WHAT IS **BEAUTIFUL** IS **GOOD** AND **USEFUL**

Master thesis Raymond van Dongelen

For Aplied Communication science University of Twente 2008

What is beautiful is good and useful!

An experimental study of the influence of visual appeal on the expectation a user has of the information quality of a website

University of Twente / Behavioral Sciences/ Communication Studies

University of Twente

Leeuwarden, 2008

Raymond van Dongelen

Committee:

Dr. T.M. van der Geest

Dr. ir. P.W. de Vries

Universiteit Twente
de ondernemende universiteit

Abstract

This study discusses the influence of visual appeal on the expectation a user has of the information quality of the website.

In the experiment 588 students participated. The participants were presented with three different information search scenarios. For each of these scenarios the participants were asked to rate the expected information quality of four websites. Each website was shown 750 ms. For each reaction the reaction time was recorded. After the rating procedure a selection procedure followed. In the selection procedure participants were asked to select four websites from a list of eight websites based on their expectation of the information quality. In both procedures half of the websites had a high visual appeal and half of the websites had a low visual appeal. After one week the rating procedure was repeated but now each website was shown 5 seconds.

The results show that the participants expected the highest quality of information on a website with a high visual appeal. The websites with a high visual appeal received a higher rating, and were selected more often. The effect of visual appeal decreased when websites were shown longer to the participants. But the effect was still there. Furthermore the results show that extreme reactions (extremely negative or extremely positive) were significantly faster than all other reactions.

The study shows that visual appeal is an important shortcut for users to determine the information quality of a website.

Samenvatting

Dit onderzoek bekijkt de invloed van de visuele aantrekkelijkheid van een website op de verwachting die een informatiezoeker heeft van de informatiekwaliteit van een website. Aan het onderzoek deden 588 studenten mee.

In een experimentele setting werd aan de deelnemers drie informatieproblemen voorgelegd. Per probleem moest de deelnemer van vier websites aangeven welke informatiekwaliteit hij van de website verwachtte. Elke website werd 750 ms getoond. Van elke beoordeling werd de reactietijd geregistreerd. Na de beoordelingsprocedure volgde de selectieprocedure. Hierbij moesten de deelnemers uit een lijst van acht websites aangeven van welke vier websites ze de hoogste informatiekwaliteit verwachtten. Bij beide procedures had de helft van de websites een hoge visuele aantrekkelijkheid en de andere helft een lage visuele aantrekkelijkheid. Na een week werd de beoordelingsprocedure nogmaals herhaald maar nu werd elke website 5 seconden getoond.

De deelnemers verwachtten de hoogste informatiekwaliteit van de aantrekkelijke websites. Deze websites kregen gemiddeld een hogere beoordeling in de beoordelingsprocedure. Ook werden de websites vaker geselecteerd in de selectieprocedure. Verder bleek dat de invloed van de visuele aantrekkelijkheid afneemt naarmate een deelnemer een website langer bekijkt. De visuele aantrekkelijkheid blijft echter een belangrijke rol spelen. De resultaten laten verder zien dat extreme reacties (extreem positief of extreem negatief) sneller worden gegeven dan alle andere reacties.

Het onderzoek laat zien dat de visuele aantrekkelijkheid van een website door gebruikers vaak wordt gebruikt als een indicator van de informatiekwaliteit van een website.

1   Foreword

I became interested in the role of visual appeal on information searchers after I observed my students browsing the Internet when searching for tutorials on programming: They quickly browsed the results given by Google, rejecting some websites before I even had the chance to read them. I wanted to know how this process works and I had the feeling that "superficial" factors such as the visual appeal of the websites visited played a role.

Originally, I planed to finish my Master's Thesis in six months. But due health problems of my daughter it became harder and harder to finish the work. I would like to thank Thea and Peter for providing structure when I needed it most. A teacher myself, I learned much from the way they set me back on track.

I am very grateful that my working environment provided me with the opportunity to pursue this degree. I would like to thank Albert Sikkema for this opportunity. Without the time he gave me it would never have been possible to finish the program. And I would like to thank my girlfriend Karen who had to listen to all my good and not-so-good ideas over countless dinners. And off course she also had to sacrifice a lot of the time we would normally have had together.

Lastly, I would like to thank Brigit van Loggem who prevented many spelling and grammar crimes I was about to commit.

