for completing the study. Of the participants, 284 (75.7%) were female. Age ranged from 18 to 47 years with a mean of $M = 25.18$ years ($SD = 6.05$). Participants were randomly assigned to one of two conditions (aesthetically pleasant website, $n = 172$, vs. aesthetically distorted website, $n = 203$). Procedures were largely identical to the previous studies. Completing the study took about 25 minutes.

3.3.1.2. Materials

The materials used in the present study are based on a mock website described in greater detail in Moshagen et al. (2009). The website provided information on health-related issues (e.g., fitness, nutrition, and physical diseases). A professional designer was requested to create an aesthetically pleasing and an aesthetically distorted version of this website. Given that the website was meant to represent an information portal, the designer was instructed to alter visual aesthetics without changing the content or the general structure of the layout, such that both resulting versions could still represent the same information portal. The designer was neither aware of the VisAWI, nor of the purpose of the current study. The aesthetically pleasing and the aesthetically distorted versions of the website differed by color, density, picture quality, and typography.

3.3.1.3. Measures

The final version of the VisAWI was used.

3.3.2. Results and discussion

Table 8 presents the means, standard deviations, and standardized mean differences as a function of the experimental condition. A MANOVA on the VisAWI facets was conducted to examine whether the aesthetically pleasing version of the website received significantly higher scores than the aesthetically distorted version. The MANOVA was significant [$F(370, 3) = 76.13, p < .01$]. Consistent with our hypothesis, post-hoc univariate ANOVAs indicated that there were significant differences between the two versions of the website on each facet (Table 8). The standardized mean differences show that the largest effect emerged on Craftsmanship with the aesthetically pleasing and the aesthetically distorted website differing by 1.60 standard deviations. The smallest effect emerged on Diversity, however; the effect size of $d = 0.49$ still indicates an effect of a medium magnitude (Cohen, 1988). These results show that each facet of the VisAWI is responsive to changes in visual aesthetics, further supporting the construct validity of the VisAWI.

4. General Discussion

On the basis of an apparent need for a well-designed and carefully validated instrument, we sought to provide a precise operational definition and to develop a new measure of perceived visual aesthetics of websites. Seven studies were described providing evidence supporting the construct validity of the VisAWI. In constructing the VisAWI, an effort was made to be as inclusive as possible in defining the construct of visual website aesthetics, so that the resulting instrument would completely describe the domain of interest. The emerged subscales contained in the VisAWI demonstrate good internal consistencies and represent facets of visual website aesthetics of both practical and theoretical interest. Simplicity and Diversity have repeatedly been treated as formal parameters of aesthetic objects throughout the history of empirical aesthetics and are also contained in Lavie and Tractinsky's (2004) measure of website aesthetics. Colors are a critical property of aesthetic objects given that they can arouse physiological, cognitive, and emotional reactions (Elliot and Maier, 2007). Finally, a website needs to be harmoniously designed and the design ideas have to be skillfully integrated, which is addressed by the Craftsmanship facet.

4.1. Emerged facets of visual website aesthetics

Mirroring the interactionist view on aesthetics outlined in the introduction, the facets refer to objective properties of websites subject to idiosyncratic evaluations. In contrast to formalized approaches to aesthetic objects that attempt to measure the degree a certain physical property (such as balance and symmetry) is realized (Ngo et al., 2000, 2003; Bauerly and Liu, 2006), VisAWI scores are based on subjective evaluations of a particular facet. For example, a high rating on the Diversity facet does not necessarily imply that the design of the website is exceedingly dynamic or creative, but merely that the realized degree of Diversity is positively valued by a given perceiver.

4.1.1 Simplicity

Simplicity comprises aspects related to the Gestalt psychologists' figural goodness concept such as unity, homogeneity, clarity, orderliness, and balance (Arnheim, 1974). Simplicity is a central building block in early attempts to formalize the aesthetic value of objects (Birkhoff, 1933; Eysenck, 1941) and is also inherent in the information theoretic approach to aesthetics

(Hochberg and McAlister, 1953; van der Helm and Leeuwenberg, 1996). The processing fluency approach (Reber et al., 2004) posits that aesthetic appraisal is a function of the perceiver's processing demands. Simple layouts can be processed more fluently and should therefore be valued positively. Accordingly, research in human-computer interaction has repeatedly demonstrated the importance of simplicity for the aesthetic appreciation of websites (Bi et al., in press; Lavie and Tractinsky, 2004; Ngo et al., 2000, 2003; Thielsch and Hirschfeld, in press).

4.1.2 Diversity

Jointly with Simplicity, this facet mirrors Fechner's (1876) notion of "unity in diversity" and thus corresponds to the complexity components in formalized approaches to aesthetic objects (Birkhoff, 1933; Eysenck, 1941). In addition to visual complexity (also called visual richness, Nasar, 1999), the Diversity facet comprises dynamics, novelty, and creativity. The psychobiological theory on aesthetics (Berlyne, 1971) posits that complexity and novelty are collative variables that determine the arousal potential of a stimulus. A stimulus that is merely simple is perceived as boring, because it results in a low arousal and eventually in a negative aesthetic response. Diversity counteracts low arousal by provoking interest and tension (Hekkert and van Wieringen, 1990; Hekkert et al., 2003) and is therefore an important component with respect to the aesthetic appraisal of websites (Lavie and Tractinsky, 2004; Haig and Whitfield, 2001; Pandir and Knight, 2006; Tuch et al., 2009).

4.1.3 Colorfulness

Wide agreement exists concerning the unique effect of colors and their composition on aesthetic appraisal in general (Arnheim, 1974; Kawabata and Zeki, 2004; Martindale and Moore, 1988; Solso, 2003) and with respect to the design of websites in particular (Cyr et al., 2010; Hall and Hanna, 2004; Hoadley, 1990; Moshagen et al., 2009). Accordingly, this facet taps aesthetic impressions stemming from the selection, placement, and combination of colors.

