Bachelor thesis:

Hedonic Quality: The Inference and perspective processing approach.

Quincy Boom

s1099477

June 2013

University of Twente

Faculty Behavioral sciences

Psychology

First supervisor: Dr. Martin Schmettow

Second supervisor: Inga Schwabe

Index

ABSTRATCT

Dit artikel is gebaseerd op een onderzoek door Tuch et al. (2012). In dit artikel wordt de rol van visuele complexiteit (VC) en prototypicaliteit (PT) op de van websites op de hedonic quality (HQ) oordeel van gebruikers. Gedurende dit onderzoek worden 76 websites gepresenteerd die variëren in VC en PT. Deze 76 websites moesten in totaal vier keer beoordeeld worden op HQ, gedurende verschillende presentatie-tijden (17 ms, 33 ms, 500 ms en zonder limiet). De resultaten van het experiment toonden aan dat VC en PT HQ oordelen van gebruikers beïnvloedden zelfs in de 17 ms conditie. Naarmate de presentatie-tijden toenamen, nam het effect van VC op HQ oordelen geleidelijk af. Deze resultaten ondersteunen de informatieverwerking stadium model van esthetische verwerking (Leder et al., 2004), door te laten zien dat informatie in verschillende stadia word beoordeeld. Wanneer presentatie-tijden toenemen blijft het effect van PT op HQ oordelen echter gelijk, dit in tegenspraak met het model van Leder et al. (2004). Websites met een lage VC en hoge PT beloond met hogere HQ scores, in het bijzonder wanneer presentatie-tijden toenamen. Over het algemeen lijkt het alsof HQ oordelen op een iets andere manier gevormd worden dan esthetische oordelen. Dit is in tegenspraak met het inferentie perspectief voorgesteld door Hassenzahl & Monk (2010).

This paper replicates a study made by Tuch et al. (2012), investigating the role of visual complexity (VC) and prototypicality (PT) as website design factors, on shaping users' hedonic quality (HQ) judgments. During this study 76 website varying in VC and PT were randomly presented and rated on perceived HQ. These 76 websites were rated four times for varying presentation-times, 17, 33, 500 and without limit. Results showed that VC and PT affected HQ judgments, even when websites were presented for 17 ms. As presentation-times increase, the effect of VC declines steadily, supporting the reasoning of the information-processing stage model of aesthetic processing (Leder et al., 2004) that information is processed in different stages. As presentation-times increase the effect of PT remains stable, contradicting the stage model by Leder et al. (2004). Websites with low VC and high PT received high HQ ratings, especially when presentation-times were higher. Overall, it seems that HQ judgment is formed in a somewhat different fashion than beauty judgment, contradicting the occurrence of an inference effect as supposed by Hassenzahl & Monk (2010).

Introduction

Computers are vastly complex systems, which are used for many purposes. Be it for relaxation, education or work, users interact with computers to fulfill these personal needs. Within the field of Human Computer Interaction (HCI) focus lies on interaction between users and computers, making it a very important mean to assess the quality of different interactive systems. A research held within HCI is UX, the purpose of which is to assess the quality of programs by measuring User Experience (UX). User Experience consists of different dimensions, namely Generic UX, Affect, Enjoyment, Aesthetics, Hedonic quality, Engagement, Motivation, Enchantment and Frustration (Bargas-avila & Hornbaek, 2011). Apart from UX, usability is the most important criterion within HCI. By measuring usability, more insight can be gained on how experience users certain computer programs and what this user experience is caused by.

Despite usability scales being vastly used, it is not exactly clear how users generate ratings on these scales. Different studies have shown that usability correlates with beauty to a remarkably high degree (Tractinsky, Katz, & Ikar, 2000; Tracktinsky, Cokhavi, Kirschenbaum, & Sharfi, 2006). Hassenzahl & Monk (2010) explain the high correlation between beauty and usability using an inference perspective. The inference perspective supposes that when judging a product, inexperienced users may use all currently available information to infer the unavailable information. Because beauty relies heavily on sensory information, it is thought by Hassenzahl & Monk (2010) to be the starting-point for many inference processes. Hassenzahl & Monk (2010) thought this high correlation between beauty and usability is mainly caused by a mediating variable, namely ‘Goodness’. Hassenzahl & Monk (2010) indicated that when forming a usability score, beforehand a beauty score is generated. Beauty scores are then used to infer a general value, named a goodness score. Finally, the unavailable usability score is inferred from goodness. Hassenzahl & Monk (2010) found that unlike other usability aspects, Hedonic Quality was directly inferred from Beauty (Figure 1).

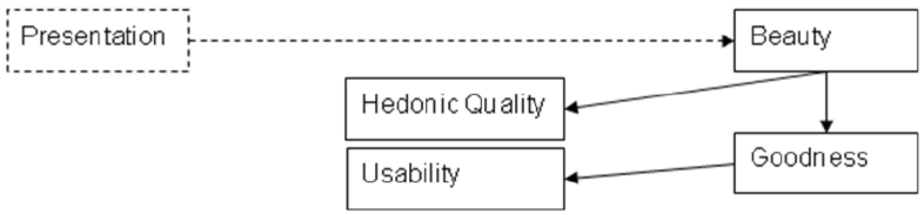

Fig. 1; Extended inference perspective by Hassenzahl & Monk (2010).

Tuch, Presslaber, Stöcklin, Opwis, & Bargas-Avila (2012) dispute this hypothesis by discussing the influence of visual complexity (VC) and prototypicality (PT) regarding first impressions of websites. The research of Tuch et al. (2012) extensively used a model portrayed by Leder et al. (2004). This model shows that the analysis of visual information occurs in different processing stages (Figure 2). Tuch et al. (2012) focused primarily on the first two stadia of the model, showing that VC and PT had a large effect on website beauty scores when websites were presented very shortly. Because VC and PT seem to influence beauty scores, it is highly likely that processing of other UX scales is similar to beauty processing.

In this current article the research of Tuch et al. (2012) was replicated. Users were shown the same websites as in Tuch et al. (2012), however in this study users rated the website according to a different UX dimension, namely Hedonic Quality. If Hedonic Quality judgment is indeed dependent of VC and PT, the role of VC and PT in UX scales would be extended.

Theoretical background

First impressions

Forming first impressions is found to be very important with regard to website perceptions of usability. The first instant an impression is formed people tend to stick by them by only seeking information that supports their initial hypothesis, while ignoring other information. This effect is called the 'confirmation bias' (Mynatt, Doherty, & Tweney, 1977). While the importance of first impressions is widely accepted, there is still a lot about first impressions that remains unknown. Lindgaard, Fernandes, Dudek, & Brown (2006) grant some clarity by conducting three studies to assess how fast people form an opinion about the visual appeal of webpages. Notable is that in Lindgaard et al. (2006) no masking was used. The first study was conducted to determine the reliability of visual appeal ratings and select website homepages to use during the second study. Study 1 showed that there was a high consistency between the visual appeal ratings collected during phase 1 and phase 2. All correlations were highly significant ($p < .001$), thus all participants' ratings were reliable.

