An Exploration of Relations Between Visual Appeal, Trustworthiness and Perceived Usability of Homepages

GITTE LINDGAARD, CATHY DUDEK, DEVJANI SEN, LIVIA SUMEGI, and PATRICK NOONAN, Carleton University

Extremely high correlations between repeated judgments of visual appeal of homepages shown for 50 milliseconds have been interpreted as evidence for a mere exposure effect [Lindgaard et al. 2006]. Continuing that work, the present research had two objectives. First, it investigated the relationship between judgments differing in cognitive demands. Second, it began to identify specific visual attributes that appear to contribute to different judgments. Three experiments are reported. All used the stimuli and viewing time as before. Using a paradigm known to disrupt processing beyond the stimulus offset, Experiment 1 was designed to ensure that the previous findings could not be attributed to such continued processing. Adopting a within-subject design, Experiment 2 investigated the extent to which judgments differing in cognitive demands (visual appeal, perceived usability, trustworthiness) may be driven by the visual characteristics of a Web page. It also enabled analyses of visual attributes that contributed most to the different judgments. Experiment 3 replicated Experiment 2 but using a between-subject design to ensure that no practice effect could occur. The results suggest that all three types of judgments are largely driven by visual appeal, but that cognitively demanding judgments are processed in a qualitatively different manner than visual appeal, and that they rely on somewhat different visual attributes. A model accounting for the results is provided.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces–*Screen Design, User-Centred Design*; D.2.2 [**Design Tools and Techniques**]: User Interfaces

General Terms: Experimentation

Additional Key Words and Phrases: Visual appeal, perceived usability, trustworthiness, masking effects

ACM Reference Format:
Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P. 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. ACM Trans. Comput.-Hum. Interact. 18, 1, Article 1 (April 2011), 30 pages.
DOI = 10.1145/1959022.1959023 http://doi.acm.org/10.1145/1959022.1959023

1. INTRODUCTION

Regardless of whether we are asked to judge personality vignettes [Anderson 1980, 1981], formulate a medical diagnosis [Eddy 1999; Lindgaard et al. 2010], or assess the appeal of a Web site [Fernandes et al. 2003; Lindgaard et al. 2006; Schenkman and Jönsson 2000], our first impression is critical for developing a more considered opinion, or for determining a subsequent action. Even when the task does not require us explicitly to assess visual appeal, it appears to affect other kinds of judgments such as perceived reliability [Basso et al. 2001], perceived usability (e.g., van der Heijden

This work was partially supported by a Canadian National Science and Engineering Research Council (NSERC)/Cognos Industry Research Chair grant, IRCSA 234087-05.
Authors' addresses: G. Lindgaard (corresponding author), C. Dudek, D. Sen, L. Sumegi, and P. Noonan, Human Oriented Technology Lab, Department of Psychology, Carleton University, Ottawa, ON, Canada; email: gitte.lindgaard@carleton.ca.

DOI 10.1145/1959022.1959023 http://doi.acm.org/10.1145/1959022.1959023

[2003] and Sonderegger and Sauer [2010]), perceived information quality [De Angeli et al. 2006; Hartmann et al. 2007, 2008], usefulness [Schultz 2007], user satisfaction [Lalomia and Sidowski 1990], and trustworthiness (e.g., Fogg et al. [2002]; Handy [1995]). The assertion that "what is beautiful is [perceived to be] usable" Tractinsky et al. [2000]) complements the similar assertion that what is [perceived to be] usable is beautiful [Hassenzahl 2004], and the claim that "attractive things work better" [Norman 2004a]. Following Thorndike's [1920] initial observation of what he called the "halo" effect, also discussed by Rosenzweig [2009] and shown in studies by Dion et al. [1972], Tractinsky et al. [2000], and most recently by De Angeli et al. [2006] and Hartmann et al. [2008], suggests that beauty matters: it influences decisions that should be independent of aesthetics.

Judgments of visual appeal of homepages are made instantly and reliably [Fernandes et al. 2003; Lindgaard et al. 2006]. One would expect assessments of more cognitively demanding characteristics such as perceived usability and trustworthiness to take longer and be less reliable with very brief viewing times and in a paradigm that allows no interaction with the homepages. Cognitively demanding judgments should be based on careful consideration and reflection [Norman 2004a]. Some researchers argue that these can only be made after at least some interaction with a product or an application (e.g., De Angeli et al. [2006] and Hassenzahl [2008]). The present research had two objectives. One was to determine if a relationship can also be demonstrated between visual appeal and other, less immediately accessible qualities, when judges are prevented from interacting with the product in question. The second objective was to begin to identify distinct visual attributes that combine to yield an immediate impression of visual appeal, perceived usability, and trustworthiness. The objectives are discussed in more detail following a review of the relevant literature that is divided into three sections. Section 1.1 discusses the biological determinants of visual appeal. Section 1.2 reviews literature on visual appeal and trustworthiness, followed by a review of visual appeal and perceived usability in Section 1.3. The objectives of the present research are then outlined in some detail in Section 1.4. This is followed by three experiments in which the relationship between judgments of visual appeal and the more cognitively demanding judgments of trustworthiness and perceived usability are investigated. Analyses of the specific visual attributes that apparently contributed to the different judgments are then presented and discussed. The general discussion, presented next, leads into a reflection on some of the limitations of the study and thoughts on future work, and finally, conclusions are presented.

1.1. Biological, Precognitive Determinants of Visual Appeal

Lindgaard and her colleagues [2006] demonstrated extremely high reliability ratings of visual appeal, both within and between participants, of a sample of 50 homepages exposed in two trials in a within-subject design for 50 msec each. They interpreted their findings in terms of a "mere exposure effect" [Zajonc 1980], claiming that participants responded to a diffuse physiologically-based affective feeling in line with Damasio's [1994, 2000] "somatic marker" hypothesis rather than responses being the outcome of a considered cognitive process. The mere exposure effect is found in experiments using a very brief stimulus exposure time, between one and 50 msec [Bornstein 1992]. Evidence suggests that it is based on affect and that it occurs in the absence of cognitive processes [Zajonc 1980, 2001]. According to Zajonc, "careful experiments have ruled out explanations of this phenomenon based on ease of recognition, and increased perceptual fluency, or subjective familiarity" [2001, p. 225]. Zajonc further argues that "if cognitive processes are not involved in a behavior. . . affective influences, which are necessarily less diverse than cognitive influences, will dominate the behavior, yielding a more homogeneous array of reactions" [2001, p. 227]. Along with evidence provided

by neuroscientists such as LeDoux [1992, 1996] and Damasio [2000], Zajonc [2000] asserts that affect and cognition are likely to be independent processes (see also Elliott and Dolan [1998]). Therefore, since stimulus exposure times falling within the time frame in which the mere exposure effect is said to occur have been shown to yield extremely highly correlated visual appeal ratings, this suggests that these are based on affective, precognitive processes.

In a study involving health-related decisions that should not involve affect at all and in which participants were free to inspect the Web sites for as long as they liked, Sillence and her colleagues [2004] showed that participants scanned and sifted information and rejected ugly Web sites within 30 seconds. Thus, while that time frame by far exceeded the exposure times associated with the mere exposure effect, the decision to reject a site was still based on affect. The authors speculate that adoption of such an affective "heuristic analysis" may be typical for situations in which people are under time pressure or lack motivation to engage in more cognitive analytical processes [Sillence et al. 2006]. Alternatively, it is possible that the instant affective somatic response may occasionally interfere with one's a priori goal to locate high-quality information when encountering a visually unappealing site. As a consumer seeking specific health information online, I should be highly motivated by my need to find the best available information, but if I reject an ugly site before I have had time even to scan, let alone sift, the information it provides, I know nothing about the quality of its content. Realistically, one cannot judge the trustworthiness of a Web site owner at a glimpse, just as one cannot judge usability without interacting with a Web site. Despite this, it is interesting to investigate the extent to which judgments imposing different cognitive demands may be related, and if they are in some way distinguishable, perhaps by being shown to rely on different visual attributes. These issues were investigated here.

1.2. Visual Appeal and Trustworthiness

In a study of trust in online health information sites, Sillence and her colleagues [2006] found that the first impression served as a kind of screening device for patients browsing the Internet for specific health advice: visually unappealing sites were rejected within a few seconds, whereas more appealing sites were scrutinized for content before being either accepted or rejected. Factors contributing to the rapid rejection included busy, cluttered, complex screen layouts, boring use of color, and other design-related issues. Along similar lines, a large consumer survey showed that assessments of credibility were based on visual design for nearly one half (46.1%) of the respondents [Fogg et al. 2002, Cyr and Trevor-Smith 2004]. Credibility was seen to relate to trustworthiness. Karvonen [2000] linked clear, clean, and simple design with trustworthiness, and in an e-commerce context, the Cheskin Research Group [1999] mentioned "design quality" as one of the six most prominent features promoting trust in Web sites. Finally, McKnight et al.'s [2002] model of trust formation assumes that consumers who develop a positive intention/willingness to trust an e-commerce Web site are more likely to complete an online transaction than consumers who do not.

The observation that visually unappealing Web sites are, at least in some circumstances, rejected very quickly suggests that initial judgments of trustworthiness are likely to be independent of the perceived or objective quality of the information presented. Egger's [2002] model of trust includes such a preinteraction phase in which initial trust is formed [Carroll and Mentis 2008]. Relying on visual appeal as a screening device for selecting Web sites is an efficient mechanism for reducing what Simon [1956] called the "problem space" quickly when one is overwhelmed by a large number of sites to choose from. It is, however, unlikely also to be effective: an ugly site may well contain information that is as useful, accessible, and accurate as its prettier competitors. With interactive experience, Web site attractiveness has been found to influence

judgments of information quality [De Angeli et al. 2006; Hartmann et al. 2007, 2008] which, one would expect, would also lead to judgments of trustworthiness. The three experiments presented here were concerned exclusively with the first impression, this immediate sense of a situation or of an object that Wright and his colleagues [2008] refer to as "connecting."

1.3. Visual Appeal and Perceived Usability

Our preference for things that are visually attractive is evident in our choice of products and applications. Even when participants *say* they prefer highly usable entertainment systems, they have still been found to *choose* [Tractinsky and Zmiri 2006] those that are higher in visual appeal than in usability. In a study of the contribution of four different aspects of Web sites to the user experience, Mahlke [2002] found that "perceived visual attractiveness" accounted for 72% of the overall variance. In a sample of hypothetical mobile phones in which participants rated perceived usability and judged price, functionality, and beauty, Hassenzahl [2008] found that over half of the participants took beauty into account in their usability ratings. Along similar lines, Diefenbach and Hassenzahl [2008, 2009, Study 2] showed that, under certain conditions, participants *prefer* a predominantly beautiful to a predominantly usable mobile phone. Furthermore, applications perceived to be easier to use do appeal more to users [Tractinsky 1997; Tractinsky et al. 2000] even when visually appealing products suffer poorer usability than less appealing ones (e.g., Mahlke [2006] and De Angeli et al. [2006]). In some cases in which more attractive applications have lead to poorer performance (e.g., Sonderegger and Sauer [2010]), users have still been found to perceive these to be more usable than less attractive versions [Ben-Bassat et al. 2006]. Thus, as van der Heijden [2003] found in his study in which he was able clearly to separate "perceived visual attractiveness" from other perceptual and experiential aspects, the visual appeal of Web sites affects users' enjoyment as well as their perceptions of ease of use. In other studies, users realized that a less attractive but more effective educational application containing exactly the same information was more usable than the more attractive version; they still preferred the attractive version. This preference was, however, reversed when the purpose of the application was deemed more serious in the sense that it was said to be intended for university students instead of younger children [De Angeli et al. 2006; Hartmann et al. 2007, 2008], providing evidence for a framing effect [Tversky and Kahneman 1981]. The present experiments carried no serious consequences for the users; they provided no particular frame of reference, nor was the stated purpose of the Web sites manipulated.