4.1.4 Craftsmanship

Craftsmanship can be characterized as the skillful and coherent integration of all relevant design dimensions. A website needs to be harmoniously designed and the artistic ideas need to be implemented with skill and care. From an art-historic perspective, Engholm (2002) describes different schools of style in web design and also points to their dependencies on available technologies. Website design is an area of rapid development and continuously changes as technology progresses (Ivory and Megraw, 2005). Therefore, it may also be necessary that websites are based on modern technologies to avoid the impression of being outdated. Moreover, aesthetic appreciation of a website may also depend on its membership to a certain school of style (Engholm, 2002), which resembles the prototypicality dimension of aesthetic objects (Hekkert and van Wieringen, 1990; Hekkert et al., 2003).

4.2. Validity

The present research also examined the emerged facets in conjunction with other constructs to assess convergent and discriminant validity patterns. Convergent validity could be established by showing reliable correlations between the VisAWI with an existing measure of visual aesthetics (Lavie and Tractinsky, 2004). Consistent with the predictions, the Simplicity facet

was strongest related to the conceptually related classic aesthetics scale, whereas the Diversity facet showed the highest correlations to the conceptually related expressive aesthetics scale. Moreover, the VisAWI scores were strongly related to overall appeal. Overall, these results provide initial evidence for the convergent validity of the VisAWI. Divergent validity was demonstrated by lower correlations to measures of perceived usability, pragmatic quality, and quality of the content of a website. The Simplicity facet exhibited the strongest relationship to evaluations of perceived usability and pragmatic quality, whereas correlations between perceived usability and the remaining facets were much weaker. Some authors argue that simplicity may act as a link between aesthetics and usability on a perceptual level (Karvonen, 2000; Lavie and Tractinsky, 2004), suggesting that simplicity seems to be a property of both, aesthetic and usable websites. In fact, the results from the validation studies are in line with this notion (for a similar result, see Lavie and Tractinsky, 2004). It is also interesting to note that the correlations between usability and the VisAWI became weaker as participants were actually navigating rather than just watching the site to be evaluated. This is consistent with the idea that visual aesthetics may produce a halo effect on evaluations of other properties of websites (Tractinsky et al., 2000).

The utility of the VisAWI in distinguishing between different websites was demonstrated by showing that aesthetically pleasant and aesthetically unpleasant websites receive very different scores on each facet and with respect to the total score. Ratings of least and most aesthetic websites differed by about two standard deviations on each facet, which is more than double the benchmark of $d = 0.80$ for a “large effect size” according to Cohen (1988). Likewise, it could be experimentally demonstrated that each facet of the VisAWI is responsive to changes in the design of a website. These results indicate that the VisAWI performs very well in discriminating various degrees of perceived visual aesthetics of websites and, consequently, may also be gainfully employed to evaluate changes of existing designs targeting at enhancing visual aesthetics. The VisAWI was also found to be related to the intention to revisit a website, providing evidence concerning the concurrent validity of the VisAWI. Finally, albeit the facets of the VisAWI are interrelated and jointly reflect a general aesthetics of websites factor, it is important to note that each facet refers to particular properties of a website and thus carries unique meaning. It was experimentally demonstrated that manipulating the color scheme of a website selectively affected ratings on the Colorfulness facet, but showed no effect on ratings on the remaining facets. Thus, it could be demonstrated that the facets are only sensitive to certain layout properties supporting the construct validity of the VisAWI.

4.3 Limitations and future research

Several limitations should be considered when interpreting the results of the present research. First, although completing the VisAWI takes less than 3 minutes, the VisAWI may still be too lengthy in some research settings, for example when needing a simple manipulation check. It would be useful to develop a brief version of the VisAWI that enable its use when assessment times must be kept to a minimum. Second, the VisAWI is limited to the assessment of visual aesthetics and cannot be used to measure aesthetics elicited through other modalities (e.g., sounds and music). Third, the present research examined the VisAWI using websites only. It would be interesting to investigate whether this instrument can also be employed to evaluate other human-computer artifacts such as interfaces of mobile phones. Fourth, neither the participants nor the selection of websites were representative for the general population of internet users and the myriad of websites, respectively. Although the VisAWI was developed with more than 2,000 participants evaluating over 70 different websites from a variety of domains, the psychometric properties and validity of the scales remain to be generalized to other groups of participants and to other websites. Fifth, it should be emphasized that the

samples of participants used in the present research share a common cultural background. This may be a source of bias given that dimensions underlying visual aesthetics of websites may vary across cultures. Replication of the factor structure within independent samples, with different cultural backgrounds and languages, is recommended. Sixth, measuring aesthetics through Likert type scales assumes that the internal state of experiencing an aesthetic encounter is subject to conscious awareness. Albeit evidence suggests that the initial aesthetic response to briefly presented websites only marginally differs from a more elaborate aesthetic response after viewing a website for a longer period of time (Lindgaard et al., 2006; Tractinsky et al., 2006), it may also be possible that the cognitive representation of an immediate aesthetic response is not accurate or that automatic and conscious processes disagree with respect to the aesthetic evaluation of a stimulus (Leder et al., 2004). Seventh, some items included in the VisAWI, particularly those of the Diversity facet, are quite abstract and may be too general to guide designers in optimizing the design of a website. Finally, some important psychometric properties (e.g., test-retest stability) were not examined in the present research. This constitutes the next step in a comprehensive psychometric evaluation of the VisAWI.

5. Summary and Conclusion

Given an apparent need for a well-designed and carefully validated assessment instrument, the present research aimed at providing a precise operational definition and a sound measure of visual aesthetics of websites. Construction of the Visual Aesthetics of Websites Inventory (VisAWI) was based on a comprehensive and broad definition of the construct, so that the instrument would completely describe the domain of interest. Four facets of perceived visual aesthetics of websites were identified and validated in a series of seven studies: Simplicity, Diversity, Colorfulness, and Craftsmanship. Simplicity and Diversity have repeatedly been treated as formal parameters, colors are a critical property of aesthetic objects, and Craftsmanship addresses the skillful and coherent integration of all relevant design dimensions. These facets jointly represent perceived visual aesthetics, but are still distinguishable from each other. Assessment of reliability and tests of the factor structure indicate that the 18 items adequately reflect the construct. Evidence for the convergent, divergent, discriminative, and concurrent validity of the VisAWI was provided. Taken together, the results suggest that the VisAWI appears to be a sound measure comprising facets of both practical and theoretical interest, which allow for a precise assessment of perceived visual aesthetics of websites.