During study 2 participants had to rate all 50 websites in random orders, during three phases. These websites differed on seven visual design characteristics. While viewing each webpage, visual appeal was rated. During the first two phases presentation times were 500 ms for each homepage. During the third phase participants could watch the homepages for as long as they desired. When comparing the scores on visual appeal to the seven design characteristics, it seemed that 5 out of 7 design characteristics correlated highly with visual appeal. The five visual characteristics that correlated highly with visual appeal also correlated highly with each other. Since 500 ms is a fairly long time to form a first impression, a third study was conducted. During Study 3 the same stimuli as in Study 2 were used, with the difference that in this study participants were randomly assigned to a 50 or 500 ms condition. Results showed significant results even in the 50ms condition. It thus seems that design characteristics influence visual appeal within 50ms when viewing websites.

In conclusion Lindgaard et al. (2006) has shown that visual appeal is influenced by the same design variables in all time-conditions. Since visual appeal in the 50ms condition is dependent of the same design variables as in the 500 ms or timeless condition, it seems that inference has not occurred despite information being less readily available. Since judgments were stable in all time-conditions, the likelihood of an inference effect occurring is very small. The absence of an inference effect in the study of Lindgaard et al. (2006) weakens assumptions made by Hassenzahl & Monk (2010). What still remained unclear however was how the relationship between design characteristics and visual appeal changed for different presentation-times, since only a 50ms and 500 ms presentation-time was used.

Information-processing stage model of aesthetic processing

The information-processing stage model of aesthetic processing proposes that there are five different processing stages when it comes to aesthetic evaluations, as shown in Figure 2.

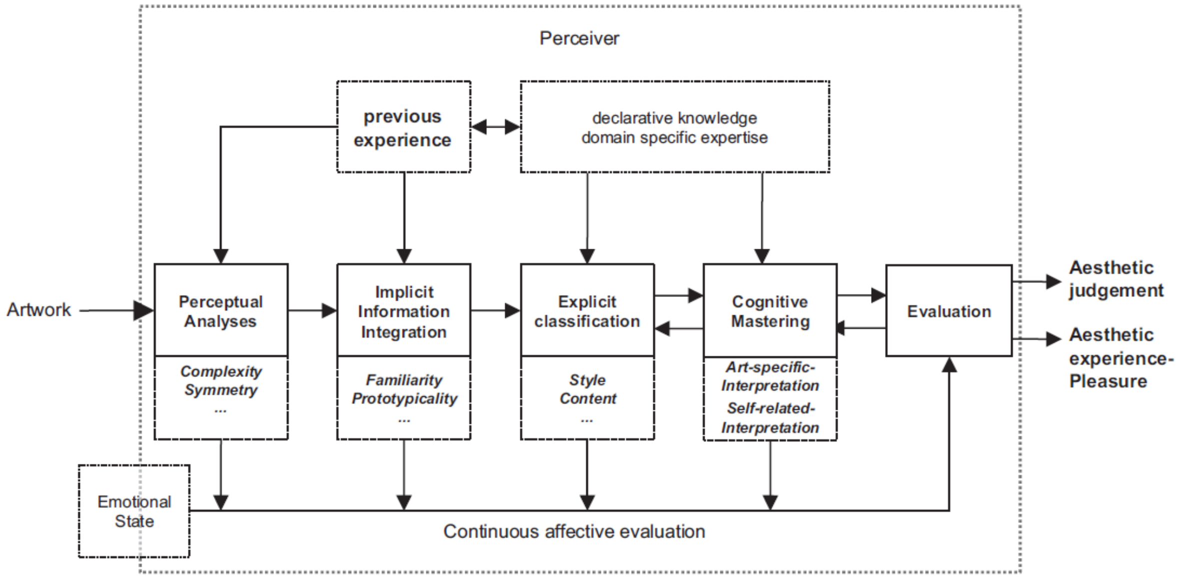

Fig. 2; Information-processing stage model of aesthetic processing (adapted from Leder et al., 2004)

Information processed in the perceptual analyses (1) and implicit information integration (2) stages are formed swiftly and do not have to become conscious thoughts in order to influence aesthetic processing. Explicit classification (3) and cognitive mastering (4) are more conscience processes which are affected by expertise and knowledge possessed by perceivers. Since expertise and knowledge play a smaller role in website perception than in art perception, explicit classification (3) and cognitive mastering (4) are considered irrelevant for the study conducted by Tuch et al. (2012). At stage five 'evaluation (5)' the processed information is evaluated leading to two different outputs of evaluation, namely aesthetic appraisal and aesthetic judgment. Aesthetic appraisal refers to how an object positively influences participants emotionally and aesthetic judgment refers to the artistic value perceivers believe the object has. Despite the model of Leder et al. (2004) originally being intended for art perception it is still useful for website perception, since variables such as complexity and prototypicality also appear in websites (Tuch et al., 2009; Tuch et al., 2012; Harper, Michailidou, & Stevens, 2009).

In Leder, Carbon, & Ripsas (2006) the information-processing stage model proposed in Leder et al. (2004) is used to assess the influence of title information on the understanding and appreciation of paintings. Leder et al. (2006) found that in titles did not affect appreciation when paintings were presented for 1s, 10s or 90s when paintings were presented for 1s descriptive titles increased understanding. In the 10s and 90s condition elaborative titles increased understanding of the paintings. These findings support the model by Leder et al. (2004), as it shows that information is processed in stages as presentation-times increase.

The information-processing stage model by Leder et al. (2004) opposes the inference perspective purposed by Hassenzahl & Monk (2010). The inference perspective by Hassenzahl & Monk (2010) shows that users process information using an inference process when information is not readily available. The information-processing stage model by Leder et al. (2004) shows that information is not inferred, information is just processed in stages. When not all information can be processed due to lack of time information is not inferred, information is just processed in lower stages than it would otherwise.

Visual complexity

The information-processing stage model of aesthetic processing does not represent a strict serial process in which all steps are strictly followed, yet it is possible to formulate hypotheses regarding timescale processing of a stimulus (Leder et al, 2006). Stage 1 and 2 within the model could thus be important in the formation of first impressions within a short timeframe. Processing within the perceptual analysis stage (1) is mostly dedicated by the visual complexity of the stimulus. Visual Complexity (VC) is a way to assess the degree of complexity within a design. A lot of research is committed into finding effects VC has on users (Reber, Schwarz, & Winkielman, 2004; Tuch et al., 2009; Tuch et al., 2012). According to Berlyne (1974) the relationship between viewer pleasure and arousal can be represented by an inverted U-curve, with medium levels of arousal being viewed as most pleasurable. Later studies try to verify these findings try to verify these results, but fail to offer any direct support (Tuch et al., 2009; Pandir & Knight, 2006). Many other effects of visual complexity have been researched. Tuch et al. (2009) indicates that VC is an important factor in web design as VC influences perceived pleasure and arousal, physiological response and recognition performance of the viewers.

Prototypicality

At the implicit information integration stage (2) aesthetic perception is mostly affected by stimulus characteristics involving the user's previous experiences, making perceived prototypicality important within this stage. Prototypicality refers to extent an objects represents a class of objects. Prototypicality thus represents mental models, which are based on experience. Many different domains show that people prefer prototypical objects, for instance in colors or paintings. Prototypical preference is even present in human relationships. It is found that when supervisors are more group prototypical, employees trust their coworkers more than wen supervisors are less group prototypical (Seppälä & Lipponen, 2012).