2 Introduction

During interaction with the Internet an information searcher has many small decisions to make (Marchionini, 1995). He has to select his information from literally millions of websites. During this search he has to select what websites to read, what information to select, and what information to trust. The sheer amount of information available makes it impossible for the searcher to systematically work trough all the available material. The searcher needs a *shortcut* to determine rapidly what information to select.

Briggs et al conceptualize the judgement of information as two distinct processes (Briggs, Burford, De Angeli, & Lynch, 2002). The first process judges the look and feel of the website, the second process is more cognitively intensive and judges the actual content of the website. If the first process does not convince the user it is worthwhile to visit the website, the second process never starts. Briggs et al. mention that visual appeal of websites is likely to play an important role in the first process. The work of Lindgaard et al. shows that the visual appeal of a website can be rapidly determined by a user (Lindgaard, Fernandes, Dudek, & Brown, 2006). A user can judge the visual appeal of a website within 500 ms and this judgement is relatively stable over time. It seems likely that the visual appeal of a website functions as a shortcut for a user to determine the quality of a website. This current study investigates the effect visual appeal has on the initial impression of the information searcher of the information quality of a website.

The research questions are:

RQ1. To what extent does the visual appeal of a website influence the expected information quality of a website?

RQ2. What is the effect of time on the relation between visual appeal and expected information quality?

3 Theoretical framework

This chapter provides a theoretical framework for this study. Central concepts are visual appeal and expected information quality. In this chapter I first review the literature, in order to arrive at a definition of these concepts. Then, the judgement process is discussed, in particular how judgements are made when a person has little resources available. Next, the relation is discussed between visual appeal and expected information quality, as is the influence of time on this relation. The chapter closes with a section that discusses the approach of the study.

3.1 Definition of visual appeal

The field of Human Computer Interaction (HCI) uses various terms to denote the visual qualities of a user interface; a standard body of terminology is lacking (Norman, 2004b). Multiple terms (visual appeal, aesthetics, beauty, attractiveness) are used in the literature to denote what seems to be the same concept.

The way visual appeal has been studied has been largely dependent on the views on visual appeal. What makes something appealing? I will divide the views on this into three categories (derived from Reber, Schwartz, & Winkielman, 2004).

The first view is the *objectivist* view. This sees visual appeal as a property of an object that will invoke a pleasurable experience in any suitable perceiver. In the field of HCI, this view translates into the study of particular attributes of a product that can make the product more or less beautiful. For examples: proportion, or symmetry of an interface (Hassenzahl, 2007).

This approach can lead to guidelines for graphical designers on how to create attractive products.

A second view is the *subjectivist* view. In this view anything can be beautiful if it pleases the senses of a particular individual. Beauty is a function of the specific characteristics of an individual; all efforts to define these characteristics will be futile (Reber et al., 2004). This approach sees beauty as something rare, a design prize that can be won only by accident (Frolich, 2004).

The last view is the *interactionist* perspective. In this view visual appeal emerges from the interaction between people and objects (Reber et al., 2004). Visual appeal is in "the processing experience of the perceiver" (Reber et al., 2004). Central are the patterns between the objects and the perceivers. This view is similar to what Hassenzahl calls the *judgemental* approach of studying visual appeal (Hassenzahl, 2007). In this approach the process of the perceiver is studied: How fast can they judge visual appeal? How stable are these judgements? What are the consequences of the judgement?

This study adopts the *interactionist* perspective in viewing visual appeal as the appraisal of an object when the user interacts with it. For the purpose of the study visual appeal will be defined as "the judgement of the attractiveness of the visible parts of a website by a visitor of a website". In contrast to other definitions the definition used in this study is aimed only at the attractiveness of the stimuli. Some researchers use broader definitions. The definition of aesthetics by Lavie & Tractinsky for example also mentions factors like "fascinating" and "creative" (Lavie & Tractinisky, 2004). These factors do not necessarily affect the visual appeal. Using a smaller definition is in line with the reasoning of Hassenzahl, who indicates that a broad definition of visual appeal (aesthetics in his document) loses some of its discriminant power (Hassenzahl, 2007).

In this document the term "visual appeal" will be used to discuss results from other studies.

3.1.1   Information quality

To arrive at a definition of expected information quality I will discuss some earlier research. Definitions of information quality roughly follow two patterns. The first focuses on the credibility of the information. Lin and Lu, for example, see information quality as the correctness, credibility and completeness of information (Lin & Lu, 2000). In a review of several papers regarding credibility, Rieh & Danielson found that most researchers view the assessment of credibility as part of the judgement of the information quality of a document (Rieh & Danielson, 2007).