Acknowledgements

The authors would like to thank Timo Dinkler, Sarah Gersie, Günther Kebeck, Sascha Serwe, and Axel Zechner for their support and valuable suggestions. Furthermore, the authors would like to thank Jennifer Nicolai and three anonymous reviewers for helpful comments on previous versions of this manuscript.

Footnotes

¹ An anonymous reviewer suggested performing an additional EFA to investigate the factorial structure prior to moving to CFA. Thus, we conducted a re-analysis using the factor-analytic methods as described in study 2. The Kaiser-Guttman criterion, the scree-plot, and the results of parallel analysis suggested the extraction of four factors, which jointly accounted for 67% of the variance. These four factors and the items associated with each factors closely mirrored the results obtained through confirmatory factor analyses.

References

- Agarwal, R., Venkatesh, V., 2003. Assessing a firm's web presence: A heuristic evaluation procedure for the measurement of usability. *Information Systems Research* 13, 168–186.
- Aladwani, A.M., Palvia, P.C., 2002. Developing and validating an instrument for measuring user-perceived web quality. *Information & Management* 39, 467–476.
- Anastasi, A., Urbana, S., 1997. *Psychological Testing*. Prentice-Hall, Upper Saddle River.
- Arnheim, R., 1974. *Art and visual perception: A psychology of the creative eye*. University of California Press, Berkeley.
- Aspillaga, M., 1991. Screen design: Location of information and its effects on learning. *Journal of Computer-Based Instruction* 18, 89-92.
- Bauerly, M., Liu, Y., 2006. Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human-Computer Studies* 64, 670-682.
- Bauerly, M., Liu, Y., 2008. Effects of symmetry and number of compositional elements on interface and design aesthetics. *International Journal of Human-Computer Interaction* 24, 275-287.
- Baumgartner, H., Homburg, C., 1996. Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing* 13, 139-161.
- 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, 210-234.
- Bentler, P.M., Bonett, D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 88, 588-606.
- Berlyne, D.E., 1971. *Aesthetics and psychobiology*. Appleton-Century-Crofts, New York.
- Bi, L., Fan, X., Liu, Y., in press. 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*.
- Birkhoff, G.D., 1933. *Aesthetic measure*. Harvard University Press, Cambridge.
- Bollen, K.A., 1989. *Structural equations with latent variables*. Wiley, Oxford, England.
- Borsboom, D., Mellenbergh, G.J., van Heerden, J., 2004. The Concept of Validity. *Psychological Review* 111, 1061-1071.
- Brady, L., Phillips, C., 2003. Aesthetics and usability: A look at colour and balance. *Usability News* 5(2). Available at: <http://www.surl.org/usabilitynews/51/aesthetics.asp>.
- Browne, M.W., Cudeck, R., 1993. Alternative ways of assessing model fit. In Bollen, K.A., Long, J.S. (Eds.), *Testing structural equation models*. Sage, Thousand Oaks, CA, pp. 136-162.
- Chang, H.H., Chen, S.W., 2009. Consumer perception of interface quality, security, and loyalty in electronic commerce. *Information & Management* 46, 411–417.

- Cliff, N., 1988. The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin* 103, 276-279.
- Cober, R.T., Brown, D.J., Levy, P.E., Cober, A.B., Keeping, L.M., 2003. Organizational web sites: Web site content and style as determinants of organizational attraction. *International Journal of Selection and Assessment* 11, 158-169.
- Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. Erlbaum, Mahwah.
- Coursaris, C.K., Sweirenga, S.J., Watrall, E., 2008. An empirical investigation of color temperature and gender effects on web aesthetics. *Journal of Usability Studies* 3, 103-117.
- Cox, J., Dale, B.G., 2002. Key quality factors in Web site design and use: An examination. *International Journal of Quality and Reliability Management* 19, 862-888.
- Creusen, M., Snelders, D., 2002. Product appearance and consumer pleasure. In Green, W.S., Jordan, P.W. (Eds.), *Pleasure with Products: Beyond Usability*. Taylor & Francis, London, pp. 69-75.
- 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, 1-21.
- Cyr, D., Kindra, G.S., Dash, S., 2008. Web site design, trust, satisfaction and e-loyalty: The Indian experience. *Online Information Review* 32, 773-790.
- De Angeli, A., Sutcliffe, A., Hartmann, J., 2006. Interaction, usability and aesthetics: What influences users' preferences? In *Proceedings of the 6th conference on Designing Interactive systems*. ACM, University Park, pp. 271-280.
- De Wulf, K., Schillewaert, N., Muylle, S., Rangarajan, D., 2006. The role of pleasure in web site success. *Information & Management* 43, 434-446.
- Dollinger, S.J., Malmquist, D., 2009. Reliability and validity of single-item self-reports. *Journal of General Psychology* 136, 231-241.
- Elliot, A.J., Maier, M.A., 2007. Color and psychological functioning. *Current Directions in Psychological Science* 16, 250.
- Engholm, I., 2002. Digital style history: The development of graphic design on the Internet. *Digital Creativity* 13, 193-211.
- Eysenck, H., 1941. The empirical determination of an aesthetic formula. *Psychological Review* 48, 83-92.
- Fava, J.L., Velicer, W.F., 1992. The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research* 27, 387-415.
- Feagin, S.L., Maynard, P., 1997. *Aesthetics*. Oxford University Press, Oxford.
- Fechner, G., 1876. *Vorschule der Ästhetik* [Preschool of aesthetics]. Breitkopf & Härtel, Leipzig.
- Fenner, D.E.W., 1996. *The aesthetic attitude*. Humanities Press, Atlantic Highlands.
- Finn, A., Kayande, U., 2003. A parsimonious web site interaction assessment scale. Available at: www.ciras.com/papers/webgscalem.doc.
- Flavián, C., Guinalíu, M., Gurrea, R., 2006. The role played by perceived usability, satisfaction and