In addition to affecting perceived usability, visual appeal has also been shown to influence actual Web site usage [Zhang and Li 2004]. In Zhang and Li's study, user perceptions of the quality of a course management system, based predominantly on visual appeal, were shown to be related both to users' intention to use it and to their actual usage patterns. Apparently, visual appeal contributes to the perceived quality of a system which, in turn, can become an antecedent to perceived trustworthiness, usability usefulness, emotional appeal, and even system usage. Some researchers go as far as to argue that visual appeal, expected usefulness and, presumably, usability drive purchasing decisions of consumer goods ranging from computers [Porteous 1996; Postrel 2001, 2002] to toasters [Bloch et al. 2003] and juicers [Buxton 2005], even to professional engineering equipment such as oscilloscopes and dispense pumps [Yamamoto and Lambert 1994]. Consumers, it is argued, increasingly make brand choices based on aesthetic value and visual design [Dumaine 1991; Schmitt and Simonson 1997], which is seen as central to the formation of the consumer/product relationship [Griffith et al. 2001; Hollins and Pugh 1990].

All of the aforesaid studies suggest that visual appeal may have implications for judgments that go well beyond the sheer look of a product, a software application, or a Web site. It is reasonable to assume that a link between perceived usability and aesthetics is based on those aspects of visual design that affect the more utilitarian features of usability. These include clear visual feedback, buttons that look and behave like buttons, semantically similar items that are grouped together, appropriate context-related terminology [Galitz 2007; Williams and Tollett 2005; Lynch and Horton 2002], and so on. These features of "good design," traditionally seen as usability-related in the HCI literature, are captured in Lavie and Tractinsky's [2004] "classic aesthetics" model. These properties, which lend a design order and harmony, are readily measurable independently of an observer [Desmet and Hekkert 2007]. Hassenzahl [2008] refers to these as "normative values," that is, rules prescribing how to accomplish a beautiful design, which designers have formalized in mathematical formulae such as the "golden ratio" (see e.g., Bezanson [2007], Hemenway [2005], Lidwell et al. [2003]). By contrast, features such as perceived creativity, imaginative and innovative design, captured in Lavie and Tractinsky's "expressive aesthetics" model, concern the subjective aesthetic experience [Frohlich 2004; McCarthy and Wright 2005], which Hassenzahl [2008] refers to as "experiential values". Lavie and Tractinsky's "expressive aesthetics" features culminate in subjective judgments of the designers' creative skills, and the scale captures the degree to which (the Gestalt impression of) beauty is in the eye of the beholder. It is easy to imagine a very beautiful object or Web site that is also quite unusable, just like one can imagine one that is very usable and quite ugly. One would therefore assume that the holistic (Gestalt) impression of beauty should have little or no effect on perceived usability unless good user interface design principles are obviously violated, thereby detracting from the classical aesthetic sense of beauty.

The preceding studies offer empirical evidence in support of a relationship between visual appeal and perceived usability, but not all findings do. In one study exploring users' experience with a Web site that was visually very appealing, appeal was just as highly valued after as before a standard usability test even though participants, on average, completed less than half the user tasks successfully [Lindgaard and Dudek 2002]. Usability ratings were significantly lower after than before the usability test. Yet, the strong impact of visual appeal seemed to reduce the importance of usability, which had only a minimal effect on participants' judgments of the interactive experience. This finding is similar to Hassenzahl's [2004] study in which he found very high test-retest reliability of visual appeal ($r = .87$) but not of usability judgments ($r = .46$), which were lower after than before participants had used the system. In both these studies, appeal judgments did not change after system usage, but usability judgments did. Likewise, Thüring and Mahlke [2007] found a weak effect of variations in visual appeal on usability ratings ($p < .10$), but variations in actual usability did not affect visual appeal ratings. As Hassenzahl [2008] notes, Lindgaard and her colleagues' [2006] study found very high correlations for (classical) aesthetics-based characteristics such as color and design, but not for the usability-related attributes of clear/confusing and simple/complex design. On the one hand, then, many studies suggest that visual appeal may determine subsequent usage and dominate judgments of unobservable properties such as perceived usability, usefulness, and trustworthiness. On the other hand, some research suggests that judgments of visual appeal may be independent of such properties. As this issue continues to generate controversy in the literature, the present research aims to contribute to the debate, concentrating solely on first impression judgments. In order to avoid the confusion and overcome the controversy surrounding the concept of aesthetics, often labeled "beauty," "attractiveness," or "visual appeal," we again used the term "visual appeal" in the present experiments. When judging visual appeal,

participants were simply asked to rate "the appeal of the Web site" on a 9-point scale.

1.4. Objectives of the Present Experiments

Our first objective was to learn more about the relationship between the three concepts of visual appeal, perceived trustworthiness, and perceived usability in a paradigm using very brief stimulus exposure times (50 msec). Experiment 1 was designed to test if Lindgaard et al.'s [2006] findings may have been unduly affected by the visual stimulus remaining accessible to the viewer beyond the stimulus offset time. It would challenge our interpretation of the earlier findings as an example of a mere exposure effect. The effect has been shown in numerous studies to begin to wane once the stimulus exposure time exceeds 50 msec [Bornstein 1992]. Therefore, if the images were still accessible after the stimulus offset at 50 msec in our earlier experiments, the results would not have amounted to a mere exposure effect. In the presence of the mere exposure effect, participants should not be able to discern visual details of the homepages. Instead, their judgment should be based on the initial diffuse physiological affect-based feeling generated by a holistic (Gestalt) first impression of visual appeal as described by Damasio's [2000] somatic markers.

To the best of our knowledge, the relationship between visual appeal and decisions imposing additional cognitive demands, that should be based on careful cognitive reflection, has not been tested in a paradigm using very brief stimulus exposure times. That was therefore the purpose of Experiment 2 in which stimulus exposure times of 50 milliseconds and the same stimuli were employed as in our earlier studies [Fernandes et al. 2003; Lindgaard et al. 2006]. In addition to judging visual appeal, participants here were also required to judge trustworthiness and perceived usability. To the extent that appeal is indeed judged on the basis of a diffuse feeling of affect, judgments of trustworthiness and perceived usability of homepages seen so briefly would only be anticipatory. This notion of anticipatory judgments is taken up in Desmet and Hekkert's [2007] discussion of nonphysical human-product interaction. These researchers argue that it is quite possible to anticipate or fantasize about interaction before it takes place, suggesting that anticipatory judgments should also be possible in the present situation in which participants were not invited to interact with the Web sites. The anticipation of human-product interaction can elicit affective responses, as found in van der Heijden's [2003] study, for example, in which he tested the Technology Acceptance Model (TAM). The TAM, introduced by Davis [1989] and adapted from the theory of reasoned action [Ajzen and Fishbein 1980], assumes that acceptance of an information system is mediated by the system's perceived (anticipated) usefulness to the individual and its perceived ease of use. The model relies on a causal chain of beliefs that influence attitude, which in turn influences intention to use. That research attests to the ecological validity of assessing perceived usability as well as trustworthiness even without actual use of a Web site as in the present experiments.

The second objective addressed by the present research was to begin to identify visual attributes that may combine into initial impressions of visual appeal, trustworthiness, and perceived usability. To enable assessment of the potential contributions of visual attributes that could be discernible with very limited stimulus exposure, it was desirable to obtain participants' judgments on all three variables. That was done in Experiments 2 and 3.

2. EXPERIMENT 1

2.1. The Nature of Masking

Masking is widely used to study the dynamics of visual information processing (e.g., Enns and Di Lollo [2000]). In backward masking paradigms, a target stimulus is

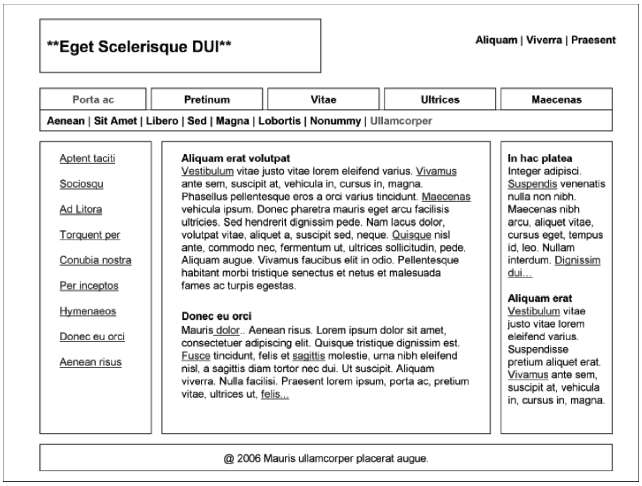

Fig. 1. The mask presented after each stimulus.

presented briefly, followed by a scrambled visual pattern that resembles the target stimulus to some extent. The purpose of backward masking is to cancel further processing of the target stimulus after its offset [Breitemeier and Ogmen 2000; Verleger et al. 2004]. According to Rieger and his colleagues [2005], a representation of the physical target stimulus properties are encoded into a sensory visual buffer where it is continually analyzed by higher processing stages and transformed into increasingly abstract codes that are then transferred into memory. The models assume that the mask overwrites the stimulus template in the sensory buffer, replacing it with a representation of the mask. The mask thus imposes constraints on the processing of the target stimulus [Breitmeier and Ogmen 2000]. Rieger et al.'s [2005] research provided empirical support for this in a study in which they integrated psychophysical and physiological data and employed conditions with and without a mask. Using stimuli comprising complex images of natural scenes, their results showed that viewers had access to the stimulus beyond the target exposure time. Therefore, when no mask is used, it would appear that the iconic trace of the target stimulus remains in the visual buffer where it decays approximately one second after the stimulus offset [Averbach and Sperling 1961; Kovacs et al. 1995; Sperling 1960]. These process models therefore challenge Lindgaard et al.'s [2006] earlier claim of a mere exposure effect in studies in which they did not use backward masking. While it was important to prevent processing of the target beyond its 50 msec exposure time, we did not want the mask to exert an overpowering influence on perception either. A predominantly medium grey mask was therefore chosen. Compared with white and black masks, grey masks have been shown to exert a medium strong effect on accurate stimulus detection in the context of letter identification tasks [Spencer and Shuntich 1970].

The design of the mask, which was used for all the stimuli, resembled a typical homepage as shown in Figure 1. The background color was a medium grey (shown here as white to increase the contrast in the smaller image), with black nonsense text and blue underlined fictitious hyperlinks. Although the design of the stimulus pages varied widely, the design of the mask was assumed to suffice for achieving the desired effect of preventing prolonged processing of the target stimuli. The maximum backward masking effect has been shown to fall symmetrically around a stimulus onset asynchrony interval ranging from 0 to 100 msec [Rieger et al. 2005]. At exposure times of about 100 msec, the masking effect is absent [Enns and Di Lollo 2000]. A

50-millisecond exposure time was therefore chosen for both the target and the mask, with the mask appearing directly upon the offset of the target. Ratings were obtained only for visual appeal.

2.2. Method

A sample of 20 native English-speaking undergraduate students took part. All had normal vision or corrected-to-normal vision, and no one was colorblind. Participants were tested individually in sessions lasting approximately 30 minutes and received course credit for their participation.

2.2.1. Apparatus. Participants were tested on a workstation with 1.6 GHz Athlon CPU, 256Mbytes of RAM, Matrox dual-head video card, and a Samsung SyncMaster 950p 19-inch monitor with a white balance calibrated at 9300° K and a gamma value of 2.1.

2.2.2. Materials. The stimuli comprised a set of 50 homepages selected from a sample of 100 in earlier studies [Fernandes et al. 2003]. Stimulus genre was not controlled, but the selected homepages would not have had wide public exposure. In the first of those studies, the 100 images were shown for 500 msec each. Participants rated the visual appeal of each image in two test phases, always shown in different random orders. The stimuli and the task requirements were identical in the two test phases. The correlation between test phases 1 and 2 was very high ($r = .97$). Ratings were transformed into z-scores from which a subset of 50 images, representing the 25 highest- and 25 lowest-rated images, was selected. These were submitted to further testing in which a new sample of participants saw the 50 images for either 500 msec or 50 msec, again in two test phases. Correlations for the two viewing conditions were very high for both test phase 1 ($r = .95$) and test phase 2 ($r = .97$), suggesting that ratings were highly reliable in both the 50 msec and the 500 msec condition. Screen shots of each homepage were taken in an Internet Explorer 6.0 browser at 1024 × 768 pixel resolution in 32-bit true color. The images looked like they were being viewed in the Internet Explorer browser. The screen shots were displayed via a software application, DirectRT™that randomized the sequence in which screen shots were shown as well as controlling the display time and recording response times.

2.2.3. Procedure. After reading a briefing form and signing an informed consent sheet, participants were seated in front of a computer, adjusting the seating height and monitor angle to their preference. A set of 20 practice homepages, not included in the data analysis, were shown to familiarize participants with the procedure. Each of the 50 homepages were shown for 50 msec in two test phases in which participants rated the visual appeal (hereafter called appeal) of each image, presented one at a time and randomized every time they were shown. Ratings were made on a 9-point scale ranging from 1 (very unappealing) to 9 (very appealing). Response times were recorded automatically using DirectRT™. The backward masking procedure described in Section 2.1 was adopted.