The second group of definitions adds relevancy of the information to the mix. Information found by the user must be both credible and relevant (Rieh & Danielson, 2007). A good example of this approach is the way Liu and Arnett operationalized information quality: They use a comprehensive list of components that add to information quality such as relevancy, accuracy, timeliness, presentation and differentiation (Liu & Arnett, 2000). Relevancy (also called usefulness) plays an important role in what users see as information quality. In a study by Rieh usefulness and goodness of information come forward as the two primary facets of information quality (Rieh, 2002).

This study will look at the effect of visual appeal on information quality. Because the role of visual appeal is theorized to have the most importance during the initial impression of the website (Briggs et al., 2002), this study will show the websites to the participants only briefly. The participants will not be able to make a complete judgement of the information quality. Therefore the participants will be asked to give a prediction about the information quality: "expected information quality".
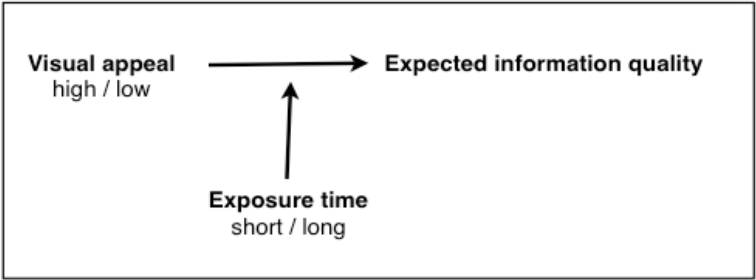
Expected information quality is defined as:

"The prediction information searchers make of the information quality of a website, based on a brief first impression. The searchers are assumed to make predictions about the usefulness of the

information and the goodness of the information. Usefulness is the extent to which the searcher thinks the information will help him to fulfil his information need. Goodness is the quality of the information and the website."

This study will use two different ways to measure expected information quality. Figure 1 shows the relation between the concepts in the *rating procedure*.

Figure 1

Conceptual model for the rating procedure

In this procedure the participants of the experiment *rate* the stimuli for expected information quality. The participants are asked to rate the goodness and the usefulness of the information.

In the *selection procedure* expected information quality is measured in a different manner. In this procedure the participants *select* stimuli for which they expect high information quality. Figure 2 shows the conceptual model for the selection procedure.

Figure 2

The research model for the selection procedure
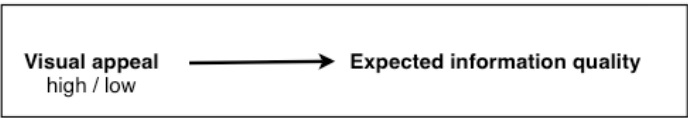
In the selection procedure the participants are assumed to also take factors such as goodness and usefulness into account.
These different approaches are implemented to improve the validity and the generalizability of the results.

If a participant of a study has to determine the usefulness of information he has to have a particular information need. A common approach is to present participants with information search tasks (see

Rieh, 2002; Wirth, Böcking, & von Pape, 2007). The participant is asked to imagine he has a certain information need. Often, tasks are selected that are diverse and easy to relate to. Wirth for example asks participants to search for information on the private life of Albert Einstein. In this study a small but diverse set of tasks is used.

3.2 The judgement process

In this study the participants will be asked to judge whether the information offered is useful and good. It has long been argued that humans in these kinds of situation can be modelled as purely economic, an economic man calculating the costs and benefits of each solution (Coleman, 1986; Mill, 1844). Simon was one of the first to reject this idea of the economic man, noting a complete lack of empirical evidence (Simon, 1955). He argues that it is not possible to make fully reasoned decisions that take into account all possible variables. Decision makers have to make decisions with less-than-optimal information available. And they are under time constraints. The human mind has a limited capacity and therefore humans have to use "approximate methods to handle most tasks" (Simon, 1990). Decision makers tend to make "satisficing" decisions: decisions that carry the risk of producing a less-than-optimal outcome, but that suffice for the purpose.

A common approach in communication science is to distinguish two cognitive systems for making judgements. Kahneman names them System 1 and System 2 (Kahneman, 2003). System 1 is fast, automatic, effortless, associative and often emotionally charged. Because it has a high capacity, it is used for most decisions. System 1 relies on simple judgemental rules, heuristics. Heuristics can be simple rules such as: "the more arguments in favour of the product, the better the product", or, "beautiful people tell the truth". Heuristics function as a cognitive shortcut, a way to make fast and effortless decisions.