- consumer trust on website loyalty. *Information & Management* 43, 1-14.
- Floyd, F.J., Widaman, K.F., 1995. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* 7, 286-299.
- Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J., Tauber, E.R., 2003. How do users evaluate the credibility of Web sites? In *Proceedings of the 2003 conference on Designing for user experiences*. ACM, San Francisco, California, pp. 1-15.
- Goldstein, G., Hersen, M., 2000. *Handbook of psychological assessment*. Elsevier, Oxford.
- Gombrich, E.H., 1995. *The story of art*. Phaidon, London.
- Gorsuch, R.L., 1997. Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment* 68, 532-560.
- Grabinger, R., 1993. Computer screen designs: Viewer judgments. *Educational Technology Research and Development* 41, 35-73.
- Haig, A., Whitfield, T., 2001. Predicting the aesthetic performance of Web sites: What attracts people? In *First International Symposium on Smart Graphics*. ACM Press, New York.
- Hall, R.H., Hanna, P., 2004. The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour & Information Technology* 23, 183-195.
- Hartmann, J., Sutcliffe, A., De Angeli, A., 2007. Investigating attractiveness in web user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, San Jose, pp. 387-396.
- 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), 1-30.
- Hassenzahl, M., 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction* 19, 319-349.
- Hassenzahl, M., Burmester, M., Koller, F., 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [AttrakDiff: A questionnaire for the measurement of perceived pragmatic and hedonic quality]. In Ziegler, J., Szwillus, G. (Eds.), *Mensch & Computer*. Teubner, Stuttgart, pp. 187-196.
- Hassenzahl, M., Tractinsky, N., 2006. User experience - a research agenda. *Behaviour & Information Technology* 25, 91-97.
- Haynes, S.N., O'Brien, W.H., 2000. *Principles and practice of behavioral assessment*. Springer, New York.
- Hekkert, P., Snelders, D., van Wieringen, P.C., 2003. 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology* 94, 111-124.
- Hekkert, P., van Wieringen, P.C., 1990. Complexity and prototypicality as determinants of the appraisal of cubist paintings. *British Journal of Psychology* 81, 483-495.
- Hoadley, E., 1990. Investigating the effects of color. *Communications of the ACM* 33, 120-125.

- Hochberg, J., McAlister, E., 1953. A quantitative approach to figural "goodness". *Journal of Experimental Psychology* 46, 361-364.
- Hoeger, R., 1997. Speed of processing and stimulus complexity in low-frequency and high-frequency channels. *Perception* 26, 1039-1045.
- Hong, S., Kim, J., 2004. Architectural criteria for website evaluation - conceptual framework and empirical validation. *Behaviour & Information Technology* 23, 337-357.
- Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179-185.
- Hu, L.T., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6, 1-55.
- Hüppe, A., 1984. *Prägnanz - ein gestalttheoretischer Grundbegriff [Prägnanz - A fundamental term in gestalt-theory]*. Profil, München.
- International Organization for Standardization, 2009. ISO FDIS 9241-210: Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Organization for Standardization, Geneva.
- Ivory, M.Y., Megraw, R., 2005. Evolution of web site design patterns. *ACM Transactions on Information Systems* 23, 497.
- Jordan, P.W., 1998. Human factors for pleasure in product use. *Applied Ergonomics* 29, 25-33.
- Karvonen, K., 2000. The beauty of simplicity. In *ACM Proceedings on the 2000 conference on Universal Usability*. ACM Press, New York, pp. 85-90.
- Kawabata, H., Zeki, S., 2004. Neural correlates of beauty. *Journal of Neurophysiology* 91, 1699-1705.
- Kim, J., Lee, J., Choi, D., 2003. Designing emotionally evocative homepages: An empirical study of the quantitative relations between design factors and emotional dimensions. *International Journal of Human-Computer Studies* 59, 899-940.
- Kim, J., Moon, J.Y., 1998. Designing towards emotional usability in customer interfaces - trustworthiness of cyber-banking system interfaces. *Interacting with Computers* 10, 1-29.
- Kirakowski, J., Claridge, N., 2001. Human centered measures of success in web site design. In Vanderdonck, J., Blandford, A., Derycke, A. (Eds.), *IHM-HCI Conference*. Editions Cépaduès, Toulouse.
- Kline, P., 2000. *The handbook of psychological testing*. Routledge, London.
- Kline, R.B., 2005. *Principles and practice of structural equation modeling*. Guilford Press, New York.
- Koffka, K., 1935. *Principles of Gestalt Psychology*. Kegan, London.
- Kurosu, M., Kashimura, K., 1995. Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. In *CHI 95 Conference companion*. ACM, Denver, Colorado, pp. 292-293.
- Lai, C., Chen, P., Shih, S., Liu, Y., Hong, J., 2010. Computational models and experimental investigations of effects of balance and symmetry on the aesthetics of text-overlaid images. *International Journal of Human-Computer Studies* 68, 41-56.