2.3. Results

All correlation analyses in this article were computed using Pearson product moment correlation coefficients and are referred to simply as "correlations" from here on. The correlation of the mean ratings of the 50 homepages in test phases 1 and 2, collapsed across participants, was highly significant $r = .73$, $p < .001$. As in our earlier experiments [Fernandes et al. 2003; Lindgaard et al. 2006] in which no mask was used, ratings were thus quite consistent from the first to the second test phase. However, they were considerably lower here than in our earlier experiments in which values of up to $r = .97$ were achieved. An additional 20 correlations were then computed to assess

the consistency of each participant's own ratings in the two test phases. Of these, 18 were significant (13 at the $p < .001$, 4 at $p < .01$, and 1 at the $p < .05$ levels). Participants' opinions were thus quite consistent from one test phase to the next as shown in both the aggregate and individual ratings. In order to determine how much participants agreed with each other, another 190 correlations were calculated separately for each test phase, correlating each participant's ratings with those of all others (i.e., participant 1's ratings were compared with all 19 others, participant 2 with 18 others, as their ratings had already been compared with those of participant 1. Participant 3 was compared with the 17 remaining, and so on, resulting in 190 correlations). In test phase 1, 62/190 (32.60%) were significant; in test phase 2, 84/190 (44.20%) were significant. Finally, the mean response time was somewhat lower in test phase 2 (1563 msec) than in test phase 1 (1648 msec), but this difference was not significant ($t < 2$).

2.4. Discussion

The previous results show clearly that ratings were consistent between the two test phases even though they accounted for less variance than in our previous studies [Fernandes et al. 2003; Lindgaard, et al. 2006]. We have two possible explanations for this difference. First, there may have been large individual differences in perceptual speed [Mount et al. 2008]. As that was not assessed a priori, it is impossible to ascertain the potential influence it may have had. The topic is interesting and deserves further investigation. Second, the mask probably did affect the findings. Yet, even if it did, the preceding results are clearly well above chance levels. It is therefore safe to conclude that any effect of the mask would have been relatively modest. No mask was therefore employed in the next two experiments.

The mere exposure effect is said to begin to wane at a 50 msec exposure time. According to Zajonc, it is the "gateway to the subliminal" [2001, p. 224], and subliminal stimuli are destined to pass below the threshold of cognitive processing. Thus, Zajonc argues that appeal can be judged purely on the basis of affect. According to Jacobson and Höfel [2007a, 2007b] aesthetic judgments involve a "two-stage process consisting of an early impression formation and a later evaluative categorization" [2007b, p. 10]. As seen in the model that follows, the impression formation process involves visual perception of a stimulus. Jacobson and Höfel label the next set of processes "central," as their EEG recordings are indicative of Central Nervous System (CNS) activity. In order to reach a decision on stimulus appeal, these central processes are engaged, and if an explicit judgment is required as in the present study, it is reported. The authors were able to show that, once a decision has been reached, the reporting per se does not add to the response time. This is also reflected in our model that follows. Although the central processes are said to involve some "reflective thinking," the nature of that thinking is unclear as it was inferred from their EEG data. It may not preclude it from being affect based in situations involving very brief exposure times.

In order to explain our results, and in agreement with other visual processing models (e.g., Kosslyn [1989] and Winn [1994]), we argue that the perceptual process begins as soon as the retina registers the stimulus, very shortly after the stimulus onset (see, e.g., Best et al. [2006] and Van Rullen and Thorpe [2001]) as shown in Figure 2. In a masking paradigm as used here, this process of iterative stimulus-perception cycles continues only until the stimulus offset, as indicated by dashed arrows in the model. While the stimulus is still displayed, central processes, here loosely referred to as "reflection" and "how do I feel?" set in. According to Damasio's [1994] concept of somatic markers, "how I feel" is based on associations between reinforcing stimuli that induce an associated physiological affective state. When we have to make decisions that involve affect such as determining how much something appeals to us, the somatic markers created by the relevant stimuli are summed to produce a net somatic state. This overall state directs

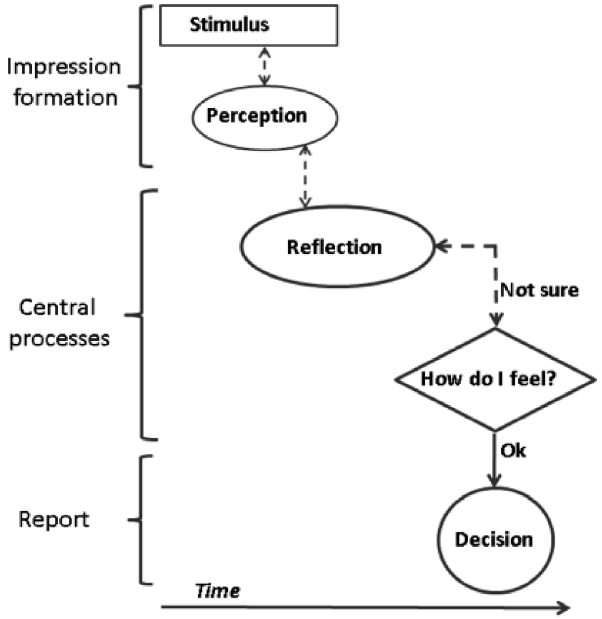

Fig. 2. A model of the processing assumed to be involved in an initial affective response.

(or biases) our decision of how to act. In the preceding model, reflection would thus involve the retrieval of a relevant somatic marker leading to the semantic translation of a physiological response into a feeling that culminates in a decision of how "good" that response "feels" (see also Wright et al. [2008]). In the present case, that decision is then expressed as an appeal rating.

3. EXPERIMENT 2

As our earlier proposed process model suggests, initial judgments of appeal are based on reflection on how "good" the given stimulus "feels." The ability to judge homepages reliably on characteristics other than appeal may require prolonged exposure to, and interaction with, the to-be-judged Web sites [De Angeli et al. 2006; Hassenzahl 2008] because assessments such as perceived usability and trustworthiness should involve additional cognitive processing. The question then is if anticipatory judgments of such stimuli, following very brief stimulus exposure times, can be made as reliably as judgments of appeal. If these other judgments are driven entirely by appeal, one would expect both ratings and response times for all three dimensions to be indistinguishable. Appealing homepages should also be perceived as more trustworthy and usable than their uglier counterparts. If, however, other judgment criteria are adopted for usability and trustworthiness, one would expect the added cognitive processing to be evident in two ways. First, ratings should be less reliable than for appeal, resulting in lower correlations between test phases 1 and 2. Second, response times should be longer because viewers would be less certain of their opinion than when judging visual appeal. Because we also wanted to assess the relative reliance on different visual attributes for the three kinds of judgments, it was important for participants to act as their own controls. Experiment 2 therefore adopted a within-subject design.

Table I. Correlations of Self- and Experimenter Entered Ratings in Test Phases 1 and 2

Dimension	Test Phase 1	Test Phase 2
Appeal	.91***	.91***
Trust	.89***	.93***
Usability	.88***	.87***

***$p < .001$

3.1. Method

A sample of 48 native English-speaking undergraduate university students took part. All had normal vision or corrected-to-normal vision, and no one was colorblind. Participants were tested individually in sessions lasting approximately 45 minutes and were paid $10 for participation. Apparatus and materials were the same as in Experiment 1. The procedure was the same also, with three exceptions. First, the mask was omitted. Second, participants rated the 50 homepages on each of three dimensions: appeal, perceived trustworthiness (hereafter called trust), and perceived usability (hereafter called usability), again shown for 50 msec each. All ratings were made on a scale ranging from 1 (very unappealing/untrustworthy/unusable) to 9 (very appealing/trustworthy/usable). The order in which the three dimensions were rated varied between participants in a Latin-square design, but participants rated the dimensions in the same order in each dimension's two test phases. That is, appeal was judged for each homepage in two successive test phases, and the same was true for trust and usability. Thus, participants saw the homepages a total of six times. Homepages were randomized every time they were shown. Finally, one half of the participants (n = 24) entered their ratings into the computer using a standard keyboard; the other half (n = 24) verbalized their ratings, which were entered by the experimenter, using the same keyboard as the participants who self-entered their ratings. These participants verbalized their ratings aloud. This was done to ensure that the self-entry requirement used in our earlier experiments would not systematically affect the ratings.

3.2. Results

3.2.1. Homepage Ratings. Participants' self-entered ratings were first correlated with experimenter-entered ratings by comparing the mean rating across all 50 homepages on each of the three dimensions (appeal, trust, and usability), in each of the two test phases (1 and 2), as shown in Table I.

The very high correlations demonstrate that the requirement to self-enter ratings into the computer did not affect these systematically. Therefore, ratings from all participants were pooled where possible in subsequent analyses.

Correlations comparing the mean ratings in test phases 1 and 2 within each dimension (appeal, trust, and usability) were then calculated, collapsed across the 48 participants. All three correlations were highly significant at the $p < .001$ level: appeal, $r = .95$; trust, $r = .96$, and usability, $r = .92$. Consistent with earlier findings in which only appeal was rated [Fernandes et al. 2003; Lindgaard et al. 2006; Tractinsky et al. 2004, 2006], judgments here were thus also highly consistent in all three dimensions from one test phase to the next.

To ensure that the these mean ratings were representative of the raw data, 48 correlations were carried out comparing each participant's first ratings with his/her own second ratings for each of the three dimensions. As can be seen in Table II, 47/48 participants' first and second ratings reached a statistically significant level on appeal, 43/48 on trust, and 42/48 on usability, demonstrating a high level of consistency within subjects.

Table II. Number of Significant Within-Participant Correlations Comparing Test Phases 1 and 2

Dimension	$p < .001$	$p < .01$	$p < .05$	Total
Appeal	41	4	2	47/48
Trust	31	8	4	43/48
Usability	38	3	1	42/48

Table III. Number of significant Correlations Indicating Agreement among Participants

Dimension	Test Phase 1	Test Phase 2
Appeal	646/1128 (57.27%)	678/1128 (60.06%)
Trust	467/1128 (41.01%)	625/1128 (55.41%)
Usability	529/1128 (46.90%)	609/1128 (54.99%)

Table IV. Correlations of Mean Ratings for the 50 Homepages in Test Phases 1 and 2

Correlation	Test Phase 1	Test Phase 2
Appeal & Trust	.62***	.84***
Appeal & Usability	.67***	.87***
Trust & Usability	.92***	.96***

***p < .001

The degree to which participants agreed with each other was then determined in the same manner as in Experiment 1. The number of significant correlations between each participant and the 47 others was counted for each dimension and each test phase (perfect agreement among all = 1128). Inspection of Table III shows that more correlations were significant at least at the p < .05 level for appeal than for trust and usability in both test phases, although this was more pronounced in test phase 1. It suggests that participants agreed more with each other on appeal than on trust and on usability. More correlations were significant in the second than in the first test phase for all three dimensions.

In order to determine if the mean ratings differed systematically between the three dimensions, a (2×3) repeated-measures ANOVA was performed for test phase (1, 2) and dimension (appeal, trust, usability), including all 48 participants. It yielded no significant differences ($F < 1$). Thus, there was no statistically traceable systematic difference in participants' decision strategies and hence no evidence for differences in the judgment strategies adopted for the three judgments. This suggests that the ratings were all driven by visual characteristics of the Web site, and hence indicating the presence of a halo effect.

We then paired the dimensions (appeal and trust; appeal and usability; usability and trust) to compare the mean ratings for the 50 homepages, again including all 48 participants. Table IV shows that all ratings were highly significant in both test phases, and that all were somewhat higher in test phase 2 than in test phase 1. The possibility that this increase represents a familiarity effect was explored later in Experiment 3. Table IV also shows that the correlations involving appeal ratings (appeal and trust; appeal and usability) were considerably lower than ratings of trust and usability (trust and usability), especially in test phase 1.

To lend credibility to these findings, we again paired the dimensions but this time for each of the 48 participants individually. Table V shows the number of correlations reaching significance level of at least p < .05 in both test phases. In test phase 1, 40/48 correlations for trust and usability were thus significant compared with 26/48 for appeal and usability, and 28/48 for appeal and trust. These numbers converged in test phase 2 in which more correlations were significant in all three calculations.

