The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments

Alexandre N. Tuch, [a,*] Eva E. Presslaber, [a] Markus Stöcklin, [a] Klaus Opwis, [a] Javier A. Bargas-Avila [b]

[a] *University of Basel, Department of Psychology, Center for Cognitive Psychology and Methodology, Switzerland*

[b] *Google/YouTube User Experience Research, Zurich, Switzerland*

Abstract

This paper experimentally investigates the role of visual complexity (VC) and prototypicality (PT) as design factors of websites, shaping users' first impressions by means of two studies. In the first study, 119 screenshots of real websites varying in VC (low vs. medium vs. high) and PT (low vs. high) were rated on perceived aesthetics. Screenshot presentation time was varied as a between-subject factor (50 ms vs. 500 ms vs. 1000 ms). Results reveal that VC and PT affect participants' aesthetics ratings within the first 50 ms of exposure. In the second study presentation times were shortened to 17, 33 and 50ms. Results suggest that VC and PT affect aesthetic perception even within 17ms, though the effect of PT is less pronounced than the one of VC. With increasing presentation time the effect of PT becomes as influential as the VC effect. This supports the reasoning of the information-processing stage model of aesthetic processing (Leder et al., 2004), where VC is processed at an earlier stage than PT. Overall, websites with low VC and high PT were perceived as highly appealing.

Key words: Web design, First impression, Visual complexity, Prototypicality, User experience, Website aesthetics

* Corresponding author. Tel.: +41 61 2673568, a.tuch@unibas.ch

1 Introduction

The question how humans come up with what is commonly called "first impressions" of objects or subjects is researched in many disciplines. In emotion research for instance it is explored how humans form first impressions of people's personalities using the visual appearance of their faces (Bar et al., 2006). The authors show that consistent first impressions can be formed very quickly, based on whatever information is available within the first 39 milliseconds. First impressions do often influence mid- and longterm human behavior (e.g. Plous, 1993; Rabin and Schrag, 1999) – therefore it is important to understand how they are formed and what factors they depend on. The influence of visual appearance on human judgments has also been researched for other topics such as for instance architecture (Akalin et al., 2009), car design (Leder and Carbon, 2005), software interfaces (Saadé and Otrakji, 2007), and websites (Lindgaard et al., 2006).

A websites' first impression is known to be a crucial moment for capturing the users interest. Within a fraction of time, people build a first visceral "gut feeling" that helps them to decide whether they are going to stay at this place or continue surfing to other sites. Research in this area has been mainly stimulated by a study of Lindgaard et al. (2006), where the authors were able to show that people are able to form stable attractiveness judgments of website screenshots within 50 milliseconds. Lindgaard et al.'s work led to a growing body of research that works toward understanding the nature of how a website's first impression is built, what factors it depends on and what consequences arise if a website triggers positive or negative judgments. Insights into these topics may enable webdesigners one day to intentionally shape positive first impressions.

In this work, we explore how the factors *visual complexity* and *prototypicality* influence first impressions of websites. We do this by applying and modifying the research paradigm of Lindgaard et al. (2006).

2 Theoretical Background

2.1 First impression

A stream of research within the field of visual aesthetics explores how initial judgments (first impressions) are built and what they depend on. One of the first publications laying heavy emphasis on websites' first impression was published by Lindgaard et al. (2006). In their first study, participants had to rate

a set of websites twice, after perceiving them for 500 ms in each trial. The authors found a high correlation of $r = .97$, leading them to conclude that people are able to form a reliable first impression within half a second. In their second study, participants performed an additional trial, where they rated the same websites after perceiving them for as long as they wished. The correlation between the ratings provided after 500 ms and unlimited perception time was again very high ($r = .98$). In their last experiment, participants had to rate a set of websites after viewing them for only 50 ms and in the second trial after 500 ms. The authors once again found a very high correlation of $r = .97$. They concluded that people are able to form a stable first impression within 50 ms of perception and that web designers should aim at providing a positive impression within this timeframe.

The studies of Lindgaard et al. (2006) led to several follow-up studies – most of them applying the same or a modified version of Lindgaard et al.'s research paradigm to various types of website screenshots and measuring different dependent variables. A common goal of these studies is to understand better what factors lead to what type of user judgments. Table 1 summarizes these follow-up studies. Overall these publications show, that users can form – in a very short time – reliable judgements of for instance attractiveness or trust, and that these judgements depend on factors such as context or visual complexity.

2.2 The process of aesthetic perception

The visual appeal, attractiveness, beauty or aesthetics (all terms are used synonymously in this paper) of user interfaces has become a topic of major interest in HCI (Bargas-Avila and Hornbæk, 2011). Numerous studies show the influence of aesthetics on for instance the relationship to usability (de Angeli et al., 2006; Hassenzahl and Monk, 2010; Tuch et al., 2012), trust and credibility (Karvonen et al., 2000; Robins and Holmes, 2008), and overall impression (Schenkman and Jönsson, 2000; Tuch et al., 2010).

Besides the general importance of aesthetics in HCI, researchers should especially be interested in how such aesthetic impressions are formed by users. It is evident that the process of human aesthetic perception is very complex, because an aesthetic judgment is shaped by several physical features of the perceived stimuli (e.g., shape, color, complexity) as well as by the individual characteristics of the perceiver like previous experiences or knowledge (Reber et al., 2004).

An elaborated approach to the process of aesthetic perception of art objects is provided by Leder et al. (2004). With the information-processing stage model

Table 1: Summary of publications in the field of first impressions

Publication	Main research topic	Screenshots; participants	Exposure time	Mask[1]	Dependent measures	Main results
Lindgaard et al. (2006), study 1	Attractiveness	100; 22	first and second trial: 500 ms	no	Visual appeal[a]	Appeal judgments after 500 ms are highly reliable.
Lindgaard et al. (2006), study 2	Attractiveness	50; 31	first and second trial: 500 ms, third: unlimited	no	Visual appeal[a]; design characteristics (third trial)[b]	Appeal judgments after 500 ms correlate highly with judgments without time restrictions.
Lindgaard et al. (2006), study 3	Attractiveness	50; 40	group #1: 50 ms, group #2: 500 ms	no	Visual appeal[a]	Reliable appeal judgments are already formed after 50 ms.
Tractinsky et al. (2006), study 1	Attractiveness	50; 40	first trial: 500 ms, second trial: 10s	no	Visual appeal[c]; objective response latency	Attractiveness ratings after 500 ms correlate highly with ratings after 10s. Extreme attractiveness ratings were provided faster by participants than moderate ones.
Tractinsky et al. (2006), study 2	Attractiveness	24; 53	first trial: 500 ms, second trial: unlimited	no	classical & expressive aesthetics[d]	High correlation between attractiveness and classical/expressive aesthetics (Lavie and Tractinsky, 2004). Low attractiveness is associated mainly with very low ratings of expressive aesthetics.
Kim and Fesenmaier (2008)	First impression[2]	50; 65	7 seconds	no	First impression[2]; informativeness; usability; credibility; inspiration; involvement; reciprocity[e]	Inspiration and usability are factors that lead to favorable first impression.
Lindgaard et al. (2008), study 1	Cultural effects reg. attractiveness	50; 72	group #1: 50 ms, group #2: 500 ms; 2 trials	no	Visual appeal[c]	There were no cultural differences regarding attractiveness ratings of US websites
Lindgaard et al. (2008), study 2	Cultural effects reg. attractiveness	50; 80	young: 50 ms, old: 500 ms; 2 trials	no	Visual appeal[c]	Chinese/Taiwanese participants rated visual appeal higher than Canadians when judging web pages of their native culture, but no differences emerged for North American web pages.
Michailidou et al. (2008)	Visual complexity and aesthetics	30; 55	first and second trial: 7s	no	Visual complexity[a]; classical & expressive aesthetics[f]	Strong correlation between visual complexity and structural elements (links, images, words and sections) as well as aesthetics (organization, clearness, cleanliness, interestingness and beautifulness).
Robins and Holmes (2008)	Attractiveness and credibility	42[3]; 20	unlimited	no	Credibility[a]	When the same content is presented using different levels of aesthetic treatment, the content with a higher aesthetic treatment was judged as having higher credibility.
Albert et al. (2009)	Trust	50; 64	first and second trial: 50 ms	yes	Trust[a]	Approximately 50% of participants were consistent in their trust assessments for the same web sites across both trials. There was a significant correlation between trust assessments of both trials, averaged across all participants.
van Schaik and Ling (2009), study 1	Attractiveness and context	50; 125	first trial: 500 ms, second trial: unlimited	no	Visual appeal (first trial)[c]; classical & expressive aesthetics (second trial)[d]	Context increases stability of judgments from perceptions after brief exposure to those after self-paced exposure. More attractive pages are preferred over less attractive ones after brief exposure, but only if no context is provided.
van Schaik and Ling (2009), study 2	Attractiveness and context	2[4]; 115	first trial: 500 ms, second trial: unlimited, third trial: site usage	no	Visual appeal (first trial)[c]; classical & expressive aesthetics (second trial)[d]; mental workload[g]; task performance	Context increases the stability of judgments from perceptions after self-paced exposure to those of after site use. After brief exposure, classically aesthetic pages that are information oriented are rated as more attractive than expressively aesthetic pages.
Thielsch and Hirschfeld (2010)	Attractiveness and usability	50; 91	unlimited	no	Perceived usability[a]; perceived aesthetics[a]	No connection between low spatial frequencies (convey information about global layout) and usability evaluations, but strong correlations between ratings of high-pass filtered websites (convey information on fine details) and those of unfiltered websites in aesthetics and usability.
Lindgaard et al. (2011), study 1	Attractiveness	50 ; 20	first and second trial: 50 ms	yes	Visual appeal[a]	Shows that results of prior studies in this field can be replicated, even if masking[1] is used.
Lindgaard et al. (2011), study 2	Attractiveness, trust, usability	50 ; 48	first and second trial: 50 ms	no	Visual appeal[a], perceived trustworthiness[a], and perceived usability[a]	Judgments of appeal, trust and usability were highly consistent from one trial to the next in aggregate and comparisons of individual data. All three judgments were driven predominantly by appeal.
Thielsch and Hirschfeld (2012)	Attractiveness	50 ; 92	50, 500 and 10000 ms	yes	Visual appeal[a]	High correlations between aesthetic responses to low-pass filtered (LF), high-pass filtered and unfiltered websites. Moderate effect of LF when stimuli are presented only once and very briefly for 50 ms.