- Lavie, T., Tractinsky, N., 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies* 60, 269-298.
- Leder, H., Belke, B., Oeberst, A., Augustin, D., 2004. A model of aesthetic appreciation and aesthetic judgements. *British Journal of Psychology* 95, 489-508.
- Lindgaard, G., Dudek, C., 2003. What is this evasive beast we call user satisfaction? *Interacting with Computers* 15, 429-452.
- 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, 115-126.
- Lindgaard, G., Whitfield, T.W.A., 2004. Integrating aesthetics within an evolutionary and psychological framework. *Theoretical Issues in Ergonomics Science* 5, 73-90.
- Ling, J., van Schaik, P., 2002. The effect of text and background colour on visual search of Web pages. *Displays* 23, 223-230.
- Ling, J., van Schaik, P., 2004. The effects of link format and screen location on visual search of web pages. *Ergonomics* 47, 907-921.
- Liu, Y., 2003a. Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. *Ergonomics* 46, 1273-1292.
- Liu, Y., 2003b. The aesthetic and the ethic dimensions of human factors and design. *Ergonomics* 46, 1293-1305.
- Loiacono, E.T., Watson, R.T., Goodhue, D.L., 2002. WebQual: A measure of Web site quality. *Marketing Theory and Applications* 13, 37-64.
- Loiacono, E.T., Watson, R.T., Goodhue, D.L., 2007. WebQual: An instrument for consumer evaluation of web sites. *International Journal of Electronic Commerce* 11(3), 51-87.
- Lorenzo-Silva, U., Ferrando, P.J., 2006. FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods* 38, 88-91.
- MacCallum, R.C., Browne, M.W., Sugawara, H.M., 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 1, 130-149.
- MacCallum, R.C., Roznowski, M., Necowitz, L.B., 1992. Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin* 111, 490-504.
- Mahlke, S., 2002. Factors influencing the experience of website usage. In *CHI '02 extended abstracts on Human factors in computing systems*. ACM, Minneapolis, pp. 846-847.
- Martindale, C., Moore, K., 1988. Priming, prototypicality, and preference. *Journal of Experimental Psychology: Human Perception and Performance* 14, 661-670.
- Mathwick, C., Malhotra, N., Ridgon, E., 2001. Experiential value: Conceptualization, measurement and application in the catalog and Internet shopping environment. *Journal of Retailing* 77, 39-56.
- Messick, S., 1980. Test validity and the ethics of assessment. *American Psychologist* 35, 1012-1027.

- Moshagen, M., Musch, J., Göritz, A.S., 2009. A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics* 52, 1311-1320.
- Motl, R.W., DiStefano, C., 2002. Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling* 9, 562-578.
- Mundorf, N., Westin, S., Dholakia, N., 1993. Effects of hedonic components and user's gender on the acceptance of screen-based information services. *Behaviour & Information Technology* 12, 293-303.
- Muthén, L.K., Muthén, B.O., 2008. *Mplus user's guide*. Muthén & Muthén, Los Angeles.
- Nasar, J.L., 1999. Perception and evaluation of residential street scenes. In Nasar, J.L., Preiser, W.F.E. (Eds.), *Directions in Person-Environment Research and Practice*. Aldershot, Ashgate.
- Ngo, D.C.L., Teo, L.S., Byrne, J.G., 2000. A mathematical theory of interface aesthetics. *Visual Mathematics* 2, 68-77.
- Ngo, D.C.L., Teo, L.S., Byrne, J.G., 2003. Modelling interface aesthetics. *Information Sciences* 152, 25-46.
- Norman, D.A., 2004. *Emotional Design*. Basic Book, New York.
- Nunnally, J.C., Bernstein, I.H., 1994. *Psychometric theory*. McGraw-Hill, New York.
- Osborne, H., Balakian, A.E., 1968. *Aesthetics and art history*. Longmans, Harlow.
- Palmer, J.W., 2002. Web site usability, design, and performance metrics. *Information Systems Research* 13, 151-167.
- Palmer, S.E., 1982. Symmetry, transformation, and the structure of perceptual systems. In Beck, F. (Ed.), *Organization and representation in perception*. Erlbaum, Hillsdale, NJ, pp. 95-144.
- Pandir, M., Knight, J., 2006. Homepage aesthetics: The search for preference factors and the challenges of subjectivity. *Interacting with Computers* 18, 1351-1370.
- Papachristos, E., Tselios, N., Avouris, N., 2006. Modeling perceived value of color in web sites. In Antoniou, G. (Ed.), *Advances in artificial intelligence*. Springer, Berlin, pp. 567-570.
- Parboteeah, D.V., Valacich, J.S., Wells, J.D., 2009. The influence of website characteristics on a consumer's urge to buy impulsively. *Information Systems Research* 20, 60-78.
- Parizotto-Ribeiro, R., Hammond, N., 2005. Does aesthetics affect the users' perceptions of VLEs? In *Proceedings of the 12th International Conference on Artificial Intelligence in Education*. Amsterdam, Denmark, pp. 25-31.
- Pomales-Garcia, C., Liu, Y., Mendez, D., 2005. Web-Based Distance Learning Technology: Does Appearance Matter. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*. pp. 755-758.
- Ramachandran, V.S., Hirstein, W., 1999. The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies* 6, 15-51.
- Rau, P.L., Gao, Q., Liu, J., 2007. The effect of rich web portal design and floating animations on visual search. *International Journal of Human-Computer Interaction* 22, 195-216.

- Reber, R., Schwarz, N., Winkielman, P., 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review* 8, 364-382.
- Roast, C., Evriviades, M., Purcell, M., Steele, B., 2002. Interaction media-using IT and liking IT. In *Pan Hellenic Conference on Human-Computer Interaction*. Patras, Greece.
- Roberts, L., Rankin, L., Silver, E., Moore, D., Plunkett, S., Washburn, D., Wilch-Ringen, B., 2003. Looks good to me. In *CHI '03 extended abstracts on Human factors in computing systems*. ACM, Ft. Lauderdale, pp. 818-819.
- Robins, R.W., Hendin, H.M., Trzesniewski, K.H., 2001. Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin* 27, 151-161.
- Rossiter, J.R., 2002. The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing* 19, 305-335.
- Santayana, G., 1955. *The sense of beauty*. Dover, New York.
- Satorra, A., Bentler, P.M., 2001. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66, 507-514.
- Schenkman, B., Jönsson, F., 2000. Aesthetics and preferences of web pages. *Behaviour & Information Technology* 19, 367-377.
- Schmidt, F.L., Hunter, J.E., 1996. Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods* 1, 199-223.
- Schmidt, K., Liu, Y., Sridharan, S., 2009. Webpage aesthetics, performance and usability: Design variables and their effects. *Ergonomics* 52, 631-643.
- Schrepp, M., Held, T., Laugwitz, B., 2006. The influence of hedonic quality on the attractiveness of user interfaces of business management software. *Interacting with Computers* 18, 1055-1069.
- Sharps, M., Nunes, M., 2002. Gestalt and feature-intensive processing: Toward a unified model of human information processing. *Current Psychology* 21, 68-84.
- Shieh, K., Lin, C., 2000. Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics* 26, 527-536.
- Simon, S.J., 2001. The impact of culture and gender on web sites: An empirical study. *ACM Database for Advances in Information Systems* 32, 18-37.
- Solso, R.L., 2003. *The psychology of art and the evolution of the conscious brain*. MIT Press, Cambridge.
- Sonderegger, A., Sauer, J., 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics* 41, 403-410.
- Spector, P.E., 1992. *Summated rating scale construction: An introduction*. Sage, Newbury Park, CA.
- Steiger, J.H., 1990. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research* 25, 173-180.
- Sutcliffe, A., de Angeli, A., 2005. Assessing Interaction Styles in Web User Interfaces. In *Proceedings*