The influence of visual complexity and prototypicality in beauty

The study by Tuch et al. (2012) consisted of two studies. In the first study, 119 website-screenshots which varied in visual complexity (VC) and prototypicality (PT) were presented in either of three different timeframes (50 ms vs. 500 ms vs. 1000 ms). After perceiving a website, participants had to rate the website's beauty on a scale. The results of this experiment showed that there is an interaction effect between VC and PT on beauty. Low VC websites received higher beauty scores than high VC websites and high PT websites received higher beauty scores than low PT websites. The interaction between VC and PT in the 50 ms condition indicates that website information was processed in both Stage 1 and Stage 2 from the 'information-processing stage model of aesthetic processing' (Figure 2). The fact that both a perceptual analyses (1) and implicit information integration (2) occurred can be explained by the relatively long exposure times of 50ms. In order to verify if presentation-times were in fact to long, a second experiment with even shorter presentation-times was conducted.

In the second part of the study, the experiment was repeated with shorter presentation times (17 ms vs. 33 ms vs. 50ms). The results confirmed the effect of VC and PT on perceived beauty in all conditions, again indicating that low VC and high PT are found most beautiful. The effects of VC and PT at the 33 ms and 50ms condition almost the same, whereas at the 17 ms condition the effect of VC on beauty was higher than the effect of PT. This supports the model of Leder et al. (2004) that VC is processed earlier than PT. It also opposes the model of Hassenzahl & Monk (2010), as perceived beauty is influenced highly by VC and PT in short presentation times opposing the existence of an inference effect. Variance within perceived beauty scores during short presentation times is explained, even though information is not easily available to the participants. When combining research conducted by Tuch et al. (2012) and Leder et al. (2006) with the model of Leder et al. (2004) it is possible to indicate approximately when each stage starts (Figure 3).

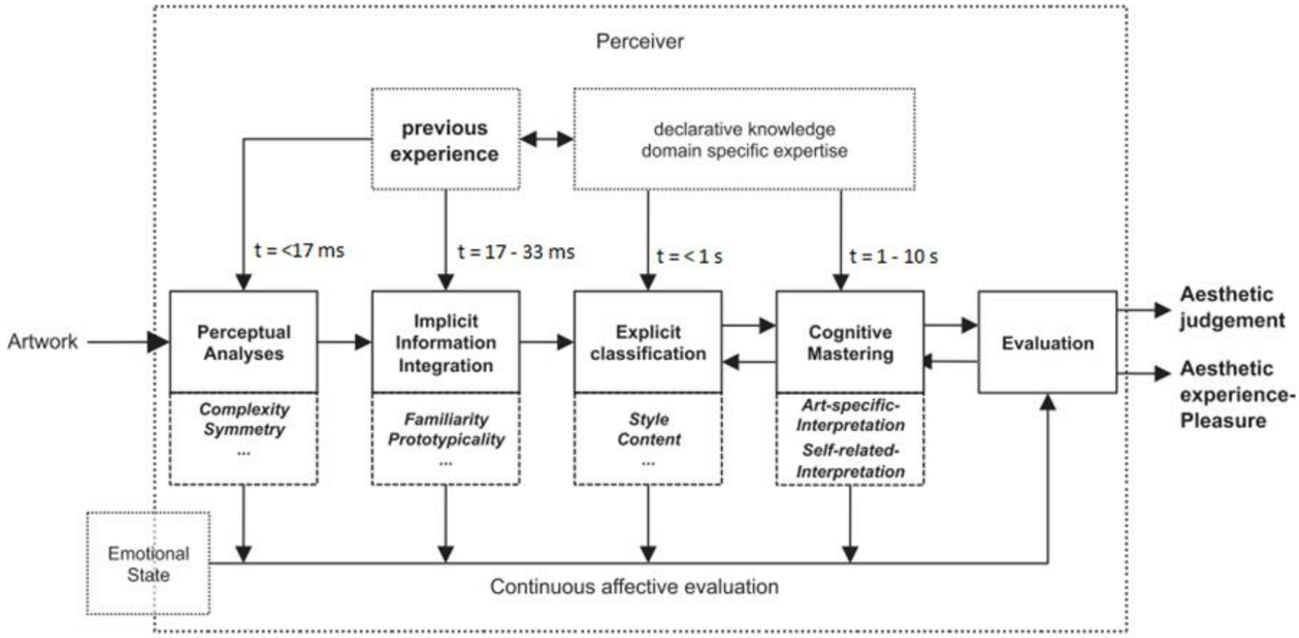

Fig. 3; Information-processing stage model of Leder et al. (2004) combined with results made by Tuch et al. (2012) and Leder et al. (2006).

Hedonic Quality

Hedonic Quality (HQ) is a subjective measure of the users' perceived quality, such as originality or innovativeness, seemingly having no direct relationship with the task related goals themselves. In order for a product to be usable it is not only important that the designs themselves function as intended, but users also need to perceive the product in the way designers had intended it. Thus Hedonic Quality clarifies the correspondence of intended and perceived product quality (Hassenzahl 2001).

Bernardo, Marimon, & Alonso-Almeida (2012) demonstrate the role of Hedonic Quality of e-commerce by examining the impact of Hedonic Quality of traveling websites on perceived value. Bernardo et al. (2012) stated that dimensions suggested by the E-S-QUAL instrument to influence e-service quality (efficiency, system availability, fulfillment and privacy) all focus on Functional Quality rather than Hedonic Quality. In order to assess the influence of Hedonic Quality in e-service quality, a structured questionnaire was administered by phone amongst 1201 consumers of online travel agencies in Spain. The questionnaire consisted of 30 items which were arranged in four constructs, namely E-service Quality, Hedonic Quality, Perceived Value and Loyalty. All items were presented as statements respondents could either agree or disagree with on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree). Results of this study show that although Functional Quality and Hedonic Quality both positively influence perceived value, Functional Quality has a higher influence than Hedonic Quality. Despite perceived value being primarily influenced by Functional Quality, Hedonic Quality contributes to perceived value as well, making it an important factor to focus on within e-commerce (Bernando et al., 2012; Lee & Kozar, 2009).

Hedonic Quality (HQ) can be divided into Hedonic Quality Stimulation (HQS) and Hedonic Quality Identification (HQI) (Hassenzahl, 2004). In this current study HQS refers to how creative or innovating the website designs are perceived. HQI relates to the nature of people to express themselves through objects. In this study it would refer to what people believe the website would express about them as a person. What the website says about how stylish or professional the person is.

Research Question

This study will identify to what extent VC and PT direct Hedonic Quality judgments. Since Tuch et al. (2012) found that VC and PT influence beauty judgments and numeral studies have found beauty and hedonic quality to be related to high extents (Hassenzahl & Monk, 2007; Schaik et al., 2012; Cogan, Parker & Zellner, 2013), we expect that VC and PT also influence Hedonic Quality judgments. This present study primarily focusses the extent to which Hedonic Quality is influenced by early information-processing stages. This study also identifies if HQI and HQS are processed in different fashions.