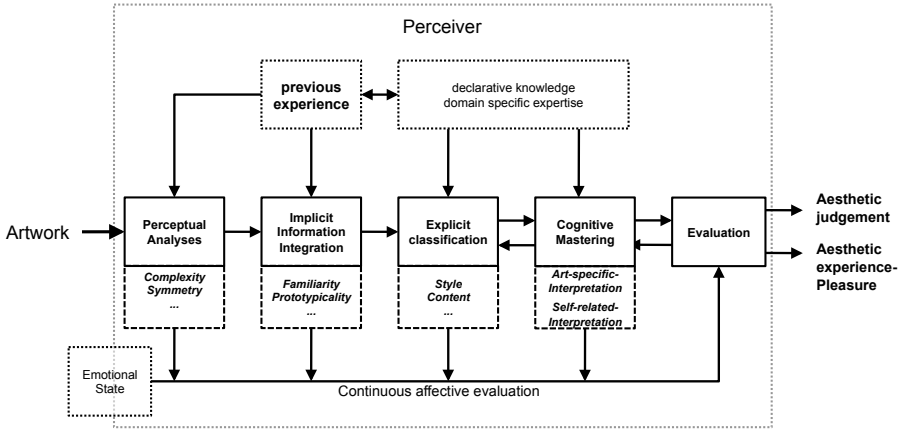
Notes. [a] 1 ad-hoc item; [b] 7 ad-hoc items; [c] 1 item from Lindgaard et al. (2006); [d] short version (6 items) based on Lavie and Tractinsky (2004); [e] items based on several sources or ad-hoc (see Kim and Fesenmaier (2008) for details); [f] 5 arbitrarily selected items from Lavie and Tractinsky (2004); [g] 1 item (Zijlstra, 1993)

[1] see section 4.1.4 for information regarding masking [2] The authors measured first impression using a 5-point Likert-type scale (terrible vs. awesome) by Crites et al. (1994)

[3] In this study 21 pairs of manipulated websites were used (high- and low-aesthetics versions of the same site); [4] In this study the stimulus material consisted of two functional websites

of aesthetic processing, Leder et al. developed a theoretical framework for research dealing with the perception of aesthetic stimuli. As illustrated in Figure 1, the model proposes five different processing stages targeted towards different cognitive analyses. The first two stages (1) perceptual analyses and (2) implicit memory integration, can be seen as intuitive and basic cognitive processes, which do not have to become conscious in order to affect perceiver's aesthetic processing ("gut feeling"). In contrast, stages three and four are higher cognitive processes that are affected by the expertise and knowledge of the perceiver, something especially relevant in the context of art perception and therefore not relevant for the topic of this paper. At the fifth stage (evaluation) the processed information are evaluated. The model suggests two different outputs of evaluation: aesthetic appraisal and aesthetic judgment. Aesthetic appraisals originate from positively valued immediate subjective experiences, whereas aesthetic judgments entail references to normative criteria and refer to the question of what should be considered being aesthetic in a more artistic sense (Leder et al., 2004; Moshagen and Thielsch, 2010). These two outcomes are not necessarily related. For instance, perceivers can judge an object as having an important artistic value (aesthetic judgment), but they are not positively affected by it (aesthetic appraisal). Further, the model assumes a relative hierarchy of the five processing stages.

Fig. 1. Information-processing stage model of aesthetic processing (adapted from Leder et al., 2004)



Even though the model does not depict a strict serial flow of information, it allows formulating hypotheses concerning time sensitive processing of a stimulus (Leder et al., 2006). Hence the first two stages of the model may play an important role in regard to first impressions (especially impressions formed in a very short period of time). With regard to the processing of a stimulus, at the first stage the perceived stimulus is analyzed perceptually, using physical features such as visual complexity or symmetry. After this perceptual analysis, at the stage of implicit memory integration, stimulus characteristics involving the perceiver's previous experiences shape the process of aesthetic perception. Variables such as prototypicality or familiarity are important at this stage.

Although the model was developed for art perception, certain aspects, such as aesthetic variables from the first two processing stages, should also be applicable for HCI research (e.g., in the context of research on first impression of websites).

Please note that this study follows an interactionist perspective on visual aesthetics (i.e. historical perspectives of object aesthetics or perceiver aesthetics such as discussed by e.g. Eysenck (1941) are disregarded).

2.3 Visual complexity

As observed by Xing and Manning (2005), it is quite difficult to define VC and there are many attempts in existing literature. However, the present study is not interested in giving a specific definition for VC. We rather follow the reasoning of Edmonds (1995), who suggests that complexity makes only sense when considered relative to a given observer. Consequently, we are primarily interested in the subjectively perceived complexity and not in a objective definition of VC (for a review on definitions of complexity see Xing and Manning, 2005).

As introduced in the context of the Leder et al. (2004) model, visual complexity (VC) plays a crucial role in the perception of visual stimuli. There are theories and several studies that have found a relationship between aesthetic preferences and complexity. According to Berlyne's aesthetic theory (Berlyne, 1974), viewers' pleasure is related to the arousal potential of a stimulus. This relationship is represented in an inverted U-curve. Berlyne showed that stimuli with a moderate arousal are pleasureable, whereas stimuli with high arousal potential are experienced as unpleasant, and stimuli with low arousal are experienced as boring. The arousal potential is linked to the potency of such collative variables as complexity, novelty, and ambiguity – these therefore become the most important predictors for perceived aesthetic preference. Berlyne's theory predicts that stimuli of a moderate degree of visual complexity will be considered pleasant, whereas both less and more complex stimuli will be considered unpleasant. However, the empirical support for this inverted U-shaped relation is mixed, several studies found a linear rather than a quadratic relation (for a critical examination see Martindale et al., 1990).

There are also several empirical studies from the filed of HCI that provide evidence for the influence of VC on aesthetic perception. Web pages of moderate complexity facilitate communication effectiveness and lead to more favorable consumer responses (Geissler et al., 2006). Furthermore, web page complexity seems to have a significant influence on users' attitudes and intentions (Bruner and Kumar, 2000; Stevenson et al., 2000). A negative correlation be-

tween complexity and pleasure in website perception was found by Pandir and Knight (2006). Further, in a study of Tuch et al. (2009), VC of web pages was related to increased experienced arousal, more negative valence appraisal, and increased facial muscle tension (musculus corrugator). Even though the aforementioned studies do not all present exactly the same pattern between complexity and viewers' ratings or responses to web pages, they indicate that, as proposed by Berlyne, complexity of objects has a major impact on viewers' perceptional state and behavior.

There is currently only one study known to us that explores the role of complexity regarding first impression and aesthetic judgments. Michailidou et al. (2008) presented 30 screenshots to users in two trials of seven seconds each. Their results show that there is a strong correlation between VC and aesthetics (see Table 1).

2.4 Prototypicality

Prototypicality (PT) is defined as "the amount to which an object is representative of a class of objects" (Leder et al., 2004, p. 496). Prototypicality is represented by mental models built through experience; a prototypical object usually represents a class of objects. Many studies found empirical evidence that people show prototypicality preferences, for instance for facial attractiveness (Etcoff, 1999), colors (Martindale, 1984), or paintings (Hekkert and Wieringen, 1990). An overview of PT literature can be found in Whitfield (2000).

In the course of time, through interaction with the Internet, users develop certain expectations of how websites look. Roth et al. (2010) showed that distinct mental models seem to exist for different web page types, i.e. people agree on many but not all web objects' location.

Hekkert et al. (2003) explored how prototypicality and novelty influence the aesthetic preference of products. They found that people prefered novel designs only as long as the novelty did not affect prototypicality. Sen and Lindgaard (2008) showed that aesthetic appeal is positively correlated to prototypicality for a set of stimuli consisting of images representing basic object categories such as guitars, chairs or cars.

There is little research that shows how prototypicality influences aesthetic judgments of websites, but it can be assumed that there may be also a preference for prototypical websites. Figure 2 shows some examples of websites of low and high prototypicality.

2.5 Response latency of aestethic ratings

In a study that explores the correlation between visual judgments given after different exposure times, Tractinsky et al. (2006) postulated a relationship between response latency and the extremity of the ratings. Their data support the conclusion that latencies of very attractive or very unattractive web pages are shorter than latencies of ratings that were placed at the middle of the scale. Response latencies and beauty ratings are related in the form of an inverted U-shaped curve. Or simply said: Users need less time to judge very ugly or beautiful websites and are slower when judging stimuli that are in between.

We emphasize these findings, because we will present evidence that these results are based on a problematic statistical procedure.

2.6 Hypotheses

Based on the presented research in this field, we decided to use Lindgaard et al.'s paradigm in combination with different degrees of VC and PT to explore how and when these factors start to influence first impression.

According to Leder's model of aesthetic perception (Leder et al., 2004), Berlyne's aesthetic theory (Berlyne, 1974), and previous empirical studies on aesthetic website perception (Tuch et al., 2009, 2011), we hypothesize that VC as well as PT have a significant impact on aesthetic judgments of the participants. More complex websites are perceived as being less beautiful than less complex websites, whereas websites of high PT are perceived as being more beautiful than websites of low PT. Furthermore, by relying on the model of Leder et al. (2004), we would expect that with regard to presentation time the magnitude of the VC effect is stable across the different time conditions, whereas the magnitude of the PT effect increases with longer presentation times. This would reflect the idea of different processing stages for the variables. Consequently, we expect a Presentation Time x PT interaction, but no Presentation Time x VC interaction. However, the model does not provide any time estimates on its processing stages, hence we are not able to predict exactly within which timeframe VC or PT affects aesthetic perception.

In the following we describe two studies aiming at investigating the effect of VC and PT on perceived beauty of web pages at different presentation times.

3 Study 1

3.1 Method

3.1.1 Design

This experiment used a three-way mixed design with the within-subject independent variables *visual complexity* (low vs. medium vs. high) and *prototypicality* (low vs. high). The between-subject independent variable was *presentation time* with three levels (50 vs. 500 vs. 1000 ms). The dependent variable was perceived beauty.