- Human-Computer Interaction - INTERACT 2005. pp. 405-417.
- Sutcliffe, A.G., Kurniawan, S., Shin, J., 2006. A method and advisor tool for multimedia user interface design. *International Journal of Human-Computer Studies* 64, 375-392.
- Szabo, M., Kanuka, H., 1999. 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, 23-42.
- Tarasewich, P., Daniel, H.Z., Griffin, H.E., 2001. Aesthetics and web site design. *Quarterly Journal of Electronic Commerce* 2, 67-82.
- Thielsch, M.T., 2008. *Ästhetik von Websites [Aesthetics of websites]*. MV Wissenschaft, Münster.
- Thielsch, M.T., Hirschfeld, G., in press. High and low spatial frequencies in website evaluations. *Ergonomics*.
- Thüring, M., Mahlke, S., 2007. Usability, aesthetics and emotions in human-technology interaction. *International Journal of Psychology* 42, 253-264.
- Tractinsky, N., 1997. Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, Atlanta, pp. 115-122.
- Tractinsky, N., Cokhavi, A., Kirschenbaum, M., Sharfi, T., 2006. Evaluating the consistency of immediate aesthetic perceptions of web pages. *International Journal of Human-Computer Studies* 64, 1071-1083.
- Tractinsky, N., Katz, A.S., Ikar, D., 2000. What is beautiful is usable. *Interacting with Computers* 13, 127-145.
- Tractinsky, N., Lowengart, O., 2007. Web-store aesthetics in e-retailing: A conceptual framework and some theoretical implications. *Academy of Marketing Science Review* 11, 1-18.
- Tuch, A.N., Bargas-Avila, J.A., Opwis, K., Wilhelm, F.H., 2009. Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies* 67, 703-715.
- van der Heijden, H., 2003. Factors influencing the usage of websites: The case of a generic portal in the Netherlands. *Information & Management* 40, 541-549.
- van der Helm, P.A., Leeuwenberg, E.L.J., 1996. Goodness of Visual Regularities: A Nontransformational Approach. *Psychological Review* 103, 429-456.
- van Schaik, P., Ling, J., 2005. Five psychometric scales for online measurement of the quality of human-computer interaction in web sites. *International Journal of Human-Computer Interaction* 18, 309-322.
- van Schaik, P., Ling, J., 2009. The role of context in perceptions of the aesthetics of web pages over time. *International Journal of Human-Computer Studies* 67, 79-89.
- von Ehrenfels, C., 1890. Über Gestaltqualitäten [On Gestalt qualities]. *Vierteljahresschrift für wissenschaftliche Philosophie* 14, 249-292.
- Wertheimer, M., 1923. Untersuchungen zur Lehre von der Gestalt [Studies on the science of Gestalt]. *Psychologische Forschung* 4, 301-350.

- Wilkinson, L. and Task Force on Statistical Inference, APA Board of Scientific Affairs, 1999. Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist* 54, 594-604.
- Winkielman, P., Halberstadt, J., Fazendeiro, T., Catty, S., 2006. Prototypes are attractive because they are easy on the mind. *Psychological Science* 17, 799.
- Yoo, B., Donthu, N., 2001. Developing a scale to measure the perceived quality of an Internet shopping site (SITEQUAL). *Quarterly Journal of Electronic Commerce* 2, 31-46.
- Zain, J.M., Tey, M., Goh, Y., 2007. Does Aesthetics of Web Page Interface Matters to Mandarin Learning? *International Journal of Computer Science and Network Security* 7(8), 43-51.
- Zajonc, R.B., 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology* 9, 1-27.
- Zhang, P., Li, N., 2005. The importance of affective quality. *Communications of the ACM* 48, 105-108.
- Zhang, P., von Dran, G., 2000. Satisfiers and dissatisfiers: A two-factor model for website design and evaluation. *Journal of the American Society for Information Science* 51, 1253-1268.
- Zheng, X.S., Chakraborty, I., Lin, J.J., Rauschenberger, R., 2009. Correlating low-level image statistics with users - rapid aesthetic and affective judgments of web pages. In *Proceedings of the 27th international conference on Human factors in computing systems*. ACM, Boston, pp. 1-10.
- Zhou, H., Fu, X., 2007. Understanding, measuring, and designing user experience: The causal relationship between the aesthetic quality of products and user affect. In Jacko, J.A. (Ed.), *Human-Computer Interaction, Part I, HCII 2007*. Springer, Heidelberg, pp. 340-349.