System 2 is slower, serial, effortful and deliberately controlled. Processing information in this way involves paying careful attention to the subject. A person has to be motivated and able to use System 2 in a certain situation. When System 2 is occupied with a demanding mental activity, it cannot be used for something else. A person will then just use System 1 for all other decisions.
Decisions made through System 1 are not necessarily of inferior quality. Heuristics can sometimes outperform complex strategies such as multiple regression (Gigerenzer & Todd, 1999). The use of heuristics may however lead to systematic biases (Tversky & Kahneman, 1974).

This study will focus on the judgements made through System 1. To force participants to use System 1, I chose to apply two restrictions. In the *rating procedure* part of the study, time is limited; in the

second part of the study (*the selection procedure*), information is limited. Both restrictions affect the *ability* of the participants to use System 2.

3.3   The relation between visual appeal and expected information quality

The beauty of a person plays an important role in how this person is valued (Dion, Berscheid, & Walster, 1972). Attractive people are judged more positively, are treated more positively, and exhibit more positive behaviours and traits (Langlois et al., 2000).

People are influenced by the attractiveness of nature, architecture and products (Tractinsky, 2004). A good example is the impact of visual appeal on buying behaviour. People are more inclined to buy attractive looking food; "the first taste is almost always with the eye" (Imram, 1999). In the field of HCI, studies have shown effects of visual appeal on perceived usability (Schenkman & Jönsson, 2000; Tractinsky, Katz, & Ikar, 2000), trust (Karvonen, 2000), credibility (Robins & Holmes, 2008) and goodness (Hassenzahl, 2004b).

Tractinsky et al (2000) show that visual appeal influences the perceived usability of a computer interface. In their study users were asked to evaluate the usability of an interface before and after use. The researchers created four types of interface, where the visual appeal and the usability of the interface were manipulated. The experiment revealed an interaction between visual appeal and usability. Appealing interfaces were perceived as usable, regardless of their actual usability. This impact of visual appeal on perceived usability, arguably the crown jewel of the HCI field, led the authors to give the article the provocative title "what is beautiful is usable".

The way usability and visual appeal are defined in the study has received criticism (Hassenzahl, 2004b). The criticism of the definition of visual appeal will be discussed later in this document, as it is a common problem within this field: how can visual appeal be manipulated in an experimental setting? The criticism of the definition of usability is that the participants in the study were likely to have interpreted usability as the goodness of the interface. According to Hassenzahl, the study shows that visual appeal influences the goodness of the system. In accordance with this criticism Hasenzahl (2004b) showed that visual appeal contributes to the goodness of the system *before* use.

Other studies show that visual appeal influences the credibility of a website. In a large survey of over 2,500 participants about website credibility, visual appeal (design look in the study) was the factor that was most often mentioned by the participants (Fogg et al., 2003). The participants were asked to compare pairs of websites as to credibility, and to explain why they thought a website was credible. Robins & Holmes (2008) studied the impact of visual appeal on credibility in an experimental setting. They asked a group of students to evaluate 42 websites. A version with the original graphics (high

aesthetic treatment) was compared to a version of the same website without graphics (low aesthetic treatment). The participants rated the websites with the high aesthetic treatment as more credible.

These studies show that visual appeal has an impact on credibility and goodness. Goodness and credibility are closely related to the concept of expected information quality that is under investigation in this study. It seems likely that visual appeal also influences expected information quality. For the user it is possible that the visual design "acts as a sign of technical refinement that lies underneath the visual layout, on the level of the infrastructure of the system behind the user interface"(Karvonen, 2000).

This results in two hypotheses:

H1. Websites with a high visual appeal are rated higher on expected information quality than comparable websites with a low visual appeal.

H2. When presented with pages with a higher visual appeal and pages with a lower visual appeal, an information searcher is more likely to select pages with a higher visual appeal.

It has been argued that the effect of visual appeal does not have to be positive. Norman argues that sometimes beauty causes users not to trust a product. Conversely, if an object is ugly it must be good, as the effort of the designer was not "wasted" on the appearance of the product (Norman, 2004a). Russo and Demoraes (2003) illustrate the point Norman makes by claiming that some products with a high visual appeal are "*hiding the harm behind the beauty*" (italics added). Previous research in HCI has shown little evidence for what is being called the dark side of beauty (Hassenzahl, 2007). I expect that that in this study visual appeal will *not* affect expected information quality in a negative way.