3.1.2 Participants

In total, n = 59 participants (45 females), mainly undergraduate psychology students of the University of Basel, took part in the experiment. They had a mean age of 25.4 years (SD = 10.3), whereas the age ranged between 18 and 62 years. Participants' mean experience in using the web was 9.5 years (SD = 2.6). All participants are used to working with computers and use the web regularly. They had no education in either visual design or web design. As compensation for taking part in the experiment, they received course credits or an equivalent of 10 US$. Participants were randomly assigned to one of the three experimental conditions (presentation times).

3.1.3 Apparatus and materials

The experiment was implemented with the software E-prime 2 (Psychology-Software-Tools, 2002) and conducted on desktop computers using 17" TFT screens with a resolution of 1440 x 900 pixels and a refresh rate of 60 Hz. The experiment was run at a resolution of 1000 x 800 pixels.

3.1.4 Stimuli selection and validation

Because websites are very heterogeneous stimuli, we decided to focus our investigations on a specific category of websites. For their study on mental models, Roth et al. (2010) extracted six different categories of websites from the 100 most visited websites of the USA, Germany, Austria and Switzerland: (1) company pages, (2) social networking sites, (3) online newspapers and news portals, (4) online shops, (5) search engines, and (6) various types. The category *company websites* was identified as being the largest one, and Roth et. al. were able to show that users have a consistent mental model of com-

pany websites. Therefore, we decided to only include company websites in our sample.

In our study we used screenshots of existing company web pages (always the main page) to ensure high ecological validity of the experiment. All screenshots were taken at the same size (1280 x 1024 pixels; for the main study the size of the screenshots was reduced to 1000 x 800 pixels). In order to define an appropriate set of stimuli for our main experiment, we used the following stimulus selection procedure: In a first step, we set up a pool of 464 company websites. These websites were listed as the most visited within the following business categories (alexa.com; August 2010): chemicals, energy, accounting, aerospace and defense, automotive, biotechnology and pharmaceuticals, financial services. In a next step, we reduced the initial pool to 270 websites by excluding sites having (1) a contentless intro page, (2) a shopping basket, (3) advertisement (banners), (4) an archive or (5) a content language other than English or German. These exclusion criteria are based on the findings of Roth et al. (2010), who showed which website-elements are typical for company sites. The remaining 270 websites were then rated by 267 participants in an online survey on visual complexity (VC; "I think this website is of high visual complexity") and prototypically (PT; "This website looks like a typical company website"). To reduce the workload for participants, the 270 sites were split into nine sets, containing 30 websites each. Participants were randomly assigned to one of the nine sets, so that each participant only had to rate 30 website screenshots. A single set was at minimum rated by 14 participants. In order to estimate the inter-rater agreement of the VC and PT appraisal, we calculated the interclass correlation (ICC) for each set separately. The average ICC coefficient across all nine sets was remarkably high for both, VC (ICC = .82; SD = .09) and PT (ICC = .78; SD = .06), suggesting a high agreement between the participants. Subsequently, the ratings were aggregated for each website so that finally every website had a score for VC and PT. By means of those scores, we allocated 120 of the 270 websites to the 6 experimental conditions (VC x PT) so that each condition contained 20 websites. To illustrate the successful allocation of the websites to the conditions, we ran an 2 x 3 MANOVA with VC and PT as independent variables and VC and PT ratings as dependent variables. As shown in Table 2, the modeling of the experimental factors was successful: VC and PT were independently manipulated. Only the expected main effects turned out to be significant (with very large effect-sizes) and no interaction between the independent variables occurred. Figure 2 shows screenshot examples for all six categories.

In order to ensure that participants really experienced a first impression of previously unknown sites, the authors aimed to selected websites that are unknown in Switzerland. To ensure this, we ran an additional online study (n = 86) evaluating all 120 web pages on *familiarity* ("How often have you visited that website?") and *attitude towards the company / brand* ("Do you know

this company / brand?") by means of a 5 point scale ranging from 1 ("not at all") to 5 ("very often / good"). To reduce the workload of the participants, the web pages were split into four sets, containing 30 pages each. A single set was at minimum rated by 18 participants. Results showed that the selected website were very unfamiliar to the participants ($M = 1.04$, $SD = .04$). Also the name of the companies or brands were not familiar to the participants ($M = 1.3$; $SD = .67$). Furthermore, the ratings did not differ between the experimental conditions. However, there were a few single websites, which achieved rating scores greater than 2.0 on the brand familiarity scale. Consequently, we excluded the familiar stimuli and reran all analyses described in section 3.3 and 4.2, but the outcome of the analyses remained the same as when all stimuli were included. We therefore rule out that the outcome of the main study is confounded by familiarity of the website or companies / brands. A list of all websites used in the study can be found in the Appendix.

Table 2
Manipulation check (results from the stimuli selection study)

	M (SD)	*M (SD)*	*M (SD)*	*F(1-2,114)*	η_p^2
Visual complexity	high	medium	low		
Visual complexity	3.32 (.36)	4.19 (.21)	5.02 (.31)	289.28	.835*
Prototypicality	3.81 (.57)	3.75 (.30)	3.78 (.78)	1.01	.017
Prototypicality	high	-	low		
Visual complexity	4.18 (.29)	-	4.08 (.29)	2.42	.021
Prototypicality	3.78 (.55)	-	4.99 (.30)	215.46	.654*
Visual complexity x prototypicality					
Visual complexity	-	-	-	.28	.005
Prototypicality	-	-	-	.36	.006

Note. * $p < .05$

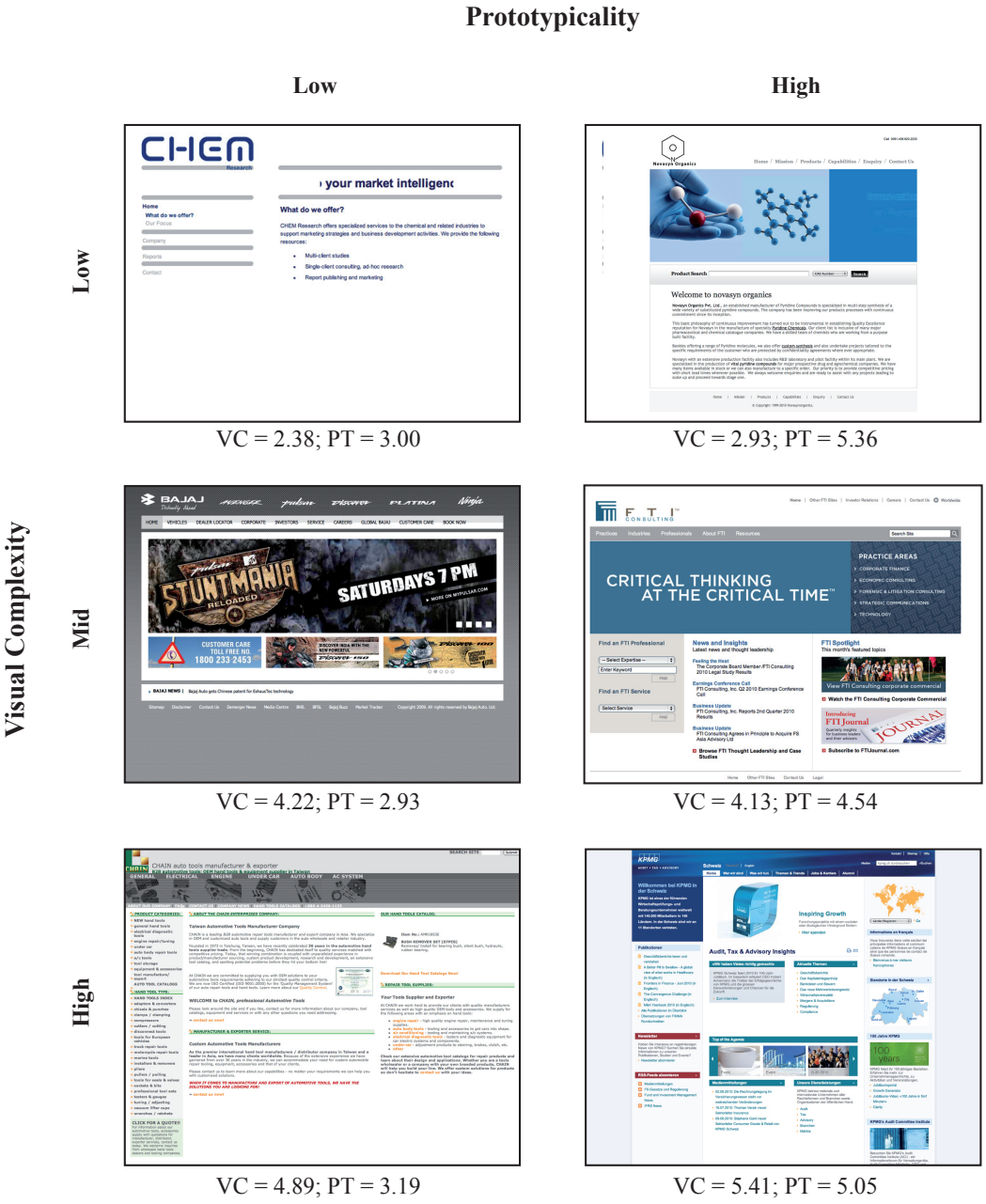
3.2 Ratings

Perceived beauty was assessed by means of a visual analog scale (VAS) with the anchors ugly and beautiful. All ratings were given using a computer mouse to click on the corresponding area of the scale. For the statistical analysis, responses on the VAS were converted to a scale ranging from 1 to 9.

3.2.1 Procedure

The experiment took place in a computer laboratory of the University, a room with 20 independent computer workstations. The participants were seated in front of a computer monitor and controlled the progress of the experiment using a computer mouse. All instructions were displayed on the screen; participants were tested in groups of 5 to 15 people, but each participant

Fig. 2. Screenshot examples for all six experimental categories.