Table V. Number of Within-Participant Significant Correlations for Paired Ratings in Test Phases 1 and 2

Correlation	Test Phase 1	Test Phase 2
Appeal & Trust	26/48	42/48
Appeal & Usability	28/48	40/48
Trust & Usability	40/48	44/48

The mere exposure effect tends to strengthen as the number of stimulus exposures increases [Bornstein 1992]. In the present experiment, one would therefore expect ratings to increase across the six instances in which participants judged the homepages. This was tested with a one-way ANOVA and found to be nonsignificant ($F < 1$), which renders the evidence for a mere exposure effect in the present data somewhat questionable.

3.2.2. Response Time Analyses. We argued earlier that certainty in one's ratings should be evident in response times (Section 3). This would be traceable in two ways. First, one would expect shorter response times at either end of the rating scale than ratings falling in the center of the scale. Second, they should be longer for trust and usability due to the greater cognitive demands compared with response times for appeal. Taking the position on the rating scale first, Tractinsky et al.'s [2004] response time data did, in fact, reveal that ratings at the center of the rating scale took longer than ratings at either end of the scale. A similar analysis performed here for each of the three dimensions and for each test phase yielded no such relationship: all analyses were nonsignificant ($F < 1$). Thus, response times did not differ systematically as a function of the position of ratings along the rating scale.

In order to compare response times on the three types of judgment, the data were normalized using a Log10 transformation [Tabachnick and Fidell 2007] on the mean response times of the 24 participants who self-entered their ratings. An analysis of the resulting geometric mean response times was conducted using a (2×3) ANOVA with repeated measures for test phase (1, 2) and dimension (appeal, trust, usability). The highly significant main effect for test phase ($F(1, 49) = 165.42$, $p < .001$) showed that response times were shorter in the second than in the first test phase. The main effect for dimension was also significant, $F(2, 98) = 3.26$, $p < .05$. Pairwise comparisons showed that response times were longer for trust than for appeal ($p < .05$). This lends some support to the argument that judgments involving more cognitive processing should take longer than judging appeal which, according to our process model proposed in Experiment 1, would be based on reflection on how "good" the response "feels." Appeal was, however, not rated faster than usability ($p > .05$), which therefore fails to support that same argument. No response time differences were found between judgments of trust and usability either ($p > .05$).

3.3. Discussion

The aforesaid results show that judgments were highly consistent from one test phase to the next in all three dimensions, achieving very high correlations in aggregate as well as in comparisons of individual data. We argued earlier that differences in the judgmental strategies adopted for trust and usability would result in lower correlations than for appeal. Conversely, an absence of statistically traceable differences in the judgmental strategies in the three dimensions could indicate that all three types of judgments were driven predominantly by appeal. As the ratings failed to provide any statistical evidence of systematic differences between the three dimensions, that appears to be the case here. The findings are thus consistent with recent research suggesting that experimental tasks in which the outcome carries no serious consequence

for the viewer are likely to be dominated by appeal [Hartmann et al. 2007, 2008]. This is indicative of a halo effect.

The preceding data does, however, provide some, albeit rather weak, evidence suggesting that judgments of trust and usability may differ qualitatively from judgments of appeal. This is partly supported by the correlations of paired dimensions, and partly by the response time data. With respect to paired dimensions, the correlation for trust and usability ratings was considerably higher than the two correlations involving appeal (appeal/usability, appeal/trust), especially in test phase 1. With respect to response times, these were significantly shorter for appeal than for trust, suggesting that participants were more certain about judging appeal than trust, although these differences were not related to the position of ratings on the rating scale. No response time differences emerged for usability relative to those for appeal. One possibility is that participants believed that "beautiful" automatically translates into "usable," as some researchers have proposed [Norman 2004a; Tractinsky et al. 2000]. Indeed, one would expect a clean design as captured in Lavie and Tractinsky's [2004] "classical aesthetics" dimension, to be more appealing than clutter. The same would be true for an interesting, creative design over a dull, boring design. Because of the anticipatory nature of the previous usability judgments, it is not surprising that response times for the two dimensions were very similar.

The significantly shorter response times in the second than in the first test phase may be attributed to increasing familiarity with the stimuli and with the task. They may thus represent a simple practice effect which can occur when a person is retested on the same stimuli. Although the odd item may be remembered, the increase in scores across the two test phases here are, according to Kaufman [1994], not due simply to recall but to familiarity with the task and the development of a particular judgment strategy. This possibility that participants developed a uniform judgment strategy that they then applied to all three types of judgment could account for the higher correlations of paired dimensions (i.e., appeal/usability, appeal/trust, usability/trust) in the second than in the first test phase. In an effort to address this potential limitation, a between-subject design was adopted in Experiment 3.

4. EXPERIMENT 3

The purpose of Experiment 3 was to avoid the potential practice effect alluded to in Experiment 2. The stimuli were thus tested on each of the three dimensions as before, but in a between-subject design including only one test phase rather than two as before. It was predicted that response times would be longer for trust and usability than for appeal ratings but that the ratings would not differ between the dimensions, as these were again expected to be driven largely by appeal.

4.1. Method

A sample of 30 native English-speaking undergraduate volunteer students meeting the same criteria as before took part. They were tested individually in sessions lasting approximately 10 minutes. Apparatus, materials, practice items, and experimental stimuli were the same as in Experiments 1 and 2, again shown for 50 msec each, and homepages were randomized differently every time they were shown. Rating scales were the same as in Experiment 2. Each participant completed one test phase in which all 50 stimuli were rated, and only on one of the three dimensions, namely appeal ($n = 10$), trust ($n = 10$), or usability ($n = 10$).

4.2. Results and Discussion

A one-way ANOVA comparing the mean ratings for all 50 homepages on the three dimensions, collapsed across participants, was conducted first. Consistent with the

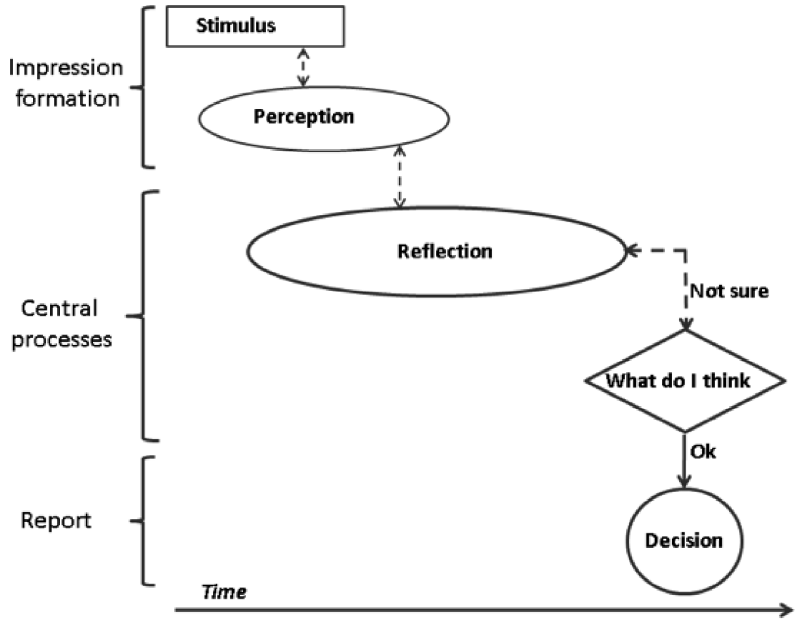

Fig. 3. The revised model accounting for judgments of trust and perceived usability.

findings of Experiment 2, it was not significant ($F < 1$). There was thus no statistical evidence suggesting that participants systematically adopted different criteria for judging appeal, usability, or trust. Instead, the ratings again reflected a halo effect, simply because there is no objective way to judge trust or usability without interacting with the application.

A comparison of the degree to which participants agreed with each other, calculated as described in Experiment 1, revealed that 28 (62.22%) of the 45 possible correlations were significant for appeal, 38/45 (84.44%) for trust, and 22/45 (48.89%) for usability.

The one-way ANOVA conducted for mean response times was highly significant, $F(2, 147) = 25.50$, $p < 001$, thus the data were not transformed. A post hoc Dunnetts C test showed that these were shorter for appeal than for trust ($p < .001$) and also shorter than for usability ($p < .001$), but that the response times for trust and usability ratings did not differ from each other ($p > .05$). This provides further support for the contention that trust and usability ratings may have required more, or at least different, processing than appeal. Trust and usability apparently shared some qualitative characteristics that differed from those typical for appeal, even though such differences were not evident in the ratings.

The model proposed in Experiment 1 allows us readily to accommodate the longer response times for trust and usability, as shown in Figure 3. Assuming a paradigm that does not employ a mask, processing of the homepage would extend beyond the stimulus offset time as indicated by the longer perceptual process compared with the previous model. In order to reach a somewhat more considered judgment for these two dimensions that impose a greater cognitive demand than when judging appeal, reflection time is also increased. Here, the reflection culminates in a distinctly cognitive decision (what do I think?) instead of the affective "how do I feel?" that characterized the appeal judgment shown in Figure 2 earlier.

Yet, since no interaction was allowed with the Web sites here, participants would still have to rely on appeal even for those more considered judgments, which would

explain the absence of differences in the ratings on the three dimensions and hence the resulting halo effect.

In an effort to begin to identify the attributes that appear to contribute to ratings of appeal, trust and usability in the sample of Web pages used in the present experiments, a more detailed analysis of visual attributes was performed next.

5. ANALYSIS OF ATTRIBUTES

Several researchers have attempted to identify attributes that contribute to visual appeal in Web sites. For example, Kim and his colleagues [2003] identified features that, when combined in certain ways, appeared predictably to evoke different emotions, and Sutcliffe [2001] proposed a set of seven heuristics for achieving attractiveness in Web design. Others have produced various lists of design elements that seem to affect the enjoyment of a Web site (e.g., Badre [2000], Callahan [2006], Cyr et al. [2006], Park et al. [2005], and Sun [2001]), but it is still not clear exactly what design elements combine to yield reliable visual appeal. To the extent that Lindgaard et al.'s [2006] interpretation of their data as a mere exposure effect was correct, it should not be possible for people to discern specific design elements such as icons, fonts, or the placement of search fields after seeing a stimulus for only 50 msec. However, the most salient aspects of a homepage such as background color [Sutcliffe 2001; Garber et al. 2008] and the high-level spatial arrangement of objects such as symmetry [Jacobson and Höfel 2003; Sutcliffe 2001], balance [Lauer and Pentak 2002; Park et al. 2005] and the amount of white space [Zdralek 2003] may contribute to that first holistic impression.

Anecdotal inspection of the background colors in the sample of homepages used here revealed only that the 10 images receiving the lowest ratings in all three dimensions predominantly had a white background whereas colors of the 10 highest rated were all quite bright and saturated. In an effort to determine the possible influence the spatial arrangement of objects and the ratio of text or graphics to background may exert on the kinds of ratings obtained here, we relied on the work by Park and her colleagues [2005]. They identified and defined a set of 11 visual attributes, which in different combinations were found to influence the perceived e-brand personality of homepages. The visual attributes they investigated were well defined and firmly grounded in the literature, which makes it possible to assess almost any homepage against these. We wanted to know if different attributes and combinations of attributes could be related specifically to higher appeal, trustworthiness, or perceived usability. Because of the large variety of homepages in our sample, it was possible to assess these meaningfully according to only four of Park et al.'s [2005] 11 attributes, namely symmetry, density, balance, and contrast. Since the ratio of graphics to text and white space varied substantially (hereafter called graphics), as did the amount of text to background (hereafter called text), we also assessed the homepages according to these two attributes. Thus, six visual attributes were explored. A definition is given of each in Appendix 1.

5.1. Method

Each of the 50 homepages was printed in color and affixed to a card together with the mean rating for each of the three dimensions in test phase 1, Experiment 2. These were selected because the same participants made all three types of judgment and also because the distribution of most- and least appealing images was very similar in the two experiments. The six visual attributes were then assessed one at a time, and separately for each of the three dimensions. Two researchers rated each homepage on a 5-point scale on every attribute, ranging from 1 = "not at all" to 5 = "vast" (amounts). Disagreements were settled by negotiation.