Tables

Table 1

Aspects of visual aesthetics of websites

Aspect	Sources	Number of items in initial pool
Animations, visual effects, movement, dynamics	Lavie and Tractinsky (2004); Rau et al. (2007); Sutcliffe and de Angeli (2005); Tarasewich et al. (2001)	7
Balance, equilibrium, symmetry	Bauerly and Liu (2006, 2008); Bi et al. (2010); Brady and Phillips (2003); Lai et al. (2010); Lavie and Tractinsky (2004); Ngo et al. (2003); Zheng et al. (2009)	6
Coherence, craftsmanship, harmony, modernity, professionalism, style	de Angeli et al. (2006); Kim et al. (2003); Hassenzahl (2004); Lavie and Tractinsky (2004); Thielsch (2008)	8
Color	Brady and Phillips (2003); Coursaris et al. (2008); Cyr et al. (2010); De Angeli et al. (2006); Hall and Hanna (2004); Kim et al. (2003); Ling and van Schaik (2002); Papachristos et al. (2006); Schrepp et al. (2006); Shieh and Lin (2000); Simon (2001); Sutcliffe and de Angeli (2005); Tarasewich et al. (2001); Thielsch (2008)	10
Complexity, diversity, variety	de Angeli et al. (2006); Ngo et al. (2003); Pandir and Knight (2006); Tuch et al. (2009)	8
Grouping, structure, order	Bauerly and Liu (2006, 2008); de Angeli et al. (2006); Lavie and Tractinsky (2004); Ling and van Schaik (2002); Schmidt et al. (2009); Schrepp et al. (2006); Schenkman and Jönsson (2000); Thielsch (2008)	6
Homogeneity, unity, regularity, uniformity	Kim et al. (2003); Ngo et al. (2003); Tarasewich et al. (2001)	7
Images, icons, graphics	de Angeli et al. (2006); Lai et al. (2010); Schenkman and Jönsson (2000); Schmidt et al. (2009); Simon (2001); Tarasewich et al. (2001)	6
Novelty, creativity, inventiveness, interestingness	Haig and Whitfield (2001); Lavie and Tractinsky (2004); Pandir and Knight (2006)	7
Proportion, cohesion	Bauerly and Liu (2006, 2008); Ngo et al. (2000, 2003)	5
Simplicity, clarity, parsimony, density	Bi et al. (in press); de Angeli et al. (2006); Karvonen (2000); Lavie and Tractinsky (2004); Ngo et al. (2003); Rau et al. (2007); Schmidt et al. (2009)	9
Text, fonts, links	Ling and van Schaik (2002); Schenkman and Jönsson (2000); Schmidt et al. (2009); Tarasewich et al. (2001)	5

Table 2

Summary of exploratory factor analysis of the initial item pool

Factor	Number of items	Range of loadings
I	12	.30-.79
II	7	.14-.68
III	9	.31-.85
IV	12	.21-.56
V	16	.33-.74
VI	21	.27-.64

Table 3

Goodness-of-fit statistics for factorial models of the VisAWI in study 3 and study 4

CFA	χ^2	df	p	CFI	RMSEA	SRMR
Study 3, CFA1	2230.94	687	< .01	0.891	0.067	0.083
Study 3, CFA 2	415.00	131	< .01	0.951	0.065	0.053
Study 4	405.57	131	< .01	0.945	0.064	0.052

Note. χ^2 = Satorra-Bentler scaled chi-square; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual. See text for details on model specifications.

Table 4

Items and factor loadings

No.	Item	Loading	
		Study 3	Study 4
Factor 1: Simplicity			
1	The layout appears too dense. (r)	0.63	0.69
5	The layout is easy to grasp.	0.77	0.80
9	Everything goes together on this site.	0.77	0.81
13	The site appears patchy. (r)	0.78	0.77
17	The layout appears well structured.	0.85	0.84
Factor 2: Diversity			
2	The layout is pleasantly varied.	0.72	0.71
6	The layout is inventive.	0.82	0.81
10	The design appears uninspired. (r)	0.86	0.84
14	The layout appears dynamic.	0.69	0.63
18	The design is uninteresting. (r)	0.85	0.80
Factor 3: Colourfulness			
3	The color composition is attractive.	0.91	0.89
7	The colors do not match. (r)	0.70	0.72
11	The choice of colors is botched. (r)	0.88	0.80
15	The colors are appealing.	0.83	0.85
Factor 4: Craftsmanship			
4	The layout appears professionally designed.	0.85	0.83
8	The layout is not up-to-date. (r)	0.77	0.70
12	The site is designed with care.	0.83	0.79
16	The design of the site lacks a concept. (r)	0.77	0.76

Note. Items were constrained to load on only one factor, i.e., the secondary loadings were equal to zero. Negatively-keyed items are indicated by (r) and are reverse-scored. The wording of each item is a translation from the German original which is available from the authors on request.

Table 5

Scale statistics of the facets and the second-order general factor

Facet	Mean	SD	α	Min	Max	d	F (df)
Simplicity	3.94	1.48	.89	2.35	5.09	2.26	3.07 (41, 470) **
Diversity	3.49	1.31	.87	2.00	4.60	2.57	3.09 (41, 470) **
Colorfulness	4.31	1.54	.89	2.98	5.33	1.86	2.09 (41, 470) **
Craftsmanship	4.36	1.42	.85	2.90	5.43	2.14	2.54 (41, 470) **
General factor	3.99	1.22	.94	2.72	4.90	1.99	2.72 (41, 470) **

Note. α = Cronbach's α estimate of internal scale consistency; d = standardized mean difference of the websites evaluated least and most aesthetic on the respective factor (study 5).

** = $p < .01$

Table 6

Correlations between the VisAWI to convergent and divergent constructs

	General factor	Simplicity	Diversity	Colorfulness	Craftsmanship
Study 5 (N = 512)					
Classic aesthetics	.79**	.82**	.52**	.61**	.71**
Expressive aesthetics	.62**	.36**	.80**	.44**	.52**
Appeal	.80**	.71**	.69**	.60**	.71**
Pragmatic quality	.65**	.77**	.43**	.41**	.57**
Quality of content	.62**	.56**	.53**	.48**	.52**
Perceived usability	.64**	.76**	.41**	.39**	.56**
Intention to revisit website	.51**	.44**	.48**	.40**	.40**
Study 6 (N = 78)					
Perceived usability	.31**	.48**	.19	.04	.41**

Note. Correlation coefficients were computed using the sum scores of the respective scales. Sum scores for the VisAWI facets were obtained by summing participants' responses to the items of the respective facet. The VisAWI general factor was determined by computing the total score of all items of the VisAWI.

** $p < .01$

Table 7

Means, standard deviations, and effect sizes of the facets and the second-order general factor as function of experimental condition (study 6)

Facet	Aesthetically distorted colors (N = 39)		Aesthetically pleasing colors (N = 39)		d	F (df)
	Mean	SD	Mean	SD		
Simplicity	5.28	0.75	5.34	0.68	0.08	0.15 (1, 76)
Diversity	3.72	1.35	3.70	1.11	-0.02	0.01 (1, 76)
Colorfulness	4.26	1.36	4.96	1.10	0.57	6.23 (1, 76)**
Craftsmanship	4.94	1.00	4.83	1.12	-0.10	0.23(1, 76)
General factor	4.55	0.90	4.71	0.81	0.19	0.64 (1, 76)

Note. Study 6 contrasted two versions of a website with the colors either being aesthetically pleasing or aesthetically distorted. d = standardized mean difference.