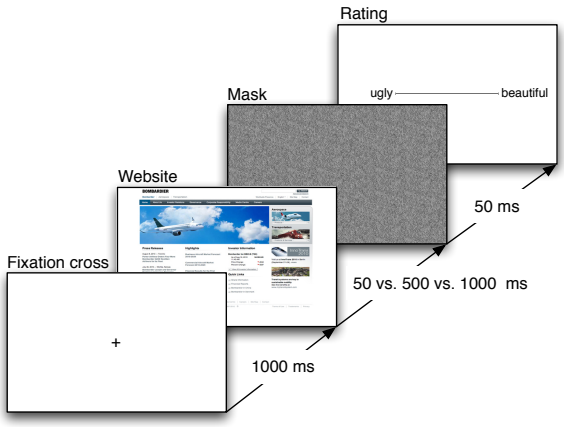


worked on a separate computer and could therefore proceed with the experiment independently from each other (workplaces were separated by wooden walls forming experimental cubicles). They were randomly assigned to one of the three experimental between-subject conditions.

Participants were presented with the 120 website screenshots one after another in random order, and they had to rate each of them immediately on perceived beauty. They were asked to rate the web pages intuitively without thinking too much. Presentation time of the website varied depending on the experi-

mental condition a participant was assigned to. To ensure that participants perceived the websites for exactly the desired duration, we used a masking procedure (c.f., Albert et al., 2009; Fei-Fei et al., 2007) to prevent participants from an extended perception of the presented stimulus, because the perception of any stimulus persist for about 250 ms after the stimulus is extinguished (Goldstein, 2009). Therefore, it is important to ensure accurate presentation times by appropriate masking. Figure 3 depicts the masked stimulus presentation trial. For the masking we used a random visual noise pattern (Rolke and Hofmann, 2007; Cantor and Thomas, 1976). The mask consisted of a random arrangement of black and withe pixels and had the same size than the web pages (1000 x 800 pixels).

Fig. 3. Procedure for a single trial used in this experiment.



3.3 Results

All data undergoing ANOVA were tested for the assumption of sphericity. Greenhouse-Geisser adjustment was used to correct for violations of sphericity. An alpha level of .05 was used for all statistical tests.

There was an unfortunate problem with the data logging of one stimulus. Due to a programming error, no rating data nor latencies were recorded for this screenshot. Hence, 119 stimuli remain for data analysis.

3.3.1 Effects of visual complexity and prototypicality on perceived beauty

To investigate the effect of VC and PT on perceived beauty for different presentation times we ran a 3 x 2 x 3 ANOVA for mixed samples, whereas VC and PT were entered as dependent sample factors and presentation time as an independent sample factor. Perceived beauty was the dependent variable. The three-way interaction was not significant ($F(3.4,\ 96.1) = 1.545$, $p = .202$, η_p^2

= .052), so all two-way interactions can be interpreted without restrictions. In regard to the two-way interactions, neither the interactions between VC and presentation time ($F(3.6, 99.9) = 2.064$, $p = .099$, $\eta_p^2 = .069$) nor between PT and presentation time reached a significant level ($F(2.0, 56.0) = 1.894$, $p = .160$, $\eta_p^2 = .063$). However, there was a significant VC x PT interaction ($F(1.7, 96.1) = 85.273$, $p < .001$, $\eta_p^2 = .604$) suggesting caution in interpreting related main effects.

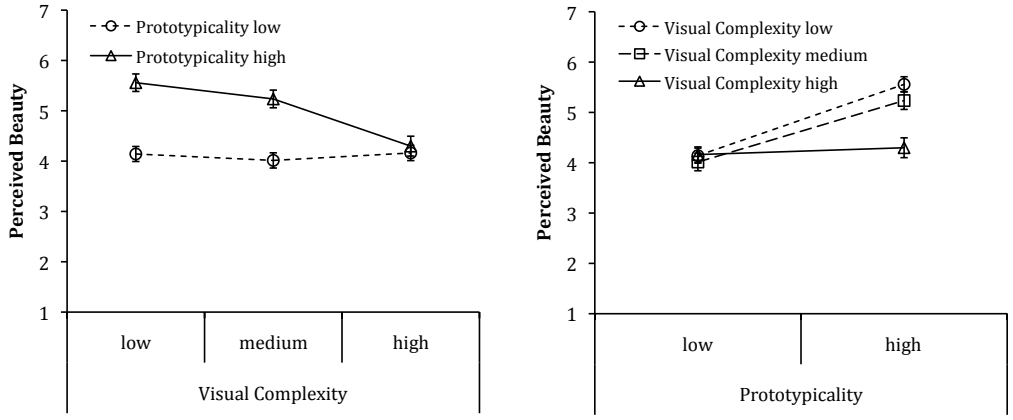
As expected, there were significant main effects for VC ($F(1.8, 99.9) = 77.607$, $p < .001$, $\eta_p^2 = .581$) and PT ($F(1.0, 56.0) = 241.365$, $p < .001$, $\eta_p^2 = .812$). More complex web pages received lower beauty ratings than less complex pages (Bonferroni corrected post-hoc test revealed that all three levels differ significantly from each other) and pages of high PT were rated as more beautiful than pages of low PT. However, these main effects must be interpreted in regard to the significant VC x PT interaction, which suggest that VC affects perceived beauty more strongly within the high PT condition, respectively the effect of PT on perceived beauty is blunted if VC is high (see Figure 4). The calculation of simple main effects confirms this interpretation: web pages of low and high PT differ significantly on each VC level, but when examining the effect sizes, the effect is much more pronounced for the low and medium (cohen's $d = 1.96$, respectively 1.79) than for the high complexity level ($d = .24$). Furthermore, all three levels of VC differ significantly from each other in the high PT condition, but not in the low PT condition. There was only a significant difference between medium and high VC, suggesting – against our expectations – that web pages of high VC are preferred over web pages of medium VC. However, effect sizes reveal that the effect in the low PT condition is relatively small ($d = .34$) compared to the effects in the high PT condition ($d = .68$ - 1.58). Finally, there was no main effect for presentation time ($F(2, 56) = 3.849$, $p = .222$, $\eta_p^2 = .052$). Descriptive statistics for all experimental conditions can be seen in Table 3.

We found support for VC and PT being relevant factors for aesthetic perception in the context of first impressions. However, the VC x PT interaction suggest that VC and PT are somehow related regarding their influence on perceived beauty. Regarding the non-significant Presentation Time x PT interaction, there is no support for the assumption raised from the model of Leder et al. (2004) that VC and PT are processed at different stages. However, it might be that the exposure time of 50 ms is already sufficient for stimulus information to be processed within the stage of implicit memory integration. Hence, a further study with shorter exposure times was conducted and is presented in the following section.

Table 3
Statistical parameters for perceived beauty for all experimental conditions

	Visual Complexity		
	low	medium	high
Prototypicality	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
		50 ms	
low	4.34 (.60)	4.37 (.66)	4.38 (.65)
high	5.52 (.79)	5.37 (.96)	4.59 (.83)
		500 ms	
low	4.18 (.79)	3.92 (.82)	4.16 (.79)
high	5.57 (.69)	5.21 (.71)	4.16 (.85)
		1000 ms	
low	3.91 (.62)	3.76 (.55)	3.95 (.60)
high	5.58 (.86)	5.13 (.70)	4.14 (.97)

Fig. 4. Interaction between visual complexity and prototypicality. Error bars represent standard errors.



4 Study 2

From study 1 we could conclude that besides VC also PT influences the beauty appraisals within 50 ms. However, in regard to the information-processing stage model of aesthetic processing (Leder et al., 2004) we would assume that the effect of VC on aesthetic perception would occur prior to the effect of PT, because the processing of VC is supposed to happen at an earlier stage (stage of perceptual analysis) than the processing of PT (stage of implicit memory integration). This suggests that the VC effect should be more pronounced with lower exposure times than the PT effect. But results from study 1 do not support this line of reasoning. One possible explanation might be that an exposure time of 50 ms is already sufficient for stimulus information to reach the stage of implicit memory integration so that the information can also be processed in regard to PT. Consequently, it would be interesting to lower the exposure time below 50 ms to see if there is a point where the influence of VC

exceeds the influence of PT on aesthetic judgment. Hence, we reran the first study by using shorter presentation times of 33 ms, respectively 17 ms.

4.1 Method

4.1.1 Design

The same design as in study 1 was used, except that this time factor VC was only varied with two levels (low vs. high). This is due to the fact that in study 1 VC and beauty were related in a linear manner, hence there was no further need for a medium VC condition. The factors VC and PT were again implemented as within-subject independent variables, whereas *presentation time* was the between-subject independent variable with three levels (17 vs. 33 vs. 50 ms) [1] . The dependent variable was perceived beauty.

4.1.2 Participants

A sample of 82 participants (57 females) took part in the experiment. They had a mean age of 27.3 years ($SD = 9.9$), whereas the age ranged between 16 and 63 years. Participants' mean experience in using the web was 9.8 years ($SD = 2.8$). All participants were experienced computer users and used the web regularly. They had no education in either visual design or web design. As compensation for taking part in the experiment, they received course credits or an equivalent of 10 US$. Participants were randomly assigned to one of the three experimental conditions (presentation times).

4.1.3 Apparatus and materials

The same experimental setup and stimuli as in study 1 were used.

4.1.4 Procedure

The experimental procedure was exactly the same as in study 1 with the exception that this time only 80 web pages were presented to the participants. This was due the elimination of the medium VC condition.

[1] These levels of presentation time are due to refresh rate of the TFT screen. Moreover, the presentation times could only be achieved approximately. The mean exposure time was 16.67 ms ($SD = .47$) for the 17 ms condition, 33.27 ms ($SD = .44$) for the 33 ms condition and 49.98 ms ($SD = .44$) for the 50 ms condition.

4.2 Results

The data were analyzed with a three-way ANOVA for mixed samples with VC and PT as within-subject independent variables and presentation time as between-subject independent variable. Perceived beauty rating was entered as the dependent variable.