Table VI. Homepage ID, Highest Mean Appeal Ratings, with Attribute Ratings of Density, Contrast, and Graphics

Homepage ID	Mean Appeal Rating	Density	Contrast	Graphics
16	**6.62**	**4**	**3**	**5**
28	**6.48**	**4**	**3**	**5**
23	6.46	3	4	5
18	6.42	2	2	5
12	6.40	3	2	5
14	6.35	2	4	4
10	6.27	2	4	5
6	6.15	3	4	5
8	6.00	3	3	4

Fig. 4. Two images rated among the highest on appeal (ID 16, 28, shown in italics).

5.2. Data Analysis

In order to explore how well ratings of appeal, trust, and usability could be predicted from the six visual attributes assessed, a series of cluster analyses were performed for each dimension. The analyses were performed using PAST, (Version 1.32), a program supporting multivariate nonparametric cluster analysis [Hammer and Harper 2005]. Paired linkages with Euclidean distances were measured in all comparisons. The results of the analyses are shown in a dendogram which groups similar Web pages together based on similarity of the visual attributes measured. All six visual attributes were included in the first round. Each attribute was then removed, one at a time, while exploring the remaining five attributes. Then two were dropped while exploring the remaining four attributes, and so on, until all combinations had been explored. The same procedure was followed for each of the three dimensions to obtain the best combination for predicting appeal, trust, and usability ratings, respectively. The best split between low- and high-rated homepages was retained. The results of the best analysis for each of the three dimensions are discussed next.

5.3. Results

5.3.1. Visual Appeal. The visual attributes of density, contrast, and graphics revealed the best cluster predicting appeal ratings. Table VI shows the relationships between these and the highest mean appeal ratings obtained in test phase 1 of Experiment 2, collapsed across all participants. The dendograms revealing the data clusters are shown and described for each dimension in Appendix 2.

Figure 4 shows the two bold faced images in Table VI. These were typical of the most appealing homepages in the sense that they demonstrate that amount of text, symmetry, and balance did not play an important role in distinguishing between the lowest and highest rated pages.

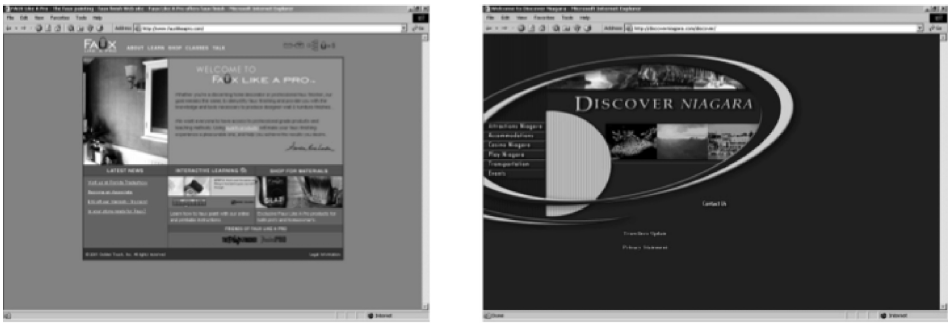

Fig. 5. Two images rated among the highest on trust (ID 23, 18, shown in italics).

Table VII. Homepage ID, Highest Mean Trust Ratings, with Attribute Ratings of Balance, Contrast, Text, and Graphics

Homepage ID	Mean Trust Rating	Balance	Contrast	Text	Graphics
23	6.63	4	4	4	5
18	6.48	5	2	3	5
28	6.46	4	3	4	5
20	6.44	3	5	5	2
17	6.02	5	2	3	4
7	6.02	5	4	4	5

One comment is worth making about image # 18 which advertises the Niagara region in Canada and also featured among the most highly rated for trust. Although an attempt had been made to select homepages that our Canadian participants would have been unlikely to encounter before, some of them may inadvertently have seen that particular image, shown to the right in Figure 5. While previous exposure could have influenced the ratings, it is noteworthy that, when the same homepages were tested on a sample of non-English speaking students in Taiwan and using the same procedure, the average appeal rating for that image, shown here for test phase 1, was very similar: Canada mean 6.48/9, Taiwan mean 6.38/9. It was among the most appealing images in both samples. It is thus reasonably safe to conclude that any bias potential that prior exposure may have caused can be considered negligible.

5.3.2. Trust. Table VII shows that homepages deemed most trustworthy contained abundant graphics, had good balance, and a moderate amount of text and contrast, save for ID 20, which was lower on contrast. Density and symmetry played no role in distinguishing the most from the least trustworthy pages. Figure 5 shows the two examples bold faced in the table.

Note that the two highest rated homepages were also among the most appealing homepages (see Table VI).

5.3.3. Usability. Table VIII shows that pages high in graphics and balance with low-to-moderate contrast were perceived to be most usable, and Figure 6 shows two typical examples, bold faced in the table. For usability ratings, density, text, or symmetry played no role in distinguishing the highest rate pages from the lowest rated.

It is interesting to note that homepages #18 and #23, which both attracted some of the highest appeal and trust ratings, again appeared among the highest usability ratings. Homepage #28 was also among the highest ratings on all three dimensions, while homepage #12 appeared among the most appealing and most usable and homepage #17 among the highest ratings for trust and usability. This overlap further suggests that ratings were dominated by appeal but also that there were some differences between

Table VIII. Homepage ID, Highest Mean Usability Ratings, with Attribute Ratings of Balance, Contrast, and Graphics

Homepage ID	Mean Usability Rating	Balance	Contrast	Graphics
18	6.69	5	2	5
15	6.56	3	3	5
28	6.44	4	3	5
12	6.15	5	2	5
17	6.10	5	2	4
23	6.10	4	4	5

Fig. 6. Two images rated among the highest on usability (ID 15, 12, shown in italics).

Table IX. Summary of Visual Attributes Characterizing the Highest Rated Images in Each Dimension

Dimension	Balance	Contrast	Density	Graphics	Symmetry	Text
Appeal	—	Mod	Mod	High	—	—
Trust	Mod-high	Mod	—	High	—	Mod-high
Usability	High	Mod	—	High	—	—

the three kinds of rating. Note also the rather saturated background color of all the six images shown before which contradicts Web design guidelines that recommend pastel colors. It is possible that people apply different judgmental criteria to a homepage than to entire Web sites, and that a preference for strong colors may only apply to the initial formation of appeal.

Table IX summarizes the preceding findings for the three dimensions, discussed in the next section.

5.4. Discussion

The just presented findings suggest that some visual attributes are perceptible in the initial, holistic Gestalt impression; they are integral to the perceptible pattern recognition in Norman's model at level one. An abundance of graphics and a moderate level of figure/ground contrast contributed to the highest mean ratings in all three dimensions. It is reasonable to assume that graphics on a homepage would typically be added for cosmetic rather than for informative purposes, thereby contributing to a positive first impression. The preference for a moderate level of contrast, which would contribute to the more utilitarian aspects of usability, is perhaps not surprising given the nature of the task. In accordance with screen design guidelines (e.g., Galitz [2007]), one would have expected preference for high figure/ground contrast, especially for the usability ratings. This may have been the case if the task had required retrieval of information from the Web sites. Clearly, the impact of individual visual attributes on both performance and preference after interaction with a Web site needs to be investigated. Preference for a moderate level of density contributed only to appeal ratings, possibly

again because participants did not navigate the site or look for specific information. Most of the images given the lowest ratings in all three dimensions were quite cluttered, whereas the design of most of the highly rated images tended to be much cleaner, therefore looking simpler. This is consistent with the notion of "classical aesthetics," affecting both visual appeal and perceived usability. Recall also that Sillence and her colleagues' [2006] found cluttered Web sites to be rejected very quickly. Apparently, the amount of "stuff" on an image contributed to the previous ratings. Interestingly, inspection of the images shown in the preceding section shows clearly that neither balance nor symmetry contributed to appeal. It is possible that the importance of these two attributes may change with longer viewing times, in line with guidelines for good Web page design [Brady and Phillips 2003; Park et al. 2005; Sutcliffe 2001], and consistent with very robust empirical findings (e.g., Baudouin and Tiberghien [2004], and Jacobson and Höfel [2002]). As Lauer and Pentak [2002] argue, balance may lend a homepage a calming look, which was apparently important for judging trust and usability but not appeal in the present context. Possible reasons for this include the fact that the sample of homepages was very heterogeneous and also that Web site genres were not controlled for. One may speculate that the purpose of a Web site could influence the degree to which inherent calmness in the design may affect its appeal. That issue still needs to be researched. Interestingly, the background color of most of the most highly rated homepages was quite bright and saturated, in contrast to Sutcliffe's recommendation that "low saturation pastel colors should be used for backgrounds" [Sutcliffe 2001, p. 189]. This suggests that people's taste may have changed since the publication of Sutcliffe's generic heuristics and certainly that more research needs to identify when more saturated background colors are deemed appropriate.

Since all the participants were young, computer-savvy adults, one may assume that they would have made decisions about trust and perceived usability of Web sites in their normal Web activities. They may therefore have come to these experiments with some implicit, albeit still unstable, criteria on which to base such judgments. Some of these are evidently design-related attributes including balance and amount of text on a page, representing Galitz's [2007] and others' good design principles, as well as Lavie and Tractinsky's [2004] classical aesthetics and Hassenzahl's [2008] normative values. Others, including the presence of graphics, represent the notion of expressive aesthetics or experiential values. While all may contribute to a feeling of pleasure that varies in intensity according to the blend in a given homepage, this alone cannot explain the differences between ratings of appeal and those of trust/usability.

6. GENERAL DISCUSSION

Summarizing the earlier findings briefly, Experiment 1 demonstrated that people judge visual appeal reliably even when processing is disrupted beyond the stimulus exposure time. Experiment 2 confirmed that people judge the appeal as well as trust and usability of homepages consistently upon viewing them for 50 msec. It also provided some evidence suggesting that trust and usability judgments may differ qualitatively from judgments of appeal. The statistically reliable differences in response times, especially in Experiment 3, confirm the claim that judgments requiring additional cognitive processing may differ qualitatively from appeal judgments when these represent the first impression of a homepage. Interestingly, differences in judgmental strategies were not evident in the ratings, suggesting that all three kinds of judgment tested here were predominantly driven by appeal. While the preceding results all suggest the presence of a halo effect, we also argued earlier that the observed differences that appear to distinguish trust and usability from appeal judgments cannot be discerned from the intensity of a feeling of pleasure alone. This argument is supported by three findings. First, the bivariate analyses comparing trust and usability in Experiment 2 resulted in a higher

correlation than those involving appeal (Tables IV and V). Second, appeal judgments were made significantly faster than the other two types in Experiment 3, although this could also simply be due to less well-defined judgment criteria for the latter. Finally, the differences in the visual attributes found to influence the three types of judgment shown in the cluster analyses also support that argument. It is very encouraging that our attempt to begin to identify some of the visual attributes that combine to render a homepage initially appealing, trustworthy, and seemingly usable suggests differences in the combinations of these. Despite the existence of several rating scales and design guidelines recommending how to accomplish appealing Web sites, the task of teasing out the importance of various visual attributes to achieve different effects has barely begun.

There is widespread agreement that aesthetic appraisal is a complex process that evolves in several phases (e.g., Clore et al. [2001], Lindgaard and Whitfield [2004], Pickford [1972], Lavie and Tractinsky [2004], Norman [2004a, 2004b], and Hassenzahl [2008]), but the boundaries that separate one phase from the next are unclear except perhaps in Jacobson and Höfel's work [2003, 2007a, 2007b] in which they refer to it as a two-stage process. The first "perception" stage is comparable to Zajonc's [2001] notion of subliminal stimulation, to Wright et al.'s [2008] notion of the "sensual thread" as an integral aspect of their "holistic approach to experience" and also to level one in Norman's [2004b] model. According to Norman, the

> "visceral level is perceptually based, given rise to immediate judgments: good or bad, safe or dangerous. . . If there is any level at which beauty is associated with the object itself, it is at the visceral level, for we have evolved to judge some objects or experiences positively, others negatively. This, however, is a beauty that is truly only skin deep, for at this level of processing, it is only what can be immediately perceived that matters: the only processing possible is simple pattern recognition through innate mechanisms" [Norman p. 314].

It represents an initial attraction referred to as the "aesthetics of appearance" [Desmet and Hekkert 2007; Hallnäs and Redström 2002; Hartmann et al. 2007; Redström 2008], that gives rise to the formation of a "general attitude." This initial impression is only refined with increased usage of the product or Web site [De Angeli et al. 2006], leading to the aesthetic, or "pleasure" of using (e.g., Buxton [2005], and Jordan [2001]).