Results of this experiment will also clarify if a beauty inference such as mentioned in Hassenzahl & Monk (2010) occurs in Hedonic Quality judgment. If Hedonic Quality judgments are dependent of VC and PT in exactly the same fashion as beauty, it would support the inference perspective by Hassenzahl & Monk (2010). It would show that beauty relates to Hedonic Quality to such a high extent that inference is likely. If Hedonic Quality judgments are dependent of VC and PT in a somewhat different fashion than beauty, the information-processing stage model by Leder et al. (2004) would gain support. As it would show that even though Hedonic Quality perception occurs in the same stages as beauty perception, Hedonic Quality judgment occurs in a different fashion. This would make beauty inference within Hedonic Quality judgments more unlikely.

Methods

Design

This experiment closely resembles research made by Tuch et al. (2012), yet the procedure within this study is somewhat different as this study consists solely of within-subject independent variables (i.e. repeated measures design). Independent variables included *visual complexity* (low or high), *prototypicality* (low or high), *condition* (17 ms, 33 ms, 500 ms or without limit) and *scale* (HQI or HQS). Judgment ratings were used as the dependent variable. This experiment was shared with a researcher examining the role of VC and PT on credibility ratings, therefore the experiment consisted of both a hedonic quality and credibility scale.

Participants

A total of n = 25 participants (11 female and 14 male) participated in the experiment. 21 participants were students on the University of Twente. 3 participants were attending an HBO study and 1 participant quit its study after high school. The average age of the participants was 21.9 years old (SD = 3.3) and age varied between 17 and 31 years. Participants whom studied behavioral sciences could receive 1 credit for participating in our research.

Materials

Free software from http://www.psychopy.org/ was used to for experiment design and data collection. The experiment was executed using a 17” TFT monitor with a 60 Hz refresh frequency. TFT-screens are used in this study as it has been shown that legibility is better on TFT-screens than on CRT-screens (Nose et al., 1999; Shieh & Lin, 2000). The experiment was performed at a resolution of 1000 x 800 pixels.

Stimuli selection

This experiment used a subset of the stimuli used by Tuch et al. (2012). Only screenshots of company websites were used, as Roth, Schmutz, Pauwels, Bargas-Avila, & Opwis (2010) has shown that people have consistent mental models of such websites. In Tuch et al. (2012) a final pool of 120 websites was used. These websites differed visual complexity (low vs. medium vs. high) and prototypicality (low vs. high). The present study however, does not use stimuli with a medium level of VC. Since Tuch et al. (2012) indicated that VC had a linear effect on aesthetics it was very likely that VC also had a linear effect on HQ judgment, making it unnecessary to also include medium levels of VC. Eventually, 76 websites were used in this study differing in VC (low and high) and PT (low and high).

Rating

To assess participants' hedonic quality judgments a visual analogue scale was developed with different anchors for HQI and HQS. For HQI the following anchors were used: Isolating – integrating, amateurish – professional, gaudy – classy, cheap – valuable, noninclusive – inclusive, takes me distant from people – brings me closer to people, unpresentable – presentable. For HQS the anchors used were: Typical – original, standard – creative, cautious – courageous, conservative – innovative, lame – exciting, easy – challenging, commonplace - new. The Dutch translations of these anchors were taken from Hassenzahl (2004).

Procedure

The experiment was conducted on one computer in a small separate room at the University of Twente using the same computer. Participants were seated in front of a computer one at a time and were asked for their gender, age and education. Instructions about the experiment were given on screen during the experiment. Participants started the experiment with a brief 'practice phase', in which several images were shown. After each image, participants had to make a judgment ranging from negative to positive on a visual analogue scale. This judgment was made using the computer mouse. After finishing the 'practice phase', the actual experimental started.

The experiment consisted of 76 images of websites, which were shown in a random order for 17 ms in the first block. Goldstein (2009) showed that stimuli persist for approximately 250 ms after stimuli disappearance. Therefore, a visual mask was displayed to ascertain the accuracy of presentation times. After presenting the visual mask, participants needed to rate a random item from the credibility or hedonic quality scale on a visual analogue scale ranging from negative to positive. After all websites were rated, the experiment was conducted with the same 76 websites in another random order. Websites were now shown for a longer presentation times, namely 33 ms, 500 ms and without limit in a second, third and fourth block respectively.

Data analyses

To test if website complexity and prototypicality had any effect on subject's Hedonic Quality judgments a regression analysis was conducted. A 2 x 3 regression analysis was performed, where judgment was entered as a dependent variable. Independent variables within this analysis were Scale (HQI or HQS), and Condition (17 ms, 33 ms and 500 ms). Scores of the limitless condition was used as a Reference Judgment (RefJ). Since participants in the limitless condition had the most time to assess a judgment, this score would be closest to their 'true judgment'. Since Visual Complexity (VC) is known to negatively affect judgment scores, the item was reversed and renamed it Visual Simplicity (VS). VS, PT and RefJ were characterized as covariates, in order to clarify to what extent VS, PT and RefJ influence HQ judgments. The assumptions for regression analysis were met, as items and independent variables were previously validated in the research by Tuch et al. (2012).

To assess if there was any association between RefJ and HQ judgment within the different conditions, a reliability analysis was conducted. In the reliability analysis RefJ was thus used as a reference for HQ judgment.

Results

Differences between HQI and HQS scales

Results showed that there were no significant differences between the HQI and HQS scale whatsoever. Only a few small contrasts were apparent between HQI and HQS. Although insignificant, a slightly different in RefJ in both scales can be observed ($\chi^2 = 1.365$, df = 3, p = .714). In the 33 ms condition the effect of RefJ on judgment was lower in the HQI scale than in the HQS scale. In both the 33 ms and 500 ms condition the effect of RefJ on judgment was higher in the HQI scale than in the HQS scale (Figure 4). Relatively this means that in the 17 ms condition judgments on the HQS scale are somewhat closer to their true score than judgments made on the HQI scale. In the 33 ms and 500 ms condition the judgments made on the HQI scale are relatively closer to the true score than judgments made on the HQS scale.

Although small, another difference between both scales appeared in the effect of PT on judgment. It seemed that the effect of PT on judgment seems slightly stronger in the HQI than in the HQS scale during all time-conditions (Figure 4). This scale difference in the Condition x PT relationship however is too small to be considered significant ($\chi^2 = 2.419$, df = 3, p = .490).

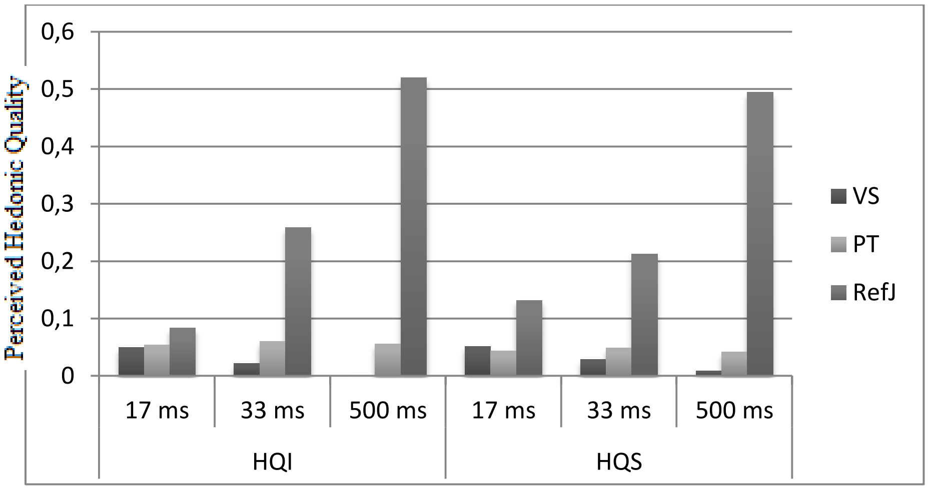

Fig. 4; Colom containing averaged judgment beta values per component, per conditie, per schaal.