** p < .01 (one-sided)

Table 8

Means, standard deviations, and effect sizes of the facets and the second-order general factor as function of experimental condition (study 7)

Facet	Aesthetically distorted website (N = 203)		Aesthetically pleasing website (N = 172)		d	F (df)
	Mean	SD	Mean	SD		
Simplicity	3.71	1.29	4.86	1.15	0.94	82.48 (1, 373)**
Diversity	3.02	1.03	3.55	1.15	0.49	22.13 (1, 373)**
Colorfulness	2.87	1.33	4.36	1.35	1.12	114.90 (1, 373)**
Craftsmanship	2.95	1.30	4.89	1.12	1.60	235.99 (1, 373)**
General factor	3.16	1.05	4.40	0.97	1.22	137.05 (1, 373)**

Note. Study 7 contrasted two versions of a website with various elements either being aesthetically pleasing or aesthetically distorted. d = standardized mean difference.

** p < .01 (one-sided)

Figures

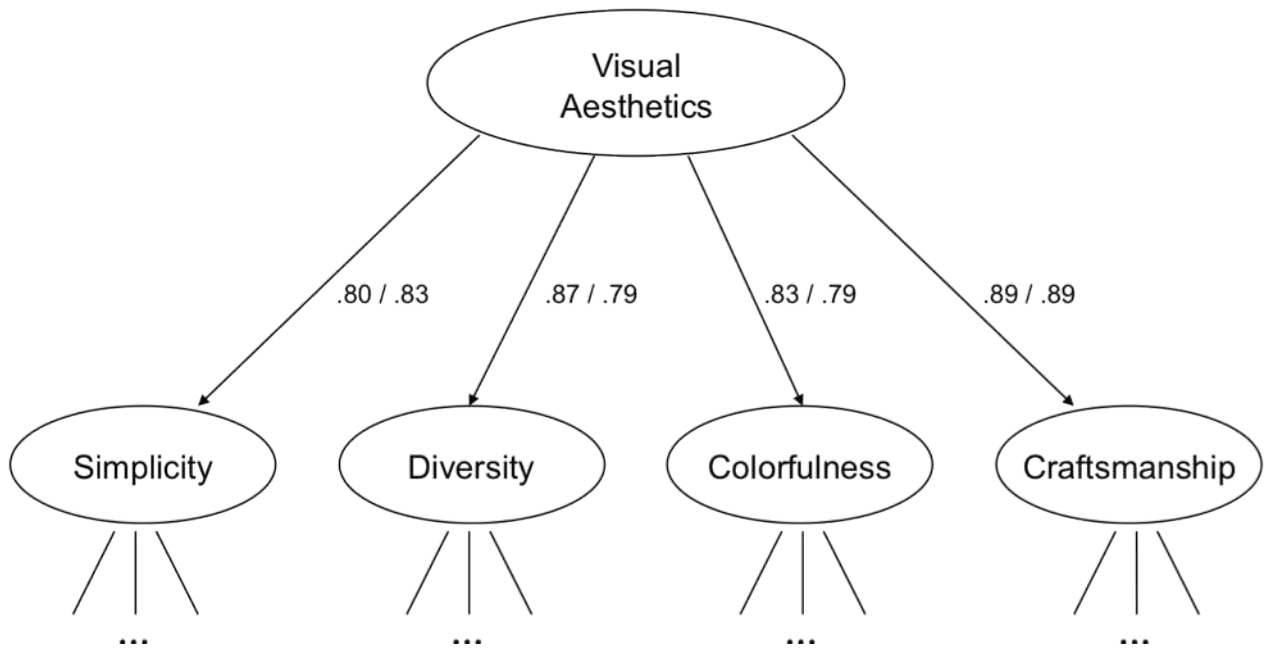


Figure 1. Structural model of the VisAWI. Second order factor loadings are presented as obtained in study 3 and study 4. Manifest indicators are omitted to enhance clarity.

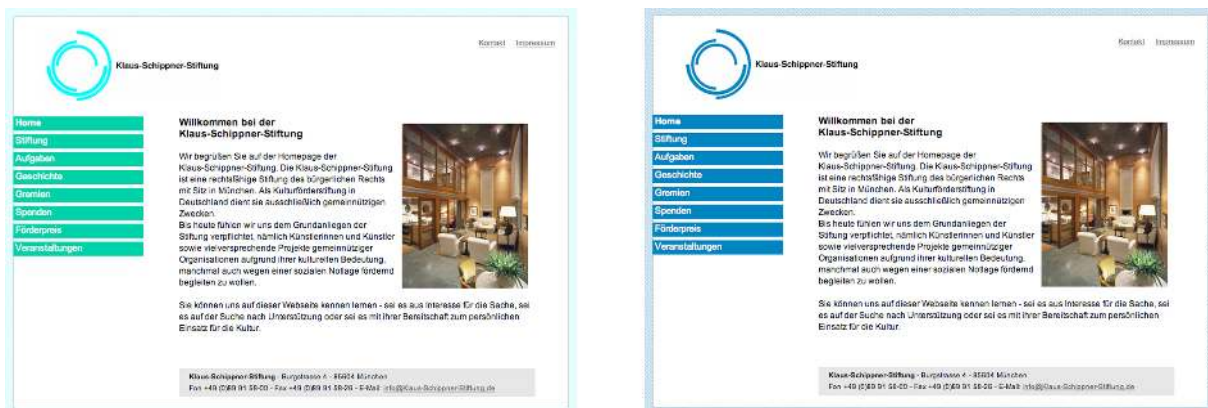


Figure 2. Websites used as stimulus materials in study 6. The aesthetically distorted color scheme is shown on the left side, and the aesthetically pleasing color scheme is shown on the right side.