In Jacobson and Höfel's model as in ours, focusing exclusively on the "aesthetics of appearance," perception is separated from the "central" processes in which reflection on the feeling the image or object culminates in a decision of how "good" it feels, at least in judgments of appeal. This process concurs with Norman's [2004b] model of emotion. According to Norman,

> "any spoken or conscious assessment of visceral responses must come from the reflective level, which means it has been subjected to possible interpretation, modification, and rationalization . . . it is only at the reflective level that full-fledged emotions reside, and this level is conscious, intellectually driven" [Norman 2004b, p. 315].

It is not entirely clear what exactly Norman means by "full-fledged" emotions as opposed to the "reflection" component in Jacobson and Höfel's model as in ours. In Norman's model, any articulation of a feeling takes us to his level 3, thereby contradicting Zajonc [1980, 2000, 2001] whose research suggests that visual appeal ratings need not involve cognitive reflection. The argument about the notion of "reflection" seems to be reducible to the definition of "cognitive reflection." In order to make a decision and assign a rating or a preference statement to an image or an object, arguably some kind of reflection is necessary even to decide on the intensity of a good feeling. Whether or when this reflection should be called "affective" or "cognitive" may, in the end, not matter, even if different kinds of judgments are found to activate different areas of the

brain. The preceding data suggests that the speed with which a decision is made in response to an immediate impression depends on the nature of the required judgment. There would be no need for Jacobson and Höfel's [2007a, 2007b] model to accommodate a special category of "affective reflection" in their "central processes." One possible way to reconcile Norman and Zajonc is that the mere exposure effect may begin to wane at viewing times that are even shorter than 50 msec. The fact that we found no evidence for a mere exposure effect in Experiment 2 leads us to suspect that it might be the case.

The observation that people are able to very quickly and reliably make judgments that differ in the amount of cognitive processing we think is important for Web designers to know. It is also important to note that different kinds of first-impression judgments appear to be driven largely by visual appeal, even if these rely on different visual attributes. Furthermore, there appear to be qualitative differences in the visual attributes to which people pay attention, and also in the amount of processing people engage in as a function of the task they are asked to complete. Although some Web design guidelines make specific recommendations concerning Web site "look and feel," research determining the judgmental strategies as well as the appropriateness of different constellations of visual attributes is still in its infancy.

6.1. Limitations and Further Research

Our data suggests that an anticipatory relationship exists among the three dimensions tested here, but they cannot reveal how it affects subsequent interaction. Nor can we determine how long the initial appeal may continue to influence other types of judgment. In order to improve the ecological validity of our research, this influence will be tested in future research in which visual appeal and objective usability will be varied systematically in selected Web sites. As the stimuli tested here were very heterogeneous, they avoided ratings being subject to judgments of stylistic appropriateness that may occur when rating, for example, news sites or banking sites. However, this strength is also a weakness in the sense that the stimuli may have inspired participants to rely more on appeal than may have been the case if examples representing one particular genre had been employed instead. This may have affected the outcomes of the cluster analyses. Similar analyses will therefore be completed on more homogeneous samples in the future. The scale used to categorize visual attributes of the homepages needs to be much more accurately defined, and more stringent judgment criteria need to be applied. We also suspect that the limited set of visual attributes included here is incomplete. This will be investigated further. Intra- and interindividual models of aesthetic judgments developed by Jacobson [2004], and models identifying the cues participants relied on most in another study of aesthetic processing [Jacobson and Höfel 2007a] revealed substantial differences between people. Thus, individual differences need to be studied in much more detail.

7. CONCLUSION

Obviously, participants cannot gain a realistic picture of objective usability or reach a true sense of trustworthiness merely from looking briefly at a homepage. One purpose with the preceding experiments was to determine if a relationship can be demonstrated between appeal and other types of judgment that would require additional processing in an experimental paradigm offering only a brief glimpse of each image and no opportunity to interact with the product. Although there are still many issues to be settled, the previous results appear to confirm such a relationship. As visual appeal clearly dominates first-impression judgments of other characteristics such as perceived usability and trustworthiness, which are thus subject to a halo effect, it does appear that judgments requiring additional processing are dealt with in a qualitative

different manner from those of visual appeal. While lending cautious support to Tractinsky et al.'s [2000] bold assertion that what is beautiful is [perceived to be] usable, the big question of dependence or independence of visual appeal, trustworthiness, and perceived usability remains unresolved. Disentangling these complex relationships remains a very interesting challenge to the HCI community.

Our second purpose was to begin to identify visual attributes that may contribute to a homepage appearing visually appealing, usable, and trustworthy. The results are encouraging in the sense that we did identify some differences in the kinds of attributes that people appear to note when judging homepages on different dimensions. In order to develop truly useful Web design guidelines, it will be essential to articulate both the cues that people rely on for different kinds of judgment and the strategies they employ when making decisions based on immediate as well as on prolonged exposure to Web sites.

APPENDIX 1: DEFINITIONS FOR THE VISUAL ATTRIBUTES APPLIED IN THE CLUSTER ANALYSES

Balance: The distribution of visual weight across the whole picture [Lauer and Pentak 2007; Ngo and Byrne 2001; Park et al. 2005]. Visual weight is influenced mainly by the sizes, colors, and locations of objects [Ngo and Byrne 2001; Wong 1987; Park et al. 2005]. The numeric weight of balance is maximized when the visual weights of elements on the right and left sides or in the upper and lower regions on a screen are equal. The 5-point scale ranged from 1=completely unbalanced (all text and graphics were presented in one quadrant of the page and the rest was empty), via 3=somewhat balanced (represented pages where text and graphics were dispersed mainly amongst two or three quadrants but not all of them, and 5=completely balanced (weight of text and graphics in each quadrant was very close to equal).

Contrast: The degree of difference between elements [Wong 1987; Park et al. 2005]. The numeric value of contrast is determined by the degree of difference among objects in size, color, and location [Wong 1987; Park et al. 2005]. This dimension ranged from 1=very low contrast (objects and text cannot be distinguished clearly from each other), via 3= medium contrast (objects and text can be distinguished from each other but no stark relations exist), to 5=contrast is at its highest (black on white background or crisp and bright images which appear to jump off the page).

Density: The ratio between the area of the background (i.e., the area not covered by objects) and the area covered by objects [Behrens 1984; Koffka 1955; Ngo and Byrne 2001]. The numeric value of density increases as the sum of the areas of the objects rises and the area of the background diminishes [Ngo and Byrne 2001]. A rating of 1=relatively empty page (much more background than text or graphics), via 3=moderate amount of visual objects (equal balance of text and or graphics to background), to 5=relatively full page (much more text and graphics than background).

Graphics: The ratio of graphics to background was assessed using a 5-point scale. A rating of 1=no graphics (text on background only), via 3=small graphics (graphics are somewhat equally represented against background), to 5=graphics take up the most amount of space possible (in relation to background; in these cases the graphic is the background).

Symmetry: The mirroring of the visual composition across a vertical or horizontal pivot line [Ngo and Byrne 2001; Park et al. 2005].

Text: The amount of text on a homepage ranged from 1 = no or very little text (up to about 10 words in total) via, 3=moderate amount of text (the equivalent of two to three sentences), to 5=vast amount of text (more than two or three paragraphs of text).

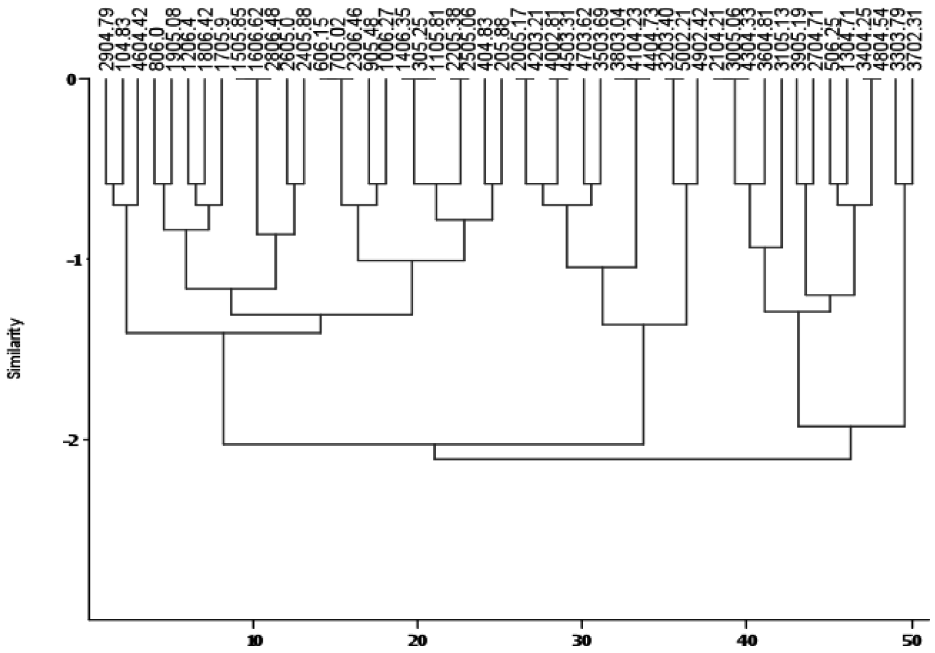

Fig. 8. Dendogram of homepages (1-50 on abscissa) and similarity on the ordinate.

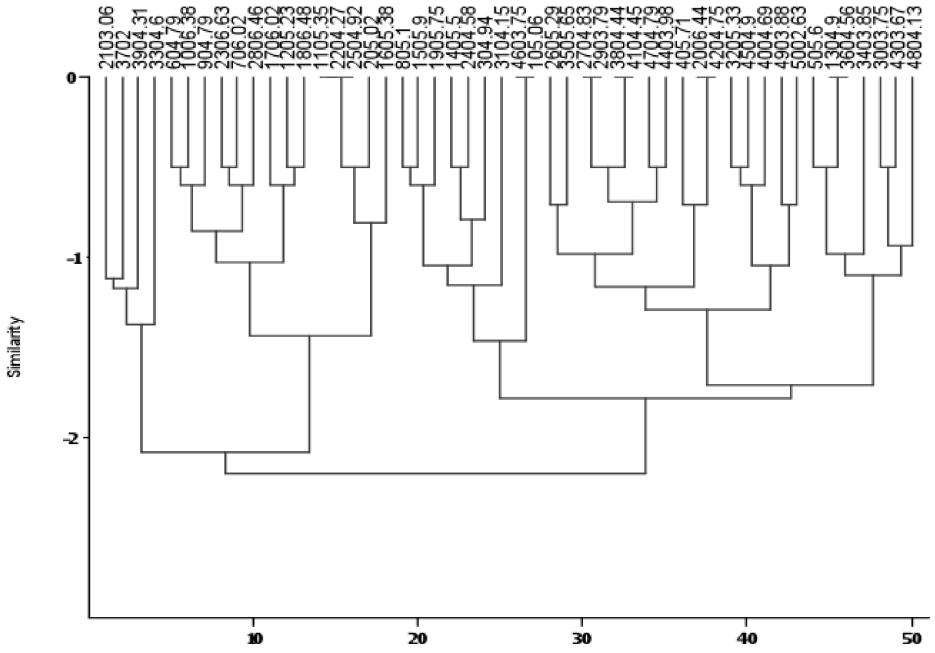

Fig. 9. Dendogram of homepages (1-50 on abscissa) and similarity on the ordinate.