H3. Websites with a higher visual appeal will *not* be rated lower on expected information quality than pages with a lower visual appeal

3.4 The influence of time

Time is a possible moderator of the relation between visual appeal and other relevant attributes (Tractinsky, 2004). Research by Lindgaard has shown that users can judge the visual appeal of a website within 500 ms and that this judgement is relatively stable over time (Lindgaard et al., 2006). The judgements made by the participants were roughly the same in the 50 ms, 500 ms and unlimited time condition. Tractinsky et al. replicated and confirmed those findings for the 500 ms condition using a different method (Tractinsky, Cokhavi, Kirschenbaum, & Sharfi, 2006).

Other factors that influence information quality are not likely to play a role when a user is exposed to a website for a short time. Briggs et al. (2002) state that factors other than the "look and feel" of a website play a role in a second evaluation process. This second process is more cognitively intensive and judges the actual content of the website. This evaluation process is only started after the user has positively evaluated the "look and feel" of the website.

There are many other factors that are likely to have an impact on information quality in this second process. Rieh mentions characteristics of the information object such as type, organisation, title, presentation, content, functionality, the source of the information and context (Rieh & Danielson, 2007). However, these factors require more time to interpret, so they will have an impact on information quality when the user has more time available. Evaluating the content, for instance, requires that the user starts reading the text, then starts interpreting graphics, and so on. For the user to interpret these signals a more thoughtful process is needed. In the words of Kahneman: the user needs to use System 2. Time however influences the ability of the user to use this system.

This study will investigate the role of time on the relation between visual appeal and expected information quality. Based on the literature I expect that the influence of visual appeal will decrease when the participants have more time available.
Hypothesis:

> H_4. The influence of visual appeal on expected information quality decreases when a user spends more time on a page.

3.5 Reaction time

Extreme judgements often have a shorter response time than more moderate judgements (Pham, Cohen, Pracejus, & Hughes, 2001; Tractinsky et al., 2006). When a person forms a more moderate opinion, he is likely to take more factors into account. Considering more factors takes more time.

Reactions times have been used as an alternative measure of the preference of a participant in experiments. The length of the reaction time indicates a direction of the judgements of the users. Tractinsky et al (2004) used reaction times to see whether there was a convergence with the ratings participants gave. Creating an alternative measure for the same concept provides evidence for construct validity (Tractinsky, 2004).

In this study, reaction times for both the expected information goodness and the expected information usefulness are measured. I expect that for both variables the more extreme judgements will be made faster.

H5. More extreme judgements of expected information quality will have a shorter reaction time than less extreme judgements.

3.6 Manipulation of visual appeal in experimental settings

Manipulating visual appeal in experimental settings is known to be problematic. When visual appeal is the independent variable, at least two versions of each stimulus must be created: one with low visual appeal and another with high visual appeal. In the paragraphs below three methods are discussed and one is selected for this study.

The first method to manipulate visual appeal is that of *rearranging screen elements*. This method is used in the study by Tractinsky et al (2000) discussed earlier. This study focused on the relation between visual appeal and usability. The researchers created a surrogate interface for an Automated Teller Machine (ATM) and manipulated visual appeal by rearranging objects on the screen. Manipulating the position of elements on the screen however also has an impact on other concepts, including usability, the concept under investigation in the study (Hassenzahl, 2004a). If for example the visual appeal is manipulated by placing screen elements randomly on the screen, the visual appeal is indeed affected; but the manipulation also greatly affects the perceived usability. It is easy to see for a user, without interacting with the system, that the system will not be very efficient to use. Such a method of manipulation is likely to have an impact on both the independent and the dependent variable in the experiment.

Another way to manipulate the visual appeal of an interface is to *leave out graphics*. Robinson & Holmes (Robins & Holmes, 2008) used this method. Of each of the websites in their study, a version was included with a high aesthetics treatment and a low aesthetic treatment. The high aesthetic treatment essentially was the website as the researchers found it on the Internet. To create the low aesthetic version, all graphics were removed. A huge advantage of this method is its simplicity; the procedure could even be automated. A problem with this method however is that the newly introduced high and low aesthetics dimension cannot be compared with the types of aesthetics that other studies use (such as the definition used in this study). Robinson & Holmes appear to assume that a high aesthetic treatment can be compared to a visually appealing product. But following the logic of Karvonen (2000) that beauty lies in simplicity, one might argue that the low aesthetic version is the more appealing version. An unwanted consequence of stripping websites of their graphics is the effect on screen layout. Removing graphics can have a huge impact on the website; screen elements may move and may not be immediately visible for the user anymore. When for instance a website uses

a graphical menu, this menu would disappear in the low aesthetic treatment. This has huge consequences for the functionality the user sees when he evaluates a website. Perceived functionality influences the information quality of a website (Rieh, 2002). So this method of manipulation also influences the dependent variable of the study.