All results of the ANOVA are displayed in Table 4. The three-way interaction was not significant, which allows to interpret all two-way without restrictions. There was no interaction between VC and presentation time, but a significant PT x Presentation Time interaction (see Figure 5). Moreover, the VC x PT interaction turned out significant. There were significant main effects for VC and PT, meaning that more complex web pages received lower beauty ratings than less complex pages and that pages of high PT were rated as more beautiful than pages of low PT. However, these main effects must be interpreted in regard to the significant two-way interactions. For the presentation time related interactions, it seems that VC strongly affects perceived beauty already at the 17 ms level, whereas the effect of PT on beauty becomes only stronger with increasing presentation time. Analyzing the different presentation time conditions separately gives further support for this interpretation: There are significant main effects for VC and PT at all three levels, but at 17 ms the magnitude of the effect is more pronounced for VC than for PT (η_p^2 = .38, respectively η_p^2 = .19), whereas at 33 ms and 50 ms the effects have approximately the same magnitude (VC: η_p^2 = .65 and PT: η_p^2 = .66, respectively VC: η_p^2 = .57 and PT: η_p^2 = .66). In regard to the VC x PT interaction, VC has a stronger effect on beauty with pages of high PT and the influence of PT on beauty is more pronounced when VC is low. This suggests again that the combination of low VC and high PT within a web page results in the highest beauty appraisals. The calculation of simple main effects supports this interpretation: comparisons between low and high PT turned out significant in both VC conditions and comparisons between low and high VC turned out significant in both PT conditions, but effects of PT were more pronounced in the low (cohen's d = .91) than in high VC condition (d = .33) and effects of VC were stronger in the high (d = .91) than the low PT condition (d = .30). Finally, there was no main effect for presentation time. Descriptive statistics for all experimental conditions and corresponding interaction plots can be seen in Table 5 and Figure 5, respectively.

The results of Study 2 confirm the influence of VC and PT on perceived beauty, suggesting that the combination of low VC and high PT within a web page leads to the most positive aesthetic appraisals. Moreover, study 2 revealed that VC and PT can be perceived even below 50 ms and within that time be used to form aesthetic appraisals. At 17 ms we can observe effects for both variables. Interestingly, at this level the effect of PT is less pronounced

that the one of VC. This supports the reasoning of Leder et al. (2004) that VC is processed at an earlier stage than PT.

Table 4
ANOVA for perceived beauty

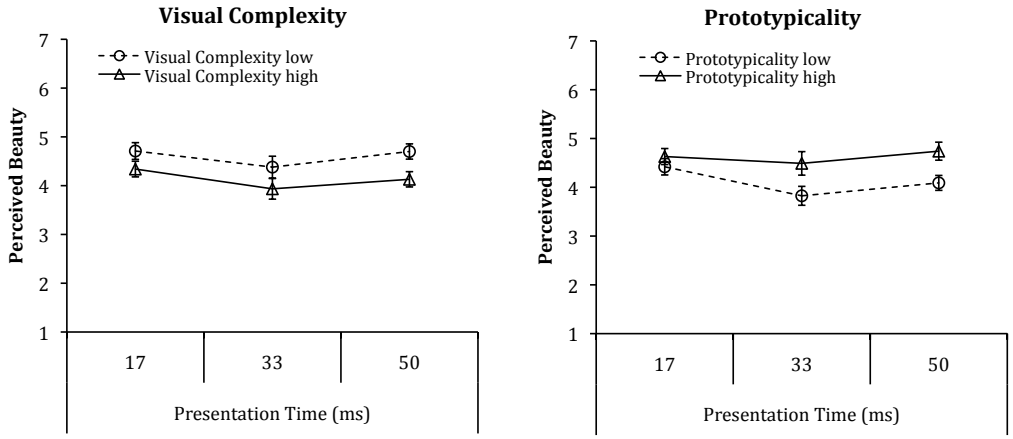
	df	*F*	η_p^2
Within-Subject			
VC	1	85.60	.52*
PT	1	97.15	.55*
VC x PT	1	30.85	.28*
VC x time	2	1.46	.04
PT x time	2	8.83	.18*
VC x PT x time	2	2.45	.06
error	79		
Between-Subject			
time	2	1.32	.03

Note. VC = visual complexity; PT = prototypicality; * p < .05

Table 5
Mean and standard deviation of perceived beauty for all experimental conditions

	Visual Complexity					
	17 ms (n = 30)		33 ms (n = 24)		50 ms (n = 28)	
	low	high	low	high	low	high
Prototypicality	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
low	4.53 (0.94)	4.31 (0.9)	3.84 (0.93)	3.81 (0.97)	4.17 (0.78)	4.01 (0.72)
high	4.89 (0.94)	4.37 (0.86)	4.92 (1.26)	4.06 (1.10)	5.23 (0.87)	4.25 (0.94)

Fig. 5. Effects of visual complexity and prototypicality on perceived beauty by different presentation times. Error bars represent standard errors.



5 Relationship between response latency and rating extremity

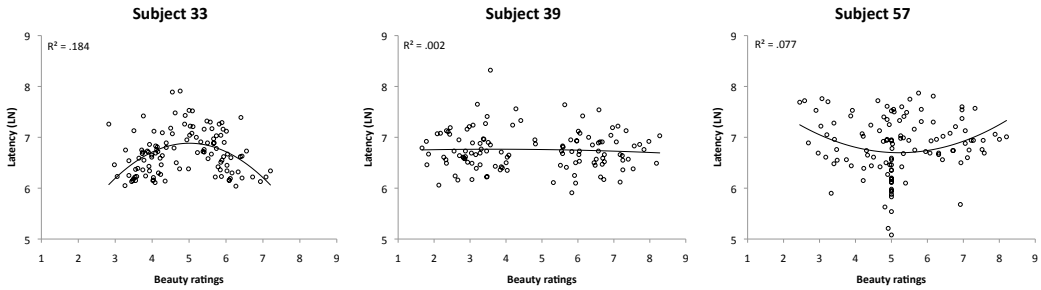
As mentioned in section 2.5, Tractinsky et al. (2006) postulated a relationship between response latency and the extremity of the rating: latencies of very attractive or very unattractive web pages are shorter than latencies of ratings that were placed at the middle of the scale.

Using the statistical procedure that was applied by Tractinsky et al. (2006), we were able to replicate their results with our data (using the data from study 1): the more extreme the rating the lower the latency. However, in the following we highlight why Tractinsky et al.'s procedure is problematic from a statistical and methodological point of view, and we suggest an alternative way to analyze the postulated relationship.

At first sight, our results seem to back up the inverted U-shaped relation between latency and beauty ratings. However, we think that the applied analysis is problematic and it does not necessarily allow conclusions to be drawn about the process of aesthetic judgment formation. This is because the response latencies of all participants of all websites were analyzed together without basing the analysis on individual participants. Hence, the inverted U-shaped curve that was identified does not necessarily reflect the rating behavior of a user; the postulated relationship could simply be an artefact of data aggregation. The major problem is that the ratings of the participants are not equally distributed among the different rating scale points, meaning that some participants might have influenced the mean response latency of a specific scale-point more than others. According to our data for instance, 16 of the total 58 ratings on the right end of the rating scale stem from a single participant. In other words, 28 percent of the ratings on that scale-point come from a single person, whereas there were seven participants who did not even provide a single rating on that point of the scale. Nevertheless, the aggregated response latencies of that scale-point are interpreted as the mean response latency at the right end of the rating scale. This issue applies for all scale-points: it is not transparent as to how many response latencies of how many participants form the mean response latencies of the different scale-points. Hence, we are ignorant of what we are actually statistically comparing. This makes it really hard to properly understand the meaning of these results.

Further, the applied between-subject ANOVA is also problematic; it contains 2261 data-points that only stem from 19 participants (500 ms condition); it is not guaranteed that the data are independent, which is crucial for such analyses. The resulting 2243 degrees of freedom error are therefore heavily inflated, which in turn makes the effect statistically "highly" significant. Despite the low p-value, the explained variance of the effect is rather low ($\eta_p^2 = .04$ for our data; $\eta_p^2 = .02$ for Tractinsky's data; both effects low).

Fig. 6. Scatterplot of the relation between response latency and beauty rating for three different participants.



Because of the above mentioned problems, we suggest investigating the relationship between latency and beauty ratings on the level of single participants. If the suggested inverted U-shaped curve reflects an aspect of the user's aesthetic perception, it must also be found on the level of individual participants and not solely as an aggregated effect. Consequently, we first log-transformed the latencies (to reduce the skewness of the distribution) and then plotted them against the ratings for each participant separately and examined the scatterplots carefully. Only very few participants showed the data pattern postulated by Tractinsky et al. (2006). Exemplarily, Figure 6 shows a participant whose data fit pretty well the inverted U-shaped curve (left), and data of participants with a poor fit (middle and right).

To estimate how frequently such inverted U-shaped patterns occur, we calculated for each participant how well their data would fit a quadratic instead of a linear function. As can be seen in Table 6, a significantly better quadratic relation between latency and beauty ratings occur only among 18 to 28 percent of the participants (only quadratic functions fitting an inverted U-shaped course were included; four participants even showed an U-shaped pattern). The averaged additional explained variances for quadratic instead of linear functions (ΔR^2) are very low for all conditions (2.0 to 3.3 %). Even for participants showing a significantly better fit to the quadratic function (*sign.* ΔR^2), the averaged explained variances from their data (ranging from 9.3 to 12.4 percent) are rather moderate.

6 Discussion

The results of study 1 and 2 clearly identify visual complexity (VC) and prototypicality (PT) as important factors for aesthetic perception of websites in the context of first impression. Websites of high visual complexity lead to a more negative first impression than websites of medium or low complexity and prototypical websites create a better first impression than less prototypical ones.

Table 6
Estimated fits (R^2) for linear and quadratic functions of the relation between response latency and beauty rating.