Influences of VS, PT and RefJ in different time-conditions

Since results on the HQI scale and the HQS scale showed no significant differences, it was decided to combine both scales into one (Figure 5). When evaluating the effect of VS on HQ judgment a Condition x VS interaction effect appeared ($\chi^2 = 55.730$, df = 2, p < .001).

Table 1; Beta values of VS per condition.

	Beta	**Lower bound**	**Upper bound**	**Chi-square**	**Sig.**
17 ms	0.050	0.038	.063	60.846	.000
33 ms	0.025	0.013	.037	16.816	.000
500 ms	0.004	-0.008	.016	.487	.485

It is apparent that the influence of VS on the HQ judgments declines as presentation-times increase. VS directs judgment to the highest degree in the 17 ms condition ($\chi^2 = 60.846$, df = 1, p < .001). In the 33 ms condition VS directs judgment to a lesser degree ($\chi^2 = 16{,}816$, df = 1, p = .000). In the 500 ms condition the effect of VS on HQ judgment is no longer significant ($\chi^2 = .485$, df = 1, p = .485). These results are summarized in Table 1.

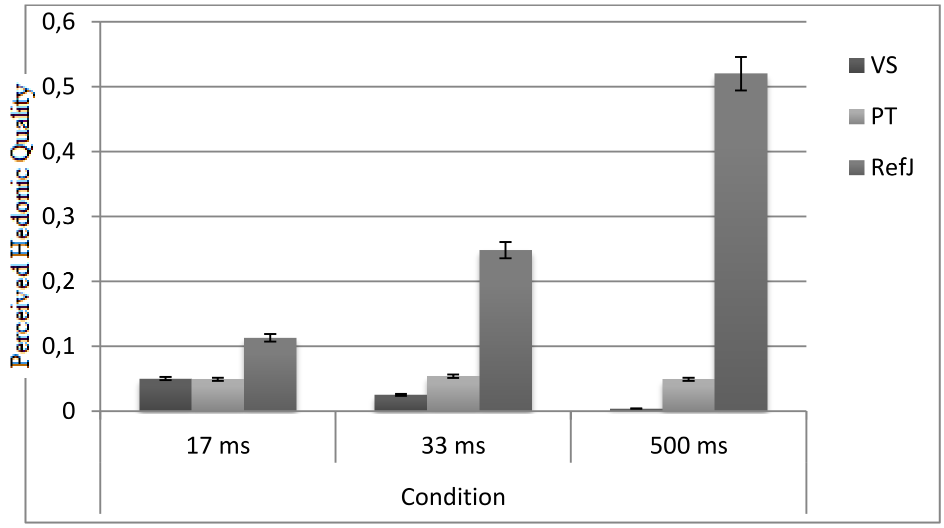

Fig. 5; Colom displaying averaged judgment beta values per component, per condition.

When examining PT in the different time-conditions, some surprising results arise. Firstly, a PT main effect was found, indicating that PT had a significant effect on HQ judgments ($\chi^2 = 182.390$, df = 1, $p < .001$). However, there was no Condition x PT interaction found ($\chi^2 = 3.174$, df = 2, $p < .205$). The effect of PT was the same in all the conditions, meaning that PT played an equally important part in the 17 ms, 33 ms and 500 ms conditions within our experiment. A VS x PT interaction-effect occurred within the data ($\chi^2 = 10{,}210$, df = 1, $p = .001$).When further examining these results it seemed that PT had a higher effect on HQ judgments when VS was high. The effect of PT on HQ judgments was lower when VS was low. To display these results, VS and PT were grouped in VS group and PT group and depicted in Figure 6.

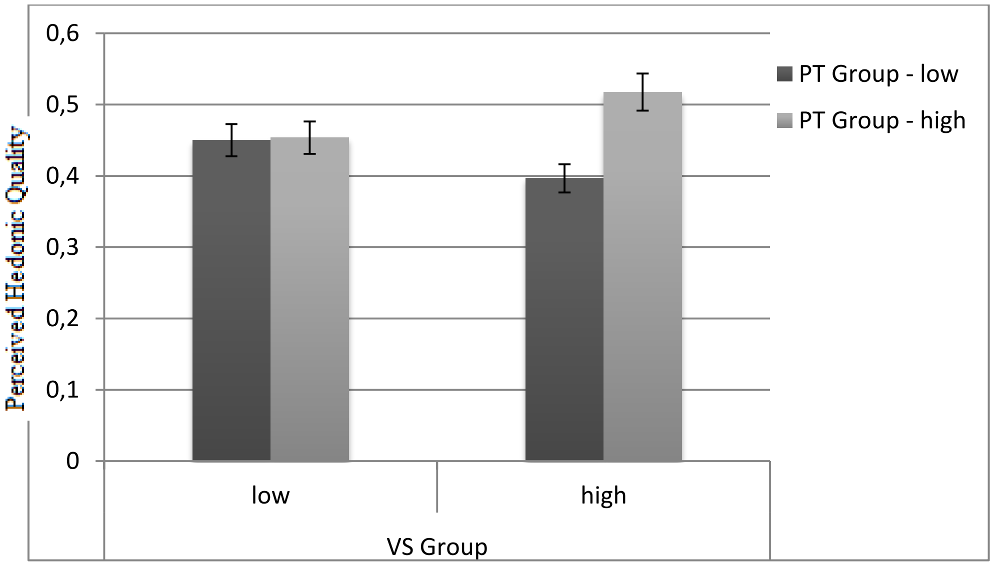

Fig. 6; Colom with averaged HQ judgments for high and low website VS and PT.

Although weaker, there was also a Condition x VS x PT interaction-effect ($\chi^2 = 9.440$, df = 2, p = .009), meaning that the interaction between VS and PT varied slightly in the different time-conditions. When accessing this variation it seems that in the 17 ms condition the combined effect of VS and PT is fairly weak. PT seems to influence HQ judgment both when VS is low and high. In the 33 ms condition the VS x PT interaction effect is very apparent, showing that PT has far less influence on HQ judgment when VS is high. In the 500 ms condition it seems that the overall effect of PT increases, making the effect of PT on judgment a bit more pronounced both when VS is low and when VS is high (Figure 7).