A third method is to use different *skins* for the same interface. A skin is a "graphic file used to change the appearance of an application's user interface" (Hassenzahl, 2004b, p. 325). Skins are available for many applications such as the MSN instant messenger and the weblog tool Wordpress, as well as many mobile phones. Hassenzahl (2004) used skins for a software mp3 player called Sonique, to study the influence of visual appeal on the perception of goodness. The availability of skins of various qualities makes it easy to change the visual appeal of the object under investigation. Because the skins are not specifically designed for the study, no biases can be introduced. Another advantage is that the skins are designed to be a unity, rather than a research artifact. A disadvantage is that skins are not available for all information applications that need to be studied. Skins are always directed at existing software applications. A website has to support the use of skins, and not many websites do. When available, the use of skins can be a very effective way to manipulate visual appeal.

This brief discussion of the three methods shows that it is very hard to manipulate the visual appeal of a software product without affecting other dimensions. Manipulations have to be performed very carefully and must be pre-tested. Crude methods such as removing the graphics will have too many consequences for other dimensions. This study takes the skins approach. Because of the lack of readily-available skins, skins had to be designed for the study. Independent designers created new versions of the websites used in the study. The designers were given a list of guidelines that will make sure that other dimensions than visual appeal are affected as little as possible. Manipulations of visual appeal were validated in a pre-study. In the pre-test all the stimuli were tested on visual appeal in a design similar to that of Lindgaard (Lindgaard et al., 2006). A description of the pre-test can be found in appendix A.

3.7 Approach

To conclude this chapter I will outline the design of the study. This study examines the effect of visual appeal on expected information quality in two ways. The effect of visual appeal on the *rating* of information quality, and the effect of visual appeal on *selection* patterns were examined.
Because the participants were faced with either limited time or limited information, they were unable to thoroughly investigate the websites. The rating of websites was executed with two exposure times for the websites: short and long. A positive effect of visual appeal on the expectation of information

quality was expected. Furthermore, longer exposure times were expected to decrease the effect of visual appeal.

The set of websites were designed in a pre-study. For each website a version with low visual appeal and a version with high visual appeal was created. All manipulations of visual appeal were checked on a large group of participants.

4 Research design

This chapter discusses the design of this study. It starts with a discussion of the selection and creation of the websites that were used, followed by an overview of the study. After that the *rating* procedure and the *selection* procedure are discussed. The chapter end with a description of the two settings in which the study was executed.

4.1 Manipulation check of the stimuli

In the experiment websites with a high and low visual appeal were used. To allow for a comparison of websites on visual appeal, 12 pairs of websites were used. Each pair consisted of a website with high visual appeal and a version with low visual appeal. The content in both websites was identical. The websites were designed and validated in a pre-study.

For the pre-study, twelve websites were selected from the Internet. Advertisements and information regarding the website owners were removed. Two graphical designers carried out the manipulations of the websites. They created a new version of each website so as to have a different visual appeal from the original version. The test leader decided whether the designer created a version with high or low visual appeal. The designers were free to do whatever they thought was necessary to manipulate the visual appeal, within a list of guidelines (see Appendix B) which was created to prevent manipulation of the perceived information quality of the website. The work of the graphical designers was validated using the guidelines. Several redesigns took place before the websites were shown to the participants. 78 participants validated the websites in a lab setting. They were exposed to the websites for 750 ms. A website was approved when the version with low visual appeal received a mean rating that was statistically lower than the mean rating of the version with high visual appeal.

Two validation procedures were needed before the test set was complete. A complete description of the pre-study can be found in appendix B.

Figure 3 shows examples of three validated websites.

Figure 3

Examples of the pairs of websites for each of the information tasks

An overview of all the websites can be found in appendix C.

The pre-test yielded twelve pairs of websites.