		50 ms*	500 ms	1000 ms
R^2 *(linear)*	*M* (*SD*)	.021 (.038)	.012 (.021)	.026 (.027)
	Range	< .001 to .131	< .001 to .071	< .001 to .099
R^2 *(quadratic)*	*M* (*SD*)	.041 (.049)	.045 (.054)	.048 (.041)
	Range	.001 to .173	.004 to .184	.019 to .114
ΔR^2	*M* (*SD*)	.020 (.064)	.033 (.052)	.021 (.026)
	Range	< .001 to .103	< .001 to .179	< .001 to .077
sign. ΔR^2	Sign. funct.	18 %	26 %	28 %
	M (*SD*)	.093 (.059)	.124 (.042)	.097 (.032)

Note. R^2 was computed for each participant separately and then averaged;
* four participants showed a sign. quadratic relation, but in the opposite direction, these cases are excluded for this table; ΔR^2 = the amount of additional variance explained by a quadratic instead of a linear function; *sign.* ΔR^2 = data pattern that significantly better fits a quadratic instead of a linear function

Both factors affect aesthetic judgments already after a perception time of 17 ms. It seems that within this time, perceivers are able to grasp information about the VC and the PT of a website and base their judgments on this information. However, the effect of PT on aesthetic perception is more pronounced within web pages of low VC, respectively the effect of VC on beauty is blunted within web pages of low PT. Hence, the combination of low VC and high PT leads to the highest beauty ratings. Moreover, there is some support for the reasoning drawn from the model of Leder et al. (2004) that VC is processed at an earlier stage than PT.

6.1 Visual complexity and prototypicality as predictors for aesthetic judgments

As expected and in line with previous research (Michailidou et al., 2008; Moshagen and Thielsch, 2010; Tuch et al., 2009, 2011), VC is a strong predictor for aesthetic judgments in general. Although the theory of Berlyne (1974) and results from experimental studies (Geissler et al., 2006) would suggest an inverted U-shaped relation between aesthetics and visual complexity, our data show a linear relation among these variables. This is not surprising as other studies dealing with websites as stimuli (Michailidou et al., 2008; Pandir and Knight, 2006; Tuch et al., 2009, 2011) also found a negative linear relationship between aesthetics and VC. In the context of Berlyne's theory, it seems that websites are in general rather complex stimuli and therefore lying on the right side of the inverted U-shaped curve. This would explain why there is a negative linear relation instead of a quadratic one. Besides VC, PT also has a strong influence on people's aesthetic judgments. In accordance with previous

findings from various research areas, our data suggest a positive relationship between PT and aesthetics. Prototypical company websites lead to better first impressions than atypical company websites. The effects of VC as well as PT on beauty appraisals were stable and very strong in both studies as can be inferred from the corresponding effect sizes (study 1: η_p^2 VC = .58 and η_p^2 PT = .81; study 2: η_p^2 VC = .52 and η_p^2 PT = .55).

Moreover, there was an interaction between VC and PT revealing that VC affects aesthetic perception stronger within websites of high PT, respectively the effect of PT on beauty is less pronounced within complex web pages than within less complex pages. It seems that as soon as VC is too high or PT too low, web pages receive lower beauty ratings, regardless of the characteristics of the other factor. Overall, websites of high PT and low VC are perceived as being the most beautiful, whereas websites of low PT and high VC trigger the worst first impression. There are some parallels to the findings of Hekkert et al. (2003), who observed that novelty only affects aesthetic judgments given high PT. There might be a similar mechanism between PT and VC.

6.2 VC and PT in the process of aesthetic perception

In regard to the information-processing stage model of aesthetic processing (Leder et al., 2004) we assumed that the effect of VC on aesthetic perception would occur prior to the effect of PT, because the processing of VC is supposed to happen at an earlier stage (stage of perceptual analysis) than the processing of PT (stage of implicit memory integration). Therefore, we expected the magnitude of the PT effect to increase with longer exposure times, whereas the magnitude of VC to remain relatively stable over the different presentation times.

In study 1 with exposure time above 50 ms we could not observe such a result pattern. However, in study 2 with exposure time below 50 ms the expected pattern occurred. Hence, our results give some support for the line of reasoning raised by Leder et al. (2004). Both, VC and PT affected aesthetic judgments already within 17 ms of exposure, but at this level the effect of VC was more pronounced than the one of PT. However, further studies with other stimulus features are needed to ensure the suggested processing stages of aesthetic perception.

6.3 The relation between response latency and beauty ratings

We aimed to replicate the findings of Tractinsky et al. (2006) on the relationship between response latencies and beauty ratings. By applying the same

analysis as Tractinsky et al. we were able to replicate their results: An inverted U-shaped relation between latencies and beauty ratings, meaning that latencies of very attractive or very unattractive web pages are shorter than latencies of ratings that are placed at the middle of the scale.

However, we point out several problematic issues about Tractinsky et al.'s analysis and show that their conclusion might result from an artifact of data averaging across subjects. Furthermore, we suggest an alternative way of analyzing the data. In contrast to Tractinsky et al., we based our analysis on the rating behavior of single participants instead of analyzing the pooled data and stimuli all together. Thereby, we could only partially back up the postulated relation. Our data suggested that only approximately 18 to 28 percent of the participants show a data pattern more or less matching a quadratic function. This means that most of the participants do not provide ratings faster at the extreme points than at the center of the scale. In the light of our results, the general claim that latency is an "unobtrusive measure of preferences and attitudes in HCI research" (Tractinsky et al., 2006, p. 1080) seems overstated. Nevertheless, there are participants who showed a relation between latency and beauty appraisal and it might be interesting to explore which characteristics lead to this phenomenon.

6.4 Limitations

There are at least four important limitations regarding this study:

(1) *Passive viewing.* The presented results are from a classical picture perception task often used in the context of first impression research. The study was conducted under a strictly controlled setting with standardized experimental conditions. Considering this, one limitation of this study is that the findings cannot be associated directly with how websites are perceived in real life. In real HCI situations (such as browsing the web), the actual interactions of users – not the passive viewing – has a major influence in shaping the overall user experience. Nevertheless, when first visiting a website, initial impressions are formed that are probably comparable to the screenshot-viewing task we employed in our study. The used paradigm enables us to look into the initial process of aesthetic judgment formation.
(2) *Demographic factors.* Another limitation of the present study is that cultural aspects of preference of website design are disregarded. The influence of VC as well as the shape of PT might vary for other cultures. Similar limitations can be stated for other demographic factors such as age (our sample consisted of a rather young population) or computer proficiency.

(3) *Website types.* The presented study was conducted with company websites. It remains to be seen if similar results can be found for other website types where users have formed stable mental models such as online shops or news websites (Roth et al., 2010).
(4) *Underlying factors of complexity and prototypicality* Complexity and prototypicality are influenced by many factors (e.g. form, color, shape, location to name just a few). In this study these underlying factors are not controlled, analyzed or understood in depth, nor can we derive conclusions about which factors lead to high or low complexity/prototypicality. Future studies should try to understand the various components that influence these constructs.

6.5 Conclusions

In this study we present initial results showing that visual complexity as well as prototypicality play a crucial role in the aesthetic judgment formation process. Users prefer websites with low visual complexity and high prototypicality. Websites of low prototypicality are generaly judged as being unattractive – in fact, this counts for websites of high as well as low complexity. Both factors already influence the aesthetic judgments after a very short presentation time of 17 ms.

This study also shows that there is great potential in the paradigm of Lindgaard et al. (2006). Using systematically selected or manipulated stimuli, it is possible to further explore and understand what happens during the process of first impressions. Further studies are needed to understand other factors such as colour, grouping, structure, or amount of text and pictures. The paradigm may also be used to expand the studies to other judgments such as trust (see e.g. Albert et al., 2009).

In sum, our study shows that designers should regard not only visual complexity, but also the factor prototypicality very carefully when designing a website. Designs that contradict what users typically expect of a website may trigger a suboptimal first impression and impair users' expectations. Latest research shows that negative product expectations lead to lower satisfaction in product interaction (Raita and Oulasvirta, 2010, 2011). This may lead to a disadvantageous negative downward spiral that should be avoided.

Acknowledgements

The authors would like to thank Anna Hanchar for conducting the experiments for study 2.

References

Akalin, A., Yildirim, K., Wilson, C., Kilicoglu, O., 2009. Architecture and engineering students' evaluations of house façades: Preference, complexity and impressiveness. Journal of Environmental Psychology 29 (1), 124–132.

Albert, W., Gribbons, W., Almadas, J., 2009. Pre-Conscious Assessment of Trust: A Case Study of Financial and Health Care Web Sites. In: Human Factors and Ergonomics Society Annual Meeting Proceedings. Vol. 53. Human Factors and Ergonomics Society, pp. 449–453.

Bar, M., Neta, M., Linz, H., 2006. Very first impressions. Emotion 6 (2), 269–278.

Bargas-Avila, J., Hornbæk, K., 2011. Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. In: Proceedings of the 2011 annual conference on Human factors in computing systems. ACM, pp. 2689–2698.

Berlyne, D., 1974. Studies in the new experimental aesthetics. Washington, D.C.: Hemisphere Pub. Corp.

Bruner, G. C., Kumar, A., 2000. Web commercials and advertising hierarchy-of-effects. Journal of Advertising Research 40 (1), 35–42.

Cantor, N., Thomas, E., 1976. Visual masking effects on duration, size, and form discrimination. Attention, Perception, & Psychophysics 19 (4), 321–327.

Crites, S., Fabrigar, L., Petty, R., 1994. Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. Personality and Social Psychology Bulletin 20 (6), 619–634.

de Angeli, A., Sutcliffe, A., Hartmann, J., 2006. Interaction, usability and aesthetics: what influences users' preferences? In: DIS '06: Proceedings of the 6th ACM conference on Designing Interactive systems. ACM Press, New York, NY, USA, pp. 271–280.

Edmonds, B., 1995. What is complexity? the philosophy of complexity per se with application to some examples in evolution. In: What is Complexity? Kluwer, Dordrecht.

Etcoff, N., 1999. Survival of the prettiest: The science of beauty. Anchor Books/Doubleday.

Eysenck, H., 1941. The empirical determination of an aesthetic formula. Psychological Review 48 (1), 83.

Fei-Fei, L., Iyer, A., Koch, C., Perona, P., 2007. What do we perceive in a glance of a real-world scene? Journal of Vision 7 (1), 1–29.

Geissler, G. L., Zinkhan, G. M., Watson, R. T., 2006. The influence of home page complexity on consumer attention, attitudes, and purchase intent. Journal of Advertising 35 (2), 69–80.

Goldstein, E., 2009. Sensation and perception, 8th Edition. Wadsworth Pub Co.