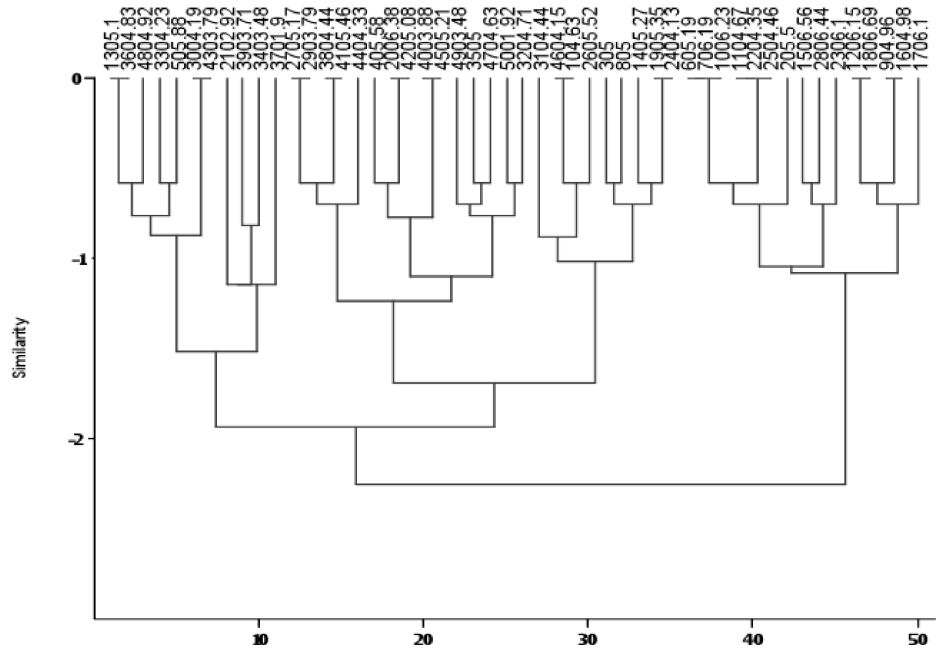

Fig. 10. Dendogram of homepages (1-50 on abscissa) and similarity on the ordinate.

APPENDIX 2: DENDOGRAMS FOR CLUSTER ANALYSES OF EACH TYPE OF JUDGMENT

(A) VISUAL APPEAL

Figure 8 shows the clusters of homepages and their associated appeal ratings; numbers on the clusters represent the homepage ID followed by a "0" and the mean visual appeal rating to two decimal places.

Note: By looking at the bottom of the dendogram it is clear that there are two separate groups of pages at about the -2 similarity mark on the ordinate. The two rightmost clusters of pages are more similar to each other than they are to the two leftmost clusters of pages. The separation between the two groups of pages occurred at approximately page 25. This is evident when looking toward the middle of the dendogram where the four clusters of pages are grouped at about the -1 mark on the ordinate.

(B) TRUSTWORTHINESS

Figure 9 shows the clusters of sites and their associated trustworthiness ratings; numbers on the clusters represent the homepage number followed by a "0" and the mean trustworthiness rating to two decimal places.

(C) PERCEIVED USABILITY

Figure 10 shows the clusters of sites and their associated perceived usability ratings; numbers on the clusters represent the homepage ID followed by a "0" and the mean perceived usability score to two decimal places.

ACKNOWLEDGMENTS

We would like to thank T. Triggs for reminding us about the iconic-trace issue. Our thanks also go to A. Monk, A. Parush, and three anonymous reviewers for very helpful comments on an earlier draft of this article.

REFERENCES

AJZEN, I. AND FISHBEIN, M. 1980. *Understanding Attitudes and Predicting Social Behavior.* Prentice Hall, Englewood Cliffs, NJ.

ANDERSON, N. H. 1980. *Foundations of Information Integration Theory*. Academic Press, London.

ANDERSON, N. H. 1981. *Methods of Information Integration Theory*. Academic Press, London.

AVERBACH, E. AND SPERLING, G. 1961. Short term storage of information in vision. In *Information Theory*, C. Cherry Ed., Butterworth, London, 196–211.

BADRE, A. 2000. The effects of cross cultural interface design orientation on World Wide Web user performance. GVU Tech. rep. http://www.cc.gatech.edu/gvu/reports/2001/.

BASSO, A. GOLDBERG, D., GREENSPAN, S., AND WEIMER, D. 2001. First impressions: Emotional and cognitive factors underlying judgments of trust in e-commerce. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*. ACM Press, New York, 143–147.

BAUDOUIN, J. Y. AND TIBERGHIEN, G. 2004. Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychol. 117*, 3, 313–332.

BEZANSON, W. 2007. *The Life of Phi: Beauty and the Golden Ratio.* Baico Publishers, Ottawa, ON, Canada.

BEHRENS, R. R. 1984. *Design in the Visual Arts.* Prentice-Hall, Upper Saddle River, NJ.

BEN-BASSAT, T., MEHER, J., AND TRACTINSKY, N. 2006. Economic and subjective measures of the perceived value of aesthetics and usability, *ACM Trans. Comput.-Hum. Interact. 13*, 2, 210–234.

BEST, L., HUNTER, A. C., AND STEWART, B. M. 2006. Perceiving relationships: A physiological examination of the perception of scatterplotts. In *Diagrams*, D. Barker-Plummer et al. Eds., Springer, Berlin, 244–257.

BLOCH, P. H., BRUNEL, F. F., AND ARNOLD, T. J. 2003. Individual differences in the centrality of visual product aesthetics: Concept and measurement. *J. Consum. Res. 29*, 551–565.

BORNSTEIN, R. F. 1992. Subliminal mere exposure effects. In *Perception Without Awareness: Cognitive, Clinical, and Social Perspectives*, R. F. Bornstein and T. S. Pittman Eds., The Guilford Press, New York, 191–210.

BRADY, L. AND PHILLIPS, C. 2003. Aesthetics and usability: A look at color and balance. *Usabil. News 5*, 1, 1–4.

BREITEMEIER, B. G. AND OGMEN, H. 2000. Recent models and findings in visual backward masking: A comparison, review, and update. *Percept. Psychophys. 62*, 8, 1572–1595.

BUXTON, B. 2005. Experience design vs. interface design. *Rotman Mag.* Winter, 47–49.

CALLAHAN, E. 2006. Cultural similarities and differences in the design of university Web sites. *J. Comput.-Mediat. Comm. 11*, 239–273.

CARROLL, J. M. AND MENTIS, H. M. 2008. The useful interface experience: The role and transformation of usability. In *Product Experience*, H. H. J. Schiffstein and P. Hekkert Eds., Elsevier, North-Holland, the Netherlands, 499–515.

CHESKIN RESEARCH GROUP. 1999. eCommerce trust study. http://www.cheskin.com/think/studies/ecommtrust.html.

CLORE, G. L., WYER, R. S. JR., DIENES, B., GASPER, K., GOHM, C., AND ISBELL, L. 2001. Affective feelings as feedback: Some cognitive consequences. In *Theories of Mood and Cognition: A User's Guidebook*, L. L. Martin and G. L. Clore Eds., Lawrence Erlbaum Associates, Mahwah, NJ, 27–62.

CYR, D., HEAD, M., LARIOS, H. AND PAN, B. 2006. Exploring human images in website design across cultures: A multi-method approach. In *Proceedings of the 5th Annual Workshop on HCI Research in MIS*. ICIS, 55–59.

CYR, D. AND TREVOR-SMITH, H. 2004. Localization of Web design: An empirical comparison of German, Japanese, and U.S. website characteristics. *J. Amer. Soc. Inf. Sci. Technol. 55*, 13, 1–10.

DAMAISO, A. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain.* Penguin Books, London.

DAMAISO, A. R. 2000. A second chance for emotion. In *Cognitive Neuroscience of Emotions*, R. D. Lane and L. Nadel Eds., Oxford University Press, New York.

DAVIS, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quart. 13*, 3, 319–340.

DE ANGELI, A., SUTCLIFFE, A., AND HARTMANN, J. 2006. Interaction, usability and aesthetics: What influences users' preferences? In *Proceedings of the Conference on Designing Interactive Systems (DIS)*. ACM Press, New York, 271–280.

DESMET, P. AND HEKKERT, P. 2007. Framework of product experience. *Int. J. Des. 1*, 1, 1–12.

DIEFENBACH, S. AND HASSENZAHL, M. 2008. Give me a reason: Hedonic product choice and justification. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. ACM Press, New York, 3051–3056.

DIEFENBACH, S. AND HASSENZAHL, M. 2009. The beauty dilemma: Beauty is valued but discounted in product choice. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM Press, New York, 1419–1426.

DION, K., BERSCHEID, E., AND WALSTER, E. 1972. What is beautiful is good. *J. Person. Social Psychol. 24*, 285–290.

DUMAINE, B. 1991. Design that sells and sells and... *Fortune*. March 11, 42–49.

EDDY, D. M. 1999. Probabilistic reasoning in clinical medicine: Problems and opportunities. In *Judgment Under Uncertainty: Heuristics and Biases,* D. Kahneman et al. Eds., Cambridge University Press, Cambridge, UK.

EGGER, F. N. 2002. Consumer trust in e-commerce: From psychology to interaction design. In *Trust in Electronic Commerce: The Role of Trust from a Legal, an Organizational, and a Technical Point of View,* J. E. J. Prins et al. Eds., Kluwer Law International, Norwell, MA. 11–44.

ELLIOTT, R. AND DOLAN, R. J. 1998. Neural responses during preference and memory judgments for subliminally presented stimuli: A functional neuro-imaging study. *J. Neurosci. 18*, 4697–4704.

ENNS, J. T. AND DI LOLLO, V. 2000. What's new in visual masking? *Trends Cogn. Sci. 4*, 9, 345–352.

FERNANDES, G., LINDGAARD, G., DILLON, R., AND WOOD, J. 2003. Judging the appeal of Web sites. In *Proceedings of the 4th World Congress on the Management of Electronic Commerce.*

FOGG, B. J., KAMEDA, T. BOYD, J., MARCHALL, J., SETHI, R., SOCKOL, M., AND TROWBRIDGE, T. 2002. Stanford-Markovsky Web credibility study: Investigating what makes Web sites credible today. Tech. rep., Persuasive Technology Lab, Stanford University, CA.

FROHLICH, D. M. 2004. Beauty as a design prize. *Hum.-Comput. Interact. 19*, 359–386.

GALITZ. W. G. 2007. *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques.* Wiley Publishing, Indianapolis, IN.

GARBER, L., HYATT, E. M., AND BOYA, U. O. 2008, The mediating effect of the appearance of nondurable consumer goods and their packaging on consumer behavior. In *Product Experience*, H. H. J. Schiffstein and P. Hekkert Eds., Elsevier, North-Holland, the Netherlands. 581–602.

GARRETT, J. J. 2003. *The Elements of User Experience: User-Centered Design for the Web.* New Riders Publishers, Indianapolis, IN.

GRIFFITH, D. A., KRAMPF, R. F., AND PALMER, J. W. 2001. The role of interface in electronic commerce: Consumer involvement with print versus on-line catalogs. *Int. J. Electron. Commerce 5*, 35–47.

HALLNÄS, L. AND REDSTRÖM, J. 2002. From use to presence: On the expressions and aesthetics of everyday computational things. *ACM Trans. Comput.-Hum. Interact. 9*, 2, 106–124.

HAMMER, O. AND HARPER, D. A. T. 2005. Past, version 1.32. http://folk.uio.no/ohammer/past/.

HANDY, C. 1995. Trust and the virtual organization. *Harvard Bus. Rev.* May-June, 40.

HARTMANN, J., SUTCLIFFE, A., AND DE ANGELI, A. 2007. Investigating attractiveness in Web user interfaces. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'07).* 387–396.

HARTMANN, J., SUTCLIFFE, A., AND DE ANGELI, A. 2008. Towards a theory of user judgment of aesthetics and user interface quality. *ACM Trans. Comput.-Hum. Interact. 15*, 4, 15-1–15-30.

HASSENZAHL, M. 2004. The interplay of beauty, goodness, and usability in interactive products. *Hum.-Comput. Interact. 19*, 4, 319–349.

HASSENZAHL, M. 2008. Aesthetics in interactive products: Correlates and consequences of beauty. In *Product Experience,* H. H. J. Schiffstein and P. Hekkert Eds., Elsevier, North-Holland, the Netherlands, 287–298.

HEMENWAY, P. 2005. *Divine Proportion: Phi In Art, Nature, and Science.* Sterling Books, New York, 20–21.

HOLLINS, B. AND PUGH, S. 1990. *Successful Product Design.* Butterworth, London.

JACOBSON, T. 2004. Individual and group modeling of aesthetic judgment strategies. *Brit. J. Psychol. 95*, 1, 41–56.

JACOBSON, T. AND HÖFEL, L. 2002. Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Percept. Motor Skills 95*, 3, 755–766.

JACOBSON, T. AND HÖFEL, L. 2003. Descriptive and evaluative judgment processes: Behavioral and electrophysiological indices of processing symmetry and aesthetics. *Cogn. Affect. Behav. Neurosci. 3*, 4, 289–299.

JACOBSON, T. AND HÖFEL, L. 2007a. Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes. *Int. J. Psychophys. 65*, 20–31.