4.2   Information search scenarios

The participants of this study judged the expected information quality of the websites. For this judgement to take place it is important that the participant was looking for information on a particular subject. To simulate this information search process, the participants were presented with three information search scenarios. They were asked to imagine having an information problem that can be solved using information from the Internet, such as the need for information on Albert Einstein in preparation of a presentation. The information search tasks cover three different topics, selected for diversity. These topics are:

- Selection of material for a presentation about Albert Einstein (taken from Wirth et al (2007))
- Getting an impression of the holiday destination "Rugen"
- Finding medical information about headaches (inspired by Rieh (2002)).

The introductory texts for each task can be found in appendix D.

For each of the tasks the participant had to perform two actions: he had to rate the information quality of the four websites belonging to the task, and he had to select four websites from a total of eight, based on his expectation of the information quality. The experiment took place in two sessions. Table 1 gives an overview of the sessions.

Table 1

Overview of the sessions

Session 1	Session 2 (one week after session 1)
Questionnaire about Internet experience	
Information scenario Einstein Rating procedure short exposure of 4 websites Selection procedure of 8 websites	Information scenario Einstein Rating procedure long exposure of 4 websites
Information scenario Rugen Rating procedure short exposure of 4 websites Selection procedure of 8 websites	Information scenario Rugen Rating procedure long exposure of 4 websites
Information scenario Headache Rating procedure short exposure of 4 websites Selection procedure of 8 websites	Information scenario Headache Rating procedure long exposure of 4 websites

4.3 Rating procedure

For each of the scenarios the participant saw four websites. The participant had to judge whether the information on the website was useful for his information need and whether the information was "good". The participants rated the expected information goodness and expected information usefulness. For each rating the reaction time was registered. These variables form the expected information quality of the website. Of each website the high visual appeal version was shown to half of the participants and the low visual appeal version was shown to the other half. A participant saw two versions with high visual appeal and two versions with low visual appeal.

To investigate the dimension of time, each participant saw the websites twice, in two separate sessions. In the first session the websites were shown for 750 ms (short exposure). In the second session, a week later, the websites were shown for 5 seconds (long exposure). This intervening time period was chosen to ensure that a participant did not remember his previous judgement. The short exposure times force heuristic processing of the websites. The period of 750 ms was based on that used by Lindgaard et al. (2006), which was 500 ms. The time period of 5 seconds was based on the study by Robins and Holmes (2008), where participants were asked to rate a page on credibility as fast as they could and the slowest rating took 4.5 seconds. Therefore 5 seconds is considered a reasonably "long" period.

4.4 Selection procedure

After rating the four websites associated with an information task, the participants were asked to select four thumbnails of websites from a list of eight thumbnails (resolution: 180 * 135 pixels). Both the lower visual appeal and higher visual appeal versions of the four websites associated with the search task were shown. At a resolution of 180 * 135 pixels the content is clearly visible, but not readable. The selection procedure was carried out only once, during the first session.

4.5 Order

To make sure there were no order effects, all websites were shown in a random order in both the selection procedure and the rating procedure. The order of the information tasks was also randomized.

4.6   Lab setting and Internet setting

To investigate whether this type of study can be executed outside a lab setting, the study was performed in two settings: a lab setting and an Internet setting. In the Internet setting potential participants were approached via e-mail, making it easier to reach a large and varied group.

Studies carried out via the Web have the following advantages (adapted by author from Reips, 2002):

- Better generalizability of the findings to more setting and situations.

- Higher external validity.

- Easier to organize.
  Participants can do the experiment in their own environment when they want to do the experiment.

- High voluntary participation.

- Reduction of experimenter effects.

- Greater openness.

Doing experiments via the Web also presents some drawbacks. The biggest problem is the lack of control. Multiple submissions can be an issue, and there is no feedback on how the experiment is unfolding. The participants might not understand the instructions in the experiment, the experiment may fail during due to technical problems, or the participant may simply drop out (Reips, 2002). For this study a test tool has been developed that was used in both settings and that prevented problems such as multiple submissions. A more detailed description of the tool can be found in appendix E.

In this study the results of the participants in the two settings are compared, to provide recommendations for future research.

5 Results

5.1 Participants

The experiment was held in two sessions of each approximately five minutes. The mean age of the participants was 23.83 (*SD*=7.082). The minimum age was 16 and the maximum age was 61. Most of the participants indicated they used the Internet daily (99%). All of the participants had used the Internet for more than a year (99%), most of them for more than five years (87%). In Table 2 an overview of the participants for the two settings can be found.