Hassenzahl, M., Monk, A., 2010. The inference of perceived usability from beauty. Human-Computer Interaction 25 (3), 235–260.

Hekkert, P., Snelders, D., Van Wieringen, P., 2003. Most advanced, yet acceptable: Typicality and novelty as joint predictors of aesthetic preference in industrial design. British Journal of Psychology 94 (1), 111–124.

Hekkert, P., Wieringen, P., 1990. Complexity and prototypicality as determinants of the appraisal of cubist paintings. British Journal of Psychology 81, 483–495.

Karvonen, K., Cardholm, L., Karlsson, S., 2000. Cultures of trust: a cross-cultural study on the formation of trust in an electronic environment. In: Proceedings of the Nordic Workshop on Secure IT Systems, Reykjavik, Iceland.

Kim, H., Fesenmaier, D., 2008. Persuasive design of destination web sites: An analysis of first impression. Journal of Travel Research 47 (1), 3–13.

Lavie, T., Tractinsky, N., 2004. Assessing dimensions of perceived visual aesthetics of web sites. International Journal of Human-Computer Studies 60 (3), 269–298.

Leder, H., Belke, B., Oeberst, A., Augustin, D., 2004. A model of aesthetic appreciation and aesthetic judgments. British Journal of Psychology 95 (4), 489–508.

Leder, H., Carbon, C., 2005. Dimensions in appreciation of car interior design. Applied Cognitive Psychology 19 (5), 603–618.

Leder, H., Carbon, C., Ripsas, A., 2006. Entitling art: Influence of title information on understanding and appreciation of paintings. Acta psychologica 121 (2), 176–198.

Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., Noonan, P., 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. ACM Transactions on Computer-Human Interaction (TOCHI) 18 (1), 1–30.

Lindgaard, G., Fernandes, G., Dudek, C., Brown, J., 2006. Attention web designers: You have 50 milliseconds to make a good first impression! Behaviour and Information Technology 25 (2), 115–126.

Lindgaard, G., Litwinska, J., Dudek, C., 2008. Judging web page visual appeal: Do east and west really differ? In: Proc. IADIS '08. pp. 157–164.

Martindale, C., 1984. The pleasures of thought: A theory of cognitive hedonics. Journal of Mind and Behavior 5 (1), 49–80.

Martindale, C., Moore, K., Borkum, J., 1990. Aesthetic Preference: Anomalous Findings for Berlyne's Psychobiological Theory. The American Journal of

Psychology 103 (1), 53–80.
Michailidou, E., Harper, S., Bechhofer, S., 2008. Visual complexity and aesthetic perception of web pages. In: Proc. SIGDOC '08. ACM, pp. 215–224.
Moshagen, M., Thielsch, M., 2010. Facets of visual aesthetics. International Journal of Human-Computer Studies 68 (10), 689–709.
Pandir, M., Knight, J., 2006. Homepage aesthetics: The search for preference factors and the challenges of subjectivity. Interacting with Computers 18, 1351–1370.
Plous, S., 1993. The psychology of judgment and decision making. McGraw-Hill New York.
Rabin, M., Schrag, J., 1999. First Impressions Matter: A Model of Confirmatory Bias. Quarterly Journal of Economics 114 (1), 37–82.
Raita, E., Oulasvirta, A., 2010. Too Good To Be Bad: The Effect of Favorable Expectations on Usability Perceptions. In: Proc. HFES '10.
Raita, E., Oulasvirta, A., 2011. Too good to be bad: Favorable product expectations boost subjective usability ratings. Interacting with Computers 23 (4), 363–371.
Reber, R., Schwarz, N., Winkielman, P., 2004. Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? Personality and Social Psychology Review 8 (4), 364.
Robins, D., Holmes, J., 2008. Aesthetics and credibility in web site design. Information Processing & Management 44 (1), 386–399.
Rolke, B., Hofmann, P., 2007. Temporal uncertainty degrades perceptual processing. Psychonomic Bulletin & Review 14 (3), 522–526.
Roth, S., Schmutz, P., Pauwels, S., Bargas-Avila, J., Opwis, K., 2010. Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages? Interacting with Computers 22 (2), 140–152.
Saadé, G., Otrakji, C., 2007. First impressions last a lifetime: effect of interface type on disorientation and cognitive load. Computers in Human Behavior 23 (1), 525–535.
Schenkman, B., Jönsson, F., 2000. Aesthetics and preferences of web pages. Behaviour and Information Technology Information Technology 19 (5), 367–377.
Sen, D., Lindgaard, G., 2008. The impact of prototypicality on aesthetic evaluation: Towards a better understanding of affective components of interface design. In: Proc. IADIS '08. pp. 263–267.
Stevenson, J. S., Bruner, G. C., Kumar, A., 2000. Webpage background and viewer attitudes. Journal of Advertising Research 40 (1), 29–34.
Thielsch, M., Hirschfeld, G., 2010. High and low spatial frequencies in website evaluations. Ergonomics 53 (8), 972–978.
Thielsch, M., Hirschfeld, G., 2012. Spatial frequencies in aesthetic website evaluations - explaining how ultra-rapid evaluations are formed. Ergonomics 55 (7), 731–742.
Tractinsky, N., Cokhavi, A., Kirschenbaum, M., Sharfi, T., 2006. Evaluating

the consistency of immediate aesthetic perceptions of web pages. International Journal of Human-Computer Studies 64 (11), 1071–1083.

Tuch, A., Bargas-Avila, J., Opwis, K., 2010. Symmetry and aesthetics in website design: It's a man's business. Computers in Human Behavior 26 (6), 1831–1837.

Tuch, A., Kreibig, S., Roth, S., Bargas-Avila, J., Opwis, K., Wilhelm, F., 2011. The role of visual complexity in affective reactions to web pages: Subjective, eye movement, and cardiovascular responses. IEEE Transactions on Affective Computing 2 (4), 230–236.

Tuch, A., Roth, S., Hornbæek, K., Opwis, K., Bargas-Avila, J., 2012. Is beautiful really usable? Toward Understanding the Relation Between Usability, Aesthetics, and Affect in HCI. Computers in Human Behavior 28, 1596–1607.

Tuch, A. N., Bargas-Avila, J. A., Opwis, K., Wilhelm, F., 2009. Visual complexity of websites: effects on users' experience, physiology, performance, and memory. International Journal of Human-Computer Studies 67 (9), 703–715.

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

Whitfield, T., 2000. Beyond prototypicality: Toward a categorical-motivation model of aesthetics. Empirical Studies of the Arts 18 (1), 1–11.

Xing, J., Manning, C., 2005. Complexity and automation displays of air traffic control: Literature review and analysis. Tech. rep., U.S. Department of Transportatioin, Office of Aerospace Medicine.

Zijlstra, F., 1993. Efficiency in work behaviour: A design approach for modern tools. University of Technology. Delft.

7 Appendix

Table 7
List of the Websites that were used in the study

Condition	Name of Company	Website URL	Familiarity of the Website [1]	Popularity of the Company / Brand [2]
			M (SD)	*M (SD)*
VC = low	rssl	http://www.rssl.com/	1.10 (.32)	1.30 (.95)
PT = low	Dow AgroSciences	http://www.dowagro.com/	1.20 (.63)	1.90 (1.52)
	Start Here, Go Places. AICPA	http://www.startheregoplaces.com/	1.00 (.00)	1.00 (.00)
	Orica	http://www.orica.com/	1.00 (.00)	1.00 (.00)
	CHEM research	http://www.chem-research.com/	1.00 (.00)	1.00 (.00)
	Benchmark Services, Inc.	http://www.oilflush.com/	1.00 (.00)	1.21 (.80)
	Progress Energy	https://www.progress-energy.com/	1.00 (.00)	1.00 (.00)
	AmeriCredit	http://www.americredit.com/	1.20 (.63)	1.50 (.85)
	breedlove and associates	http://www.mybreedlove.com/	1.10 (.32)	1.20 (.63)
	Exchange Consulting Group	http://www.exchangeconsulting.com/	1.00 (.00)	1.00 (.00)
	leadsource	http://www.leadsource.com/	1.07 (.27)	1.07 (.27)
	Action Fast Bail Bonds	http://www.actionfastbailbonds.com/	1.00 (.00)	1.00 (.00)
	Honda	http://www.honda.com/	1.10 (.32)	4.10 (1.66)
	aptera	http://www.aptera.com/	1.00 (.00)	1.00 (.00)
	Classic Unimogs / Classic 4x4	http://classicunimogs.com/	1.00 (.00)	1.00 (.00)
	Blizzard	http://www.blizzardplows.com/	1.00 (.00)	1.29 (.64)
	Sno-Gate	http://www.sno-gate.com/	1.00 (.00)	1.00 (.00)
	Yellowstone Track Systems, Inc.	http://www.yellowstonetrack.com/	1.00 (.00)	1.14 (.65)
	Donmar Sunroofs & Accessories	http://www.donmar.com/	1.00 (.00)	1.00 (.00)
	vehicle controls	http://www.vehiclecontrols.com/	1.00 (.00)	1.00 (.00)
	Group Mean		*1.04 (.07)*	*1.29 (.70)*
VC = low	Quintiles	http://www.quintiles.com/	1.00 (.00)	1.00 (.00)
PT = high	Sherwin-Williams	http://www.sherwin-williams.com/	1.00 (.00)	1.00 (.00)
	engro corp	http://engro.com/	1.00 (.00)	1.00 (.00)
	Novasyn Organics	http://www.novasynorganics.com/	1.20 (.63)	1.20 (.63)
	Sabic	http://www.sabic.com/corporate/en	1.00 (.00)	1.00 (.00)
	Hebei Yanuo	http://www.yanuo.com/	1.00 (.00)	1.00 (.00)
	Ameresco	http://www.ameresco.com/	1.00 (.00)	1.00 (.00)
	Northeast Utilities System	http://www.nu.com/	1.00 (.00)	1.05 (.22)
	National Heat Exchange Cleaning Corporation	http://www.nationalheatexchange.com/	1.00 (.00)	1.00 (.00)
	PG&E	http://www.pge.com/	1.00 (.00)	1.00 (.00)
	Chevrolet	http://www.chevrolet.com/#cruze	1.00 (.00)	2.33 (1.97)
	AIAM	http://www.globalautomakers.org/	1.00 (.00)	1.00 (.00)
	GEM	http://www.polarisindustries.com/	1.00 (.00)	1.14 (.36)
	tesla	http://www.teslamotors.com/	1.50 (1.27)	2.10 (1.60)
	MAFS	https://www.usemafs.com/	1.00 (.00)	1.00 (.00)
	Honda Motor	http://powersports.honda.com/	1.00 (.00)	2.92 (2.02)
	JVC USA	http://www.jvc.com/	1.40 (1.26)	4.30 (1.34)
	Pioneer	http://www.pioneerelectronics.com/	1.14 (.36)	3.38 (1.72)
	Allete	http://www.allete.com/	1.00 (.00)	1.00 (.00)
	Fantasy Junction	http://www.fantasyjunction.com/	1.00 (.00)	1.00 (.00)
	Group Mean		*1.06 (.14)*	*1.52 (.97)*

Note. [1] range: 1 ("not at all") to 5 ("very often"); [2] range: 1 ("not at all") to 5 ("very good"); the language of all websites was English.