JACOBSON, T. AND HÖFEL, L. 2007b. Electrophysiological indices of processing symmetry and aesthetics: A result of judgment categorization or judgment report?. *J. Psychophys. 21*, 1, 9–21.

JORDAN, P. W. 2001. Pleasure with products: Human factors for body, mind, and soul. In *Human Factors in Product Design: Current Practice and Future Trends,* W.S. Green and P. W. Jordan Eds., Taylor & Francis, London, 206–217.

KARVONEN, K. 2000. The beauty of simplicity. In *Proceedings of the Conference on Universal Usability.* ACM Press, New York, 85–90.

KAUFMAN, A. S. 1994. Tests of intelligence. In *The Encyclopedia of Human Intelligence,* vol. 2, R. J. Sternberg Ed., Macmillan Publishing Company, 828–833.

KIM, J., LEE, J., AND CHOI, D. 2003. Designing emotionally evocative homepages: An empirical study of the quantitative relations between design factors and emotional dimensions. *Int. J. Hum.-Comput. Studies 59*, 899–940.

KOFFKA, K. 1955. *Principles of Gestalt Psychology.* Routledge and Kegan Publishing, London.

KOSSLYN, S. M. 1989. Understanding charts and graphs. *Appl. Cogn. Psychol. 3*, 185–226.

KOVACS, G., VOGELS, R., AND ORBAN, G. A. 1995. Cortical correlate of backward masking. *Proc. Nat. Acad. Sci. USA 92*, 5587–5591.

LALOMIA, M. J. AND SIDOWSKI, J. B. 1990. Measurements of computer satisfaction, literacy, and aptitudes: A review. *Int. J. Hum.-Comput. Interact. 2*, 3, 231–253.

LAUER, D. A. AND PENTAK, S. 2002. *Design Basics.* Wadsworth, Sydney, Australia. Chapter 5, Balance.

LAUER, D.A. AND PENTAK, S. 2007. *Design Basics.* Thomson Wadsworth, Sydney, Australia. Chapter 5, Balance.

LAVIE, T. AND TRACTINSKY, N. 2004. Assessing dimensions of perceived visual aesthetics of Web sites. *Int. J. Hum.-Comput. Studies 6*, 269–298.

LEDOUX, J. 1992. Emotion and the amygdale. In *The Amygdala: Neurobiological Aspects of Emotion, Mystery, and Mental Dysfunction,* J. P. Aggleton Ed., John Wiley and Sons, New York, 339–351.

LEDOUX, J. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* Simon and Schuster, New York.

LIDWELL, W., HOLDEN, K., AND BUTLER, J. 2003. *Universal Principles of Design: A Cross-Disciplinary Reference.* Rockport Publishers, Gloucester, MA.

LINDGAARD, G. AND DUDEK, C. 2002. User satisfaction, aesthetics and usability: Beyond reductionism. In *Proceedings of the IFIP 17th World Computer Congress.* 231–246.

LINDGAARD, G., FERNANDES, G., DUDEK, C., AND BROWN, J. 2006. Attention Web designers: You have 50 milliseconds to make a good first impression!, *Behav. Inf. Technol. 25*, 115–126.

LINDGAARD, G., FOLKENS, J., PYPER, C., FRIZE, M., AND WALKER, R. 2010. Contributions of psychology to the design of diagnostic decision support systems. In *Proceedings of the 2nd IFIP TC 13 Human-Computer Interaction Symposium (HCIS'10). (Held as part of WCC'10)*. 15–26.

LINDGAARD, G. AND WHITFIELD, A. 2004. Integrating aesthetics within an evolutionary and psychological Framework. *Theor. Issues Ergon. Sci. 5*, 1, 73–90.

LYNCH, P. J. AND HORTON, S. 2002. *Web Style Guide: Basic Design Principles for Creating Web Sites*, 2nd ed. Yale University Press, New Haven, CT.

MAHLKE, S. 2002. Factors influencing the experience of website usage. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'02): Changing the World, Changing Ourselves, Extended Abstracts.* ACM Press, New York, 846–847.

MAHLKE, S. 2006. Studying user experience with digital audio players. In *Proceedings of the 5th International Conference on Entertainment Computing (ICEC'06).* 358–361.

MCCARTHY, J. AND WRIGHT, P. 2005. Putting 'felt-life' at the centre of human-computer interaction (HCI). *Cogn. Technol. Work 7*, 262–271.

MCKNIGHT, D. H., CHOUDHURY, V., AND KACMAR, C. 2002. Developing and validating trust measures for e-commerce: An integrative typology. *Inf. Syst. Res. 13*, 3, 334–359.

MOUNT, M. K., OH, I.-S., AND BURNS, M. 2008. Incremental validity of perceptual speed and accuracy over general mental ability. *Personn. Psychol. 61*, 113–139.

NGO, D. C. L. AND BYRNE, J. G. 2001. Application of an aesthetic evaluation model to data entry screens. *Comput. Hum. Behav. 17*, 149–185.

NORMAN, D. A. 2004a. *Emotional Design: Why We Love (or Hate) Everyday Things.* Basic Books, New York.

NORMAN, D. A. 2004b. Introduction to this special section on beauty, goodness, and usability. *Hum.-Comput. Interact. 19*, 311–318.

PARK, S. E., CHOI, D., AND KIM, J. 2005. Visualizing e-brand personality: Exploratory studies on visual attributes and e-brand personalities in Korea. *Int. J. Hum.-Comput. Interact. 19*, 1, 7–34.

PICKFORD, R.W. 1972. *Psychology and Visual Aesthetics.* Hutchinson Educational, London.

PORTEOUS, J. D. 1996. *Environmental Aesthetics: Ideas, Politics, and Planning.* Routledge, London.

POSTREL, V. 2001. *Can good looks guarantee product success?* The New York Times. July 12.

POSTREL, V. 2002. *The Substance of Style.* Harper Collins, New York.

REDSTRÖM, J. 2008. Tangled interaction: On the expressiveness of tangible user interfaces. *ACM Trans. Comput.-Hum. Interact. 15*, 4, 16-1– 6-17.

RIEGER, J. W., BRAUN, C., BÜLTHOFF, H. H., AND GEGENFURTNER, K. R. 2005. The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study. *J. Vis. 5*, 275–286.

ROSENZWEIG, P. 2009. *The Halo Effect... and the Eight Other Business Delusions that Deceive Managers.* Simon & Schuster, New York. 50–64.

SCHENKMAN, B. N. AND JÖNSSON, F. U. 2000. Aesthetics and preferences of homepages. *Behav. Inf. Technol. 19*, 367–377.

SCHMITT, B. H. AND SIMONSON, A. 1997. *Marketing Aesthetics: The Strategic Management of Brands, Identity and Image.* Free Press, New York.

SCHULTZ, L. 2007. Effects of graphical elements on perceived usefulness of a library. http://www.tarleton.edu/schultz/finalprojectinternetsvcs.htm.

SILLENCE, E., BRIGGS, P., FISHWICK, L., AND HARRIS, P. 2004. Timeline analysis: A tool for understanding the selection and rejection of health websites. In *Proceedings of the Conference HCI'04: Design for Life*. Vol. 2. 113–116.

SILLENCE, E., BRIGGS, P., HARRIS, P., AND FISHWICK, L. 2006. A framework for understanding trust factors in Web-based health advice. *Int. J. Hum.-Comput. Studies 64*, 697–713.

SIMON, H. A. 1956. Rational choice and the structure of the environment. *Psychol. Rev. 63*, 129–138.

SONDEREGGER, A. AND SAUER, J. 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Appl. Ergon. 41*, 403–410.

SPENCER, T. J. AND SHUNTICH, R. 1970. Evidence for an interruption theory of backward masking. *J. Experiment. Psychol. 85*, 198–203.

SPERLING, G. 1960. The information available in brief visual presentations. *Psychol. Monogr. Gen. Appl. 74*, 11, 1–30.

SUN, H. 2001. Building a culturally-competent corporate Web site: An exploratory study of cultural markers in multilingual Web design. In *Proceedings of the 19th Annual International Conference on Computer Documentation.* ACM Press, New York, 95–201.

SUTCLIFFE, A. 2001. Heuristic evaluation of website attractiveness and usability. In *Interactive Systems: Design, Specification, and Verification.* Lecture Notes in Computer Science, vol. 2220. Springer, 183–198.

TABACHNICK, B. G. AND FIDELL, L. S. 2007. *Experimental Design Using ANOVA.* Duxbury Press, Belmont, CA.

THORNDIKE, E. L. 1920. A consistent error in psycho-logical ratings. *J. Appl. Psychol. 4*, 1, 25–29.

THÜRING, M. AND MAHLKE, S. 2007. Usability, aesthetics, and emotions in human-technology interaction. *Int. J. Psychol. 42*, 4, 253–264.

TRACTINSKY, N. 1997. Aesthetics and apparent usability: Empirically cultural and methodological issues. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems.* ACM Press, New York, 115–122.

TRACTINSKY, N., KATZ, A. S., AND IKAR, D. 2000. What is beautiful is usable. *Interact. Comput. 13*, 127–145.

TRACTINSKY, N., COKHAVI, A., AND KIRSCHENBAUM, M. 2004. User ratings and response latencies to evaluate the consistency of immediate aesthetic perceptions of homepages. In *Proceedings of the 3rd Annual Workshop on HCI Research in MIS.*

TRACTINSKY, N., COKHAVI, A., AND KIRSCHENBAUM, M. AND SHARFI, T. 2006. Evaluating the consistency of immediate aesthetic perceptions of homepages. *Int. J. Hum.-Comput. Studies 64*, 1071–1083.

TRACTINSKY, N. AND ZMIRI, D. 2006. Exploring attributes of skins as potential antecedents of emotion in HCI. In *Aesthetic Computing,* P. Fishwick Ed., MIT Press, Cambridge, MA, 405–421.

TVERSKY, A. AND KAHNEMAN, D. 1981. The Framing of decisions and the psychology of choice. *Sci. 211*, 453–458.

VAN DER HEIJDEN, H. 2003. Factors influencing the usage of websites: The case of a generic portal in the Netherlands. *Inf. Manag. 40*, 6, 541–549.

VAN RULLEN, R. AND THORPE, S. J. 2001. The time course of visual processing: From early perception to decision-making. *J. Cogn. Neurosci. 13*, 4, 454–461.

VERLEGER, R., JASKOWSKI, P., AYDEMIR, A., VAN DE LUBBE, R. H. J., AND GROEN, M. 2004. Qualitative differences between conscious and non-conscious processing? On inverse priming induced by masked arrows. *J. Experiment. Psychol. Gen. 133*, 494–515.

WILLIANS, R. AND TOLLETT, J. 2005. *The Non-Designer's Web Book* 3rd Ed. Peachpit Press, Berkeley, CA.

WINN, W. 1994. Contributions of perceptual and cognitive processes to the comprehension of graphics. In *Comprehension of Graphics,* W. Schnotz, and R. W. Kulhavy Eds., Elsvier Science, The Netherlands, 1–27.

WONG, W. 1987. *Principles of Color Design.* Van Nostrand Reinhold, New York.

WRIGHT, P., WALLACE, J., AND MCCARTHY, J. 2008. Aesthetics and experience-centered design. *ACM Trans. Comput.-Hum. Interact. 15*, 4, 18-1–18-21.

YAMAMOTO, M. AND LAMBERT, D. R. 1994. The impact of product aesthetics on the evaluation of industrial products. *J. Prod. Innov. Manag. 11*, 309–324.

ZAJONC, R., 1980. Feeling and thinking: Preferences need no inferences. *Amer. Psychol. 35*, 2, 151–175.

ZAJONC, R. B. 2000. Feeling and thinking: Closing the debate over the independence of affect. In *Feeling and Thinking: The Role of Affect in Social Cognition*, J. P. Forgas Ed., Cambridge University Press, Cambridge, UK, 207–220.

ZAJONC, R. B. 2001. Mere exposure: A gateway to the subliminal. Current Direct. *Psychol. Sci. 10*, 6, 224–228.

ZDRALEK, J. 2003. White space: How much nothing should there be? Unpublished MA thesis, Carleton University, Ottawa, Canada. www.deziner.com/Papers_files/thesis_content.pdf.

ZHANG, P. AND LI, N. 2004. Love at first sight or sustained effect? The role of perceived effective quality on users' cognitive reactions to information technology. In *Proceedings of the 25th International Conference on Information Systems (ICIS).* 283–296.

Received May 2009; revised September 2010; accepted September 2010