Table 2

Overview of the participants for the two exposure times in the two settings

Exposure time	Setting	Male	Female	Total
Short exposure	Lab setting	97	17	114
	Internet setting	150	324	474
	Total	247	341	588
Long exposure	Lab setting	55	12	67
	Internet setting	97	191	288
	Total	152	203	355

Table 2 shows that only part of the participants were exposed twice to the websites. 355 participants took part in both sessions. The other participants took part only in the short exposure condition.

5.1.1 Participants in the lab setting

The participants in the lab setting were all students of the NHL University, enrolled in two different programs. They were asked to participate in the experiment at the beginning of a lecture. The second session was held exactly one week after the first session; all students that participated in the first session were asked to participate again.

The experiment took place in a small classroom equipped with 10 identical computers. At the beginning of each session, the test leader briefly introduced the study. The participants were told this study was aimed at developing a deeper understanding of the "first impression" of websites.

5.1.2 Participants in the Internet setting

The participants in the Internet setting were students from the NHL University enrolled in 49 different programs and from the University of Twente enrolled in two different programs.

The students received an invitation to participate by e-mail. The invitation described the study as a study on the first impression of websites. One week after completing the first session, the participant received a second e-mail, inviting them to participate in the second session of the study. Students who did not enter the second session within two weeks received a reminder.

A total of 14561 invitations were sent. The response rate was 3.4 %. This low percentage was mainly caused by low response at the NHL University. At 12.5%, the response rate at the University of Twente was normal. There are several explanations for this low response rate at the NHL University. Because of the amount of spam e-mail that users at the NHL University receive, many students ignore their e-mails. Another problem was that many e-mail addresses turned out to belong to former students. A final problem was that the test tool required a current version of the Adobe Flash Player, which not all participants had installed on their systems. If potential participants did not have this version they would have to carry out an additional action to take part in the experiment.

5.1.3   Number of judgements made in the rating procedure

The participants rated each screen on expected information goodness and expected information usefulness. These two variables are combined in the variable "expected information quality". Expected information quality is the sum of the two variables. The scale ranges from 1 (low quality) to 13 (high quality).

Table 3 shows the number of judgements for the short exposure and long exposure conditions in the Lab setting and the Internet setting.

Table 3

Number of judgements for each setting and each condition

Setting	Short exposure	Long exposure	Total
Lab setting	1372 (n=114)	808 (n=67)	2180 (n=588)
Internet setting	5684 (n=474)	3452 (n=288)	9136 (n=288)
Total	7056 (n=588)	4260 (n=355)	11316 (n=588)

Table 3 shows that there were far more judgements in the short exposure condition. For the comparison of the short and long exposure conditions an ANOVA test was needed. One of the assumptions of the ANOVA test is not met, by the discrepancy between the number of judgements in the short exposure and the long exposure conditions. To see whether the reactions of the group that did *not* participate in the long exposure condition are comparable to the group that did participate in the long exposure condition, I compared the mean rating of the groups for expected information quality.

The group that only participated in the short exposure condition gave a lower rating for expected information quality. This result was statistically significant: $t(6319)=-.2770$, $p<.01$. Therefore only the results of the participants that took part in both conditions were used.

5.2 Effect of visual appeal on expected information quality

5.2.1 Results of the rating procedure

The first step in the experiment was the rating by the participants of four websites per information task. The main question in this step was whether the visually appealing websites were rated higher on expected information quality. In addition, it was examined whether "time" moderated the results.

Table 4 shows the mean ratings for the different conditions. The table shows that high visual appeal websites were rated higher than websites with a low visual appeal. This difference can be seen in both the short exposure and the long exposure conditions. Remarkable is that the overall judgement of the websites did not differ between the two conditions. In the short exposure a mean rating of 8.008 was given, in the long exposure a rating of 8.007 was given.

Table 4

Mean ratings for expected information quality of the websites with a high and low visual appeal

Appeal	Time	Mean expected information quality
High visual appeal	Short exposure	9.213 (SD=2.346)
	Long exposure	8.930(SD=2.374)
	Total	9.071(SD=2.342)
Low visual appeal	Short exposure	6.817 (SD=2.978)
	Long exposure	7.092(SD=2.724)
	Total	6,955 (SD=2.865)
Total	Short exposure	8.008(SD=2.938)
	Long exposure	8.007(SD=2.705)
	Total	8.008(SD=2.823)

Note: Expected information quality ranges from 1 (low quality) to 13 (high quality)

Figure 4 visualizes the ratings of the websites.

Figure 4