Table 7
List of the Websites that were used in the study (Continued)

Condition	Name of Company	Website URL	Familiarity of the Website [1]	Popularity of the Company / Brand [2]
			M (SD)	*M (SD)*
VC = medium	Cynotech	http://www.cyanotech.com/	1.00 (.00)	1.05 (.22)
PT = low	Seventh Generation	http://www.seventhgeneration.com/	1.00 (.00)	1.00 (.00)
	SynChem, Inc.	http://www.synchem.com/	1.00 (.00)	1.10 (.44)
	Interspec, LLC	http://www.interspecllc.net/	1.00 (.00)	1.00 (.00)
	Dow Corning	http://www.dowcorning.com/	1.05 (.22)	1.20 (.89)
	FEM Elecctric Association, Inc.	http://www.femelectric.coop/	1.00 (.00)	1.00 (.00)
	Ge Energy	http://www.ge-energy.com/	1.00 (.00)	2.36 (1.6)
	BizEE	http://www.bizeesoftware.com/	1.00 (.00)	1.00 (.00)
	OB10 e-Invoicing	http://www.ob10.com/Country/US/	1.00 (.00)	1.00 (.00)
	Experian	http://www.experian.com/	1.40 (1.26)	1.20 (.63)
	Islamic Banking & Finance	http://islamicbankingandfinance.com/	1.00 (.00)	1.29 (1.07)
	Bisk Educaton	http://www.bisk.com/	1.00 (.00)	1.00 (.00)
	Lucas Oil Products Inc.	http://www.lucasoil.com/	1.20 (.63)	1.20 (.63)
	AutoRevo	http://www.autorevo.com/	1.00 (.00)	1.00 (.00)
	Autotomorrow.com	http://www.autotomorrow.com/	1.40 (1.26)	1.20 (.63)
	bajaj	http://www.bajajauto.com/	1.00 (.00)	1.00 (.00)
	Plows Unlimited	http://www.plowsunlimited.com/archive/	1.00 (.00)	1.00 (.00)
	Powermadd	http://www.powermadd.com/	1.00 (.00)	1.00 (.00)
	Spartan Motors	http://www.spartanmotors.com/	1.00 (.00)	1.00 (.00)
	Dr. T's Accounting Problems and Business Examples	http://www.drtaccounting.com/	1.1 (.32)	1.10 (.32)
	Group Mean		*1.06 (.13)*	*1.13 (.30)*
VC = medium	Intacct AICPA business solutions	http://us.intacct.com/	1.00 (.00)	1.14 (.65)
PT = high	Praxair	http://www.praxair.com/	1.00 (.00)	1.00 (.00)
	Behr	http://www.behr.com/Behr/home	1.00 (.00)	1.00 (.00)
	Valspar	http://www.valsparpaint.com/	1.40 (1.26)	1.2 (.63)
	Deloitte	http://www.deloitte.com/	1.00 (.00)	2.00 (1.70)
	Ballard	http://www.ballard.com/	1.00 (.00)	1.00 (.00)
	FuelCell Energy	http://www.fuelcellenergy.com/	1.00 (.00)	1.00 (.00)
	Powerspan	http://www.powerspan.com/	1.00 (.00)	1.00 (.00)
	National Rural Utilities Cooperative Finance Corporation	http://www.nrucfc.coop/	1.00 (.00)	1.00 (.00)
	FTI Consulting	http://www.fticonsulting.com/	1.00 (.00)	1.00 (.00)
	J.P. Morgan	http://www.jpmorgan.com/	1.00 (.00)	2.20 (1.62)
	ocra worldwide	http://www.ocra.com/	1.14 (.53)	1.29 (1.07)
	paychex	http://paychex.com	1.00 (.00)	1.00 (.00)
	Bombardier	http://www.bombardier.com/	1.05 (.22)	2.65 (1.73)
	Edelbrock	http://www.edelbrock.com/	1.00 (.00)	1.00 (.00)
	ducati	http://www.ducati.com/	1.07 (.27)	3.14 (1.96)
	Harley-Davidson USA	http://www.harley-davidson.com/	1.08 (.29)	2.58 (1.98)
	ANSA	http://ansaautomotive.com/	1.00 (.00)	1.20 (.42)
	EBD Group	http://www.ebdgroup.com/	1.00 (.00)	1.00 (.00)
	PPD	http://www.ppdi.com/	1.00 (.00)	1.00 (.00)
	Group Mean		*1.04 (.09)*	*1.42 (.68)*

Note. [1] range: 1 ("not at all") to 5 ("very often"); [2] range: 1 ("not at all") to 5 ("very good"); the language of all websites was English.

Table 7
List of the Websites that were used in the study (Continued)

Condition	Name of Company	Website URL	Familiarity of the Website [1]	Popularity of the Company / Brand [2]
			M (SD)	*M (SD)*
VC = high	Airgas	http://www.airgas.com/	1.00 (.00)	1.00 (.00)
PT = low	Sensient Technologies Corporation	http://www.sensient-tech.com/	1.00 (.00)	1.00 (.00)
	Abraxas Energy Consulting	http://www.abraxasenergy.com/	1.00 (.00)	1.00 (.00)
	First European	http://www.first-european.co.uk/	1.00 (.00)	1.05 (.22)
	American Express	https://www.americanexpress.com/	1.08 (.29)	3.17 (1.53)
	Chase	https://www.chase.com/	1.00 (.00)	1.29 (.83)
	Bank of America	https://www.bankofamerica.com/	1.14 (.48)	2.67 (1.74)
	Bureau Van Dijk	http://www.bvdinfo.com/	1.00 (.00)	1.10 (.32)
	Geico	http://www.geico.com/	1.30 (.95)	1.20 (.63)
	SNL Financial	http://www.snl.com/	1.20 (.63)	1.22 (.67)
	Lloyds TSB	http://www.lloydstsb-offshore.com/	1.29 (1.07)	1.64 (1.34)
	Killer IRS Representative Joe Mastriano, CPA	http://www.taxproblem.org/	1.30 (.95)	1.20 (.63)
	Honeywell	http://honeywell.com/Pages/Home.aspx	1.00 (.00)	1.57 (1.28)
	Military Suppliers & News	http://www.armedforces-int.com/	1.05 (.22)	1.50 (1.15)
	izmocars	http://www.izmocars.com/	1.00 (.00)	1.00 (.00)
	eBizAutos	http://www.ebizautos.com/	1.00 (.00)	1.00 (.00)
	Freedom off Road	http://www.freedomoffroad.com.au/	1.00 (.00)	1.00 (.00)
	Horschel Companies	http://www.hbpllc.com/	1.00 (.00)	1.29 (1.07)
	Snowcare for Tropps	http://projectevergreen.com/scft/	1.00 (.00)	1.21 (.80)
	Chain auto tool manufacturer & exporter	http://www.chain-auto-tools.com/	1.30 (.95)	1.30 (.95)
	Group Mean		*1.08 (.12)*	*1.37 (.57)*
VC = high	Gen-Probe	http://www.gen-probe.com/	1.00 (.00)	1.00 (.00)
PT = high	GlaxoSmithKline	http://www.gsk.com/	1.30 (.67)	3.30 (1.77)
	Bayer Crop Science	http://www.bayercropscience.com/	1.00 (.00)	3.07 (1.73)
	Air Procuts	http://www.airproducts.com/	1.00 (.00)	1.00 (.00)
	Genecor	http://www.genencor.com/	1.00 (.00)	1.07 (.27)
	Symyx (now Accelrys)	http://accelrys.com/	1.00 (.00)	1.00 (.00)
	FPL	http://www.fpl.com/	1.20 (.63)	1.20 (.63)
	KPMG	http://www.kpmg.com/	1.17 (.39)	1.83 (1.59)
	BIC Alliance	http://www.bicalliance.com/	1.00 (.00)	1.00 (.00)
	D&B	http://www.dnb.com/	1.30 (.95)	1.20 (.63)
	BusinessFinance	http://businessfinancemag.com/	1.00 (.00)	1.19 (.51)
	Morgan Stanley Smith Barney	https://www.morganstanleyclientserv.com/	1.30 (.95)	2.20 (1.62)
	HSBC	http://www.hsbc.com/1/2/	1.15 (.37)	2.71 (1.74)
	taxback.com	http://www.taxback.com/	1.00 (.00)	1.00 (.00)
	Air Transport Association	http://www.airlines.org/pages/home.aspx	1.00 (.00)	1.00 (.00)
	United Technologies	http://www.utc.com/Home	1.00 (.00)	1.00 (.00)
	Mar-Kee Group	http://markeegroup.com/	1.30 (.95)	1.10 (.32)
	AdPerfect	http://www.adperfect.com/	1.00 (.00)	1.00 (.00)
	CPEcredit Interactive Online CPE Courses	http://cpecredit.com/scripts/index.asp	1.00 (.00)	1.00 (.00)
	Group Mean		*1.09 (.13)*	*1.44 (.76)*

Note. [1] range: 1 ("not at all") to 5 ("very often"); [2] range: 1 ("not at all") to 5 ("very good"); the language of all websites was English.