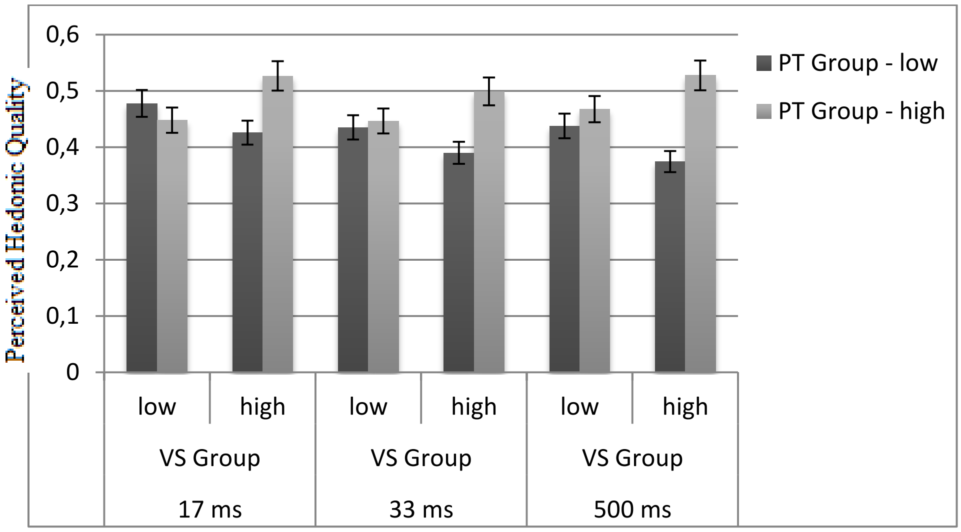

Fig. 7; Colom with average HQ judgments per condition for high and low website VS and PT.

When analyzing the association between of RefJ and judgment, a Condition x RefJ interaction was detected ($\chi^2 = 400.979$, df = 3, p < .001).

Table 2: Reliability of RefJ and HQ judgment per condition.

	Cronbach's Alpha	**Cronbach's Alpha based on standardized items**	**Number of Items**
17 ms	.246	.246	2
33 ms	.460	.462	2
500 ms	.703	.703	2

RefJ associated more with judgment increased as presentation-times got higher (Table 2). In the 17 ms condition the influence of RefJ on HQ judgment was at its lowest ($\alpha = .246$). In the

33 ms condition the influence of RefJ on HQ judgment grew ($\alpha = .460$) reaching its peak at 500 ms ($\alpha = .703$). Since RefJ scores were closest to ‘true judgment scores’ it appeared that judgments got closer to their ‘true judgments’ as presentation-times increased.

Discussion

Results of this study clearly point out that visual complexity (VC) and prototypicality (PT) are important factors in the Hedonic Quality (HQ) perception of websites. When forming a first impression, websites with lower visual complexity are judged more favorably than websites with high visual complexity. The effect of visual complexity diminishes if participants are confronted with the websites for longer periods of time. In general, prototypical websites are judged more positively than less prototypical websites. However a VC x PT interaction is also present within the data. When VC is high the effect of PT is more pronounced than when VC is low. Thus it seems that the combination of high PT and low VC amounts to the highest HQ judgments. These results are the same results reached by Tuch et al. (2012).

Information-processing stage model of aesthetic processing

This study offers some support for the model by Leder et al (2004), primarily due to findings concerning website VC and RefJ. This study showed that the effect of VC on judgment was highest when presentation time was lower. According to Leder et al. (2004) judgment is made in progressively higher stages as presentation-times increase. In higher stages judgment is directed by other factors, diminishing the effects present in lower stages. When viewing the Leder et al. (2004) model, it would be explain why the effect of VC on judgment to decrease as presentation-times increase. RefJ also offers support for the model of Leder et al. (2004). In this study it seemed that, as presentation-times increase the influence of RefJ on HQ judgment seems to also increase. HQ judgments made in higher presentation-times thus seem to get closer to the HQ reference judgment made at 5000ms. Since scores gradually seem to get closer to the RefJ it appears that as perception moves to higher stages, HQ judgments are gradually being influenced by the similar factors as the RefJ. Supporting the model by Leder et al. (2004), stating that judgment is influenced by different variables as presentation-times increase.

The only finding in this study that lies in contrast with the model of Leder et al. (2004) is the influence of PT on HQ judgment. Not only did this study find PT to effect HQ judgments in websites presented for only 17 ms, this study also found website PT to be an equally accurate predictor for HQ judgment within all time-conditions. The early occurrence of the PT effect could have been caused by the TFT-screens the experiments were conducted on. Since the 17 ms condition is very short for frames to be presented on the screen, it is possible that websites in this condition were actually presented longer than 17 ms. Still the influence of using TFT-screens in the conduction of this experiment might not be high, as Tuch et al. (2012) also conducted their experiment using TFT-screens and found no PT influence on judgment in the 17 ms condition. Either way, it is impossible for this study to distinguish whether the VC effect occurred before the PT effect as both effects were apparent in the shortest time-condition. For following research it would thus be necessary to use CRT screens and also use a time-condition shorter than 17 ms, possible clarifying when VC and/or PT starts to influence HQ judgments.

With regards to the relationship between VC and PT this study also found some interesting results. It seemed that the VC x PT interaction changed in the different time-conditions. In the 17 ms condition the influence of PT on HQ judgment hardly differed when VC was high or low. However in the 33 ms condition a clear distinction arose, indication that PT influenced judgments only when VC was high. In the 500 ms condition PT affected HQ judgments more in both the low and high VC condition (Figure 7).

It seems as if the VC x PT interaction gains strength as the influence of VC on HQ judgment decreases. It would thus seem as if the effect of VC does not decline as presentation times, it is just combined into a broader VC x PT interaction. This VC x PT interaction is weaker in the 17 ms condition probably because participants were not able to combine VC and PT information this fast. In the 33 ms and 500 ms condition the participants have adequate time to view VC and PT as a whole, creating a clear VC x PT interaction.

(uitleggen hoe dat mogelijk is)

HQI and HQS scales

Results indicate that there are few differences between the HQI and HQS scale. The only apparent differences were caused by PT and by RefJ. It seemed that the influence of PT on HQI scores was higher than on HQS scores. It thus seems that PT has a bit more influence in how participants identify themselves in a website, than on perceived website creativity or innovation. This difference was a bit surprising, as the HQS scale seemed more compatible with PT.

This can possibly be caused by the fact that the HQS consists has a wider notion of the definition of creativity and innovation than PT does.

The effect of RefJ on HQI scores and HQS scores differed somewhat in the different conditions. In the 17 ms condition the effect of RefJ on HQI scores was lower than on HQS scores. Since personally identifying oneself with a website is a bit more demanding task than viewing website innovation. It is thus very likely that participants in the 17 ms condition were only able to rate items on the HQI scale using more superficial means. Since it was a bit easier to rate items on the HQS scale, there was more room for other variables to take over judgment. In the 33 ms and 500 ms conditions the influence of RefJ on HQI scores was higher than on HQS scores. It thus seems that when given the time, HQI scores gets closer to 'true judgment' faster than HQS scores. Since website HQI scores are processed faster, it would thus seem that users slightly prioritize personal identification with a website over how stimulation the website is found. This shows that in website perception the need to personal identification exceeds the need for stimulation.

Perceptive processing compared to inference perspective

A fair amount of the results in this study are comparable to findings made by Tuch et al. (2012), showing that visual complexity and prototypicality also play an important role in the process of hedonic quality judgments. This study showed that websites with low visual complexity and high prototypicality were more preferred by users. Prototypicality seems to influence website judgment more when visual complexity within the website is low. Both visual complexity and prototypicality seem to influence hedonic quality judgment even when websites are only presented for 17 ms. This research also offered some support to the information integration stage model developed by Leder et al. (2004), showing that hedonic quality judgment is indeed processed in stages.

This study disputes the occurrence of a beauty inference effect suggested by Hassenzahl & Monk (2010). Firstly this study shows that hedonic quality judgment occurs in stages based on presentation-times. With varying presentation-times hedonic quality scores can be explained with use visual complexity and prototypicality, as suggested by the model of Leder et al. (2004). Secondly this study shows that visual complexity and prototypicality influence hedonic quality in a different fashion than beauty. When considering the inference perspective it is expected that the shortest condition would be most prone to a beauty inference, since information in that condition is least easily obtained. This study however shows that in the 17 ms condition unlike with beauty judgment, hedonic quality judgment is guided partly by prototypicality. If inference would have in fact taken place, the prototypicality effect in the 17 ms condition should be non-existent in hedonic quality judgment just as in beauty judgment. The discrepancy between hedonic quality judgment and beauty judgment shows that it is unlikely a beauty inference has taken place in this instance.

In sum, this study shows both visual complexity and prototypicality are important web design characteristics. Complex designs or designs not typically expected of a websites could make bad first impressions, causing decreased user expectations. Research has shown that negative expectations can lead to lower satisfaction (Raita & Oulasvirta, 2011). Considering visual complexity and prototypicality when designing a website could thus prevent user dissatisfaction.

Future research

There is still a lot about Hedonic Quality perception which remains unclear, leaving room for further research. Since not all stages of the model by Leder et al. (2004) were used, further research is necessary in the other stages to explore other factors affecting hedonic quality judgment. If for example timelines depicted in Figure 3, are all implemented in a research it is possible to see how hedonic quality judgment changes over these stages.

A combined research for beauty processing and hedonic quality processing should be conducted. This research should research the similarities between beauty and hedonic quality processing and the speed at which beauty judgment and hedonic quality judgment get closer to their true judgment. If beauty and hedonic quality processing differ in many respects the possibility of an inference effect becomes even more unlikely.

Other UX scales need to be measured using the model by Leder et al. (2004). The model by Leder et al. (2004) would be strengthened if stages also appear in other UX scales. If processing occurs in a different fashion within UX scales, the existence of an inference effect would become even more unlikely.

Acknowledgements

I would like to thank Alexandre N. Tuch for allowing the use of stimuli previously used in his own research.

I would also like to thank Marc Hassenzahl for use of his inference model.

References

Bernardo, M., Marimon, F., & Alonso-Almeida, M.D.M. (2012). Functional quality and hedonic quality: A study of the dimensions of e-service quality in online travel agencies. *Information & Management*, *49*(7-8), 342–347. doi:10.1016/j.im.2012.06.005.

Cogan, E., Parker, S., & Zellner, D.A. (2013). Beauty beyond compare: effects of context extremity and categorization on hedonic contrast. Journal of experimental psychology. Human perception and performance, 39(1), 16–22. doi:10.1037/a0031020.

Goldstein, E.B. (2009). Sensation and Perception. Arizona: Wadsworth, ISBN-10: 0495601497, ISBN-13: 9780495601494.

Harper, S., Michailidou, E., & Stevens, R. (2009). Toward a definition of visual complexity as an implicit measure of cognitive load. *ACM Transactions on Applied Perception*, *6*(2), 1–18. doi:10.1145/1498700.1498704.

Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, *19*(4), 319–349. doi:10.1207/s15327051hci1904_2.

Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481–499. doi:10.1207/S15327590IJHC1304_07.

Hassenzahl, M., & Monk, A. (2010). Human – Computer Interaction The Inference of Perceived Usability From Beauty The Inference of Perceived Usability From Beauty, (March 2013), 37–41. doi:10.1080/073700242010500139.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British journal of psychology (London, England : 1953)*, *95*(Pt 4), 489–508. doi:10.1348/0007126042369811.

Leder, H., Carbon, C.C., & Ripsas, A.L. (2006). Entitling art: Influence of title information on understanding and appreciation of paintings. Acta psychologica, 121(2), 176–98. doi:10.1016/j.actpsy.2005.08.005

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology, 25*(2), 115–126. doi:10.1080/01449290500330448.

Mynatt, C.R., Doherty, M.E., Tweney, R.D., Mynatt, C.R., Doherty, M.E., & Ryan, D. (1977). Quarterly Journal of Experimental Confirmation bias in a simulated research environment : An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology 29*, 85-95.

Nose, T., Ikeda, N., & Kanoh, H. (1999). LCD Legibility as a Function of Resolution, *IEICE translations E82-C* (10), 1792–1797.

Raita, E. & Oulasvirta, A. (2011). Too good to be bad: Favorable product expectations boost subjective usability ratings. Interacting with Computers, 23(4), 363–371. doi:10.1016/j.intcom.2011.04.002.

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 : an official journal of the Society for Personality and Social Psychology, Inc, 8(4), 364–82. doi:10.1207/s15327957pspr0804_3.

Roth, S. P., Schmutz, P., Pauwels, S. L., Bargas-Avila, J.A., & 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. doi:10.1016/j.intcom.2009.10.004.

van Schaik, P., Hassenzahl, M., & Ling, J. (2012). User-Experience from an Inference Perspective. ACM Transactions on Computer-Human Interaction, 19(2), 1–25. doi:10.1145/2240156.2240159.

Shieh, K.K., & Lin, C.C. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics*, *26*(5), 527–536. doi:10.1016/S0169-8141(00)00025-1.

Seppälä, T., Lipponen, J., & Pirttilä-Backman, A.M. (2012). Leader fairness and employees' trust in coworkers: The moderating role of leader group prototypicality. *Group Dynamics: Theory, Research, and Practice*, *16*(1), 35–49. doi:10.1037/a0026970.

Tractinsky, N., Katz, A.., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*(2), 127–145. doi:10.1016/S0953-5438(00)00031-X.

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. doi:10.1016/j.ijhcs.2006.06.009.

Tuch, A.N., Bargas-Avila, J.A., Opwis, K., & Wilhelm, F.H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, *67*(9), 703–715. doi:10.1016/j.ijhcs.2009.04.002.

Tuch, A.N., Presslaber, E.E., Stöcklin, M., Opwis, K., & Bargas-Avila, J.A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies*, *70*(11), 794–811. doi:10.1016/j.ijhcs.2012.06.003.

Appendix

Output of the Regression analysis:

Model Information

Dependent Variable		Judgement
Probability Distribution		Normal
Link Function		Identity
	1	Subj
Subject Effect	2	trial
	3	AnchorLow
Working Correlation Matrix Structure		Exchangeable

Case Processing Summary

	N	Percent
Included	4200	100,0%
Excluded	0	0,0%
Total	4200	100,0%

Correlated Data Summary

		Subj	25
Number of Levels	Subject Effect	trial	168
		AnchorLow	14
Number of Subjects			4200
Number of Measurements per Subject	Minimum		1
	Maximum		1
Correlation Matrix Dimension			1

Categorical Variable Information

			N	Percent
Factor	Scale	HQI	2100	50,0%
		HQS	2100	50,0%
		Total	4200	100,0%
	Condition	0.017s	1400	33,3%
		0.033s	1400	33,3%
		0.5s	1400	33,3%

	Total	4200	100,0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Judgement	4200	,00	1,00	,4720	,26983
	VS	4200	2,32	5,88	4,1230	,88723
Covariate	PT	4200	2,29	5,55	4,3855	,74546
	referenceJudgement	4200	,00	1,00	,4691	,28500

Goodness of Fit[a]

	Value
Quasi Likelihood under Independence Model Criterion (QIC)[b]	304,878
Corrected Quasi Likelihood under Independence Model Criterion (QICC)[b]	298,717

Dependent Variable: Judgement
Model: Scale * Condition * VS, Scale * Condition * PT, Scale * Condition * referenceJudgement, Condition * VS, Condition * PT, Condition * referenceJudgement, Condition * VS * PT, VS * PT, VS, PT, referenceJudgement
a. Information criteria are in small-is-better form.
b. Computed using the full log quasi-likelihood function.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
Scale * Condition * VS	,652	3	,885
Scale * Condition * PT	2,419	3	,490
Scale * Condition * referenceJudgement	1,365	3	,714

Condition * VS	32,848	2	,000
Condition * PT	3,174	2	,205
Condition * referenceJudge-ment	93,521	2	,000
Condition * VS * PT	9,440	2	,009
VS * PT	10,210	1	,001
VS	55,730	1	,000
PT	182,390	1	,000
referenceJudgement	273,095	1	,000

Dependent Variable: Judgement

Model: Scale * Condition * VS, Scale * Condition * PT, Scale * Condition * referenceJudgement, Condition * VS, Condition * PT, Condition * referenceJudgement, Condition * VS * PT, VS * PT, VS, PT, referenceJudgement

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test	
			Lower	Upper	Wald Chi-Square	df
[Scale=2] * [Condition=1] * VS	,077	,0110	,055	,099	48,877	1
[Scale=2] * [Condition=2] * VS	,035	,0110	,014	,057	10,455	1
[Scale=2] * [Condition=3] * VS	-,001	,0110	-,023	,020	,010	1
[Scale=3] * [Condition=1] * VS	,079	,0124	,054	,103	40,436	1
[Scale=3] * [Condition=2] * VS	,042	,0108	,021	,063	14,884	1
[Scale=3] * [Condition=3] * VS	,006	,0106	-,014	,027	,359	1
[Scale=2] * [Condition=1] * PT	,070	,0100	,050	,090	48,654	1
[Scale=2] * [Condition=2] * PT	,068	,0097	,049	,087	48,596	1
[Scale=2] * [Condition=3] * PT	,055	,0100	,035	,075	29,997	1
[Scale=3] * [Condition=1] * PT	,061	,0104	,040	,081	34,101	1
[Scale=3] * [Condition=2] * PT	,057	,0097	,038	,076	35,120	1

[Scale=3] * [Condition=3] * PT	,041	,0094	,023	,059	19,142	1
[Scale=2] * [Condition=1] * referenceJudgement	,088	,0441	,002	,175	4,003	1
[Scale=2] * [Condition=2] * referenceJudgement	,261	,0424	,178	,344	38,069	1
[Scale=2] * [Condition=3] * referenceJudgement	,520	,0410	,440	,600	161,176	1
[Scale=3] * [Condition=1] * referenceJudgement	,135	,0445	,048	,222	9,184	1
[Scale=3] * [Condition=2] * referenceJudgement	,214	,0422	,132	,297	25,844	1
[Scale=3] * [Condition=3] * referenceJudgement	,495	,0393	,418	,572	158,213	1
[Condition=1] * VS	0[a]	.	.	.	.	.
[Condition=2] * VS	0[a]	.	.	.	.	.
[Condition=3] * VS	0[a]	.	.	.	.	.
[Condition=1] * PT	0[a]	.	.	.	.	.
[Condition=2] * PT	0[a]	.	.	.	.	.
[Condition=3] * PT	0[a]	.	.	.	.	.
[Condition=1] * referenceJudgement	0[a]	.	.	.	.	.
[Condition=2] * referenceJudgement	0[a]	.	.	.	.	.
[Condition=3] * referenceJudgement	0[a]	.	.	.	.	.
[Condition=1] * VS * PT	-,010	,0027	-,015	-,005	14,060	1
[Condition=2] * VS * PT	-,005	,0025	-,010	,000	3,687	1
[Condition=3] * VS * PT	,001	,0024	-,004	,006	,143	1
VS * PT	0[a]	.	.	.	.	.
VS	0[a]	.	.	.	.	.
PT	0[a]	.	.	.	.	.
referenceJudgement	0[a]	.	.	.	.	.
(Scale)	,061					

Dependent Variable: Judgement

Model: Scale * Condition * VS, Scale * Condition * PT, Scale * Condition * referenceJudgement, Condition * VS, Condition * PT, Co referenceJudgement, Condition * VS * PT, VS * PT, VS, PT, referenceJudgement

a. Set to zero because this parameter is redundant.

Output of the Reliability Analysis on RefJ and judgment:

Condition = 0.017s

Case Processing Summary[a]

		N	%
Cases	Valid	1400	100,0
	Excluded[b]	0	,0
	Total	1400	100,0

a. Condition = 0.017s

b. Listwise deletion based on all variables in the procedure.

Reliability Statistics[a]

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,246	,246	2

a. Condition = 0.017s

Inter-Item Correlation Matrix[a]

	Judgement	referenceJudgement
Judgement	1,000	,140
referenceJudgement	,140	1,000

a. Condition = 0.017s

Inter-Item Covariance Matrix[a]

	Judgement	referenceJudgement
Judgement	,069	,011
referenceJudgement	,011	,081

a. Condition = 0.017s

Summary Item Statistics[a]

a. Condition = 0.017s

Condition = 0.033s

Case Processing Summary[a]

		N	%
Cases	Valid	1400	100,0
	Excluded[b]	0	,0
	Total	1400	100,0

a. Condition = 0.033s

b. Listwise deletion based on all variables in the procedure.

Reliability Statistics[a]

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,460	,462	2

a. Condition = 0.033s

Inter-Item Correlation Matrix[a]

	Judgement	referenceJudgement
Judgement	1,000	,300
referenceJudgement	,300	1,000

a. Condition = 0.033s

Inter-Item Covariance Matrix[a]

	Judgement	referenceJudgement
Judgement	,067	,022
referenceJudgement	,022	,081

a. Condition = 0.033s

Summary Item Statistics[a]

a. Condition = 0.033s

Condition = 0.5s

Case Processing Summary[a]

		N	%
Cases	Valid	1400	100,0
	Excluded[b]	0	,0
	Total	1400	100,0

a. Condition = 0.5s

b. Listwise deletion based on all variables in the procedure.

Reliability Statistics[a]

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,703	,703	2

a. Condition = 0.5s

Inter-Item Correlation Matrix[a]

	Judgement	referenceJudgement

Judgement	1,000	,542
referenceJudgement	,542	1,000

a. Condition = 0.5s

Inter-Item Covariance Matrix[a]

	Judgement	referenceJudgement
Judgement	,082	,044
referenceJudgement	,044	,081

a. Condition = 0.5s

Summary Item Statistics[a]

a. Condition = 0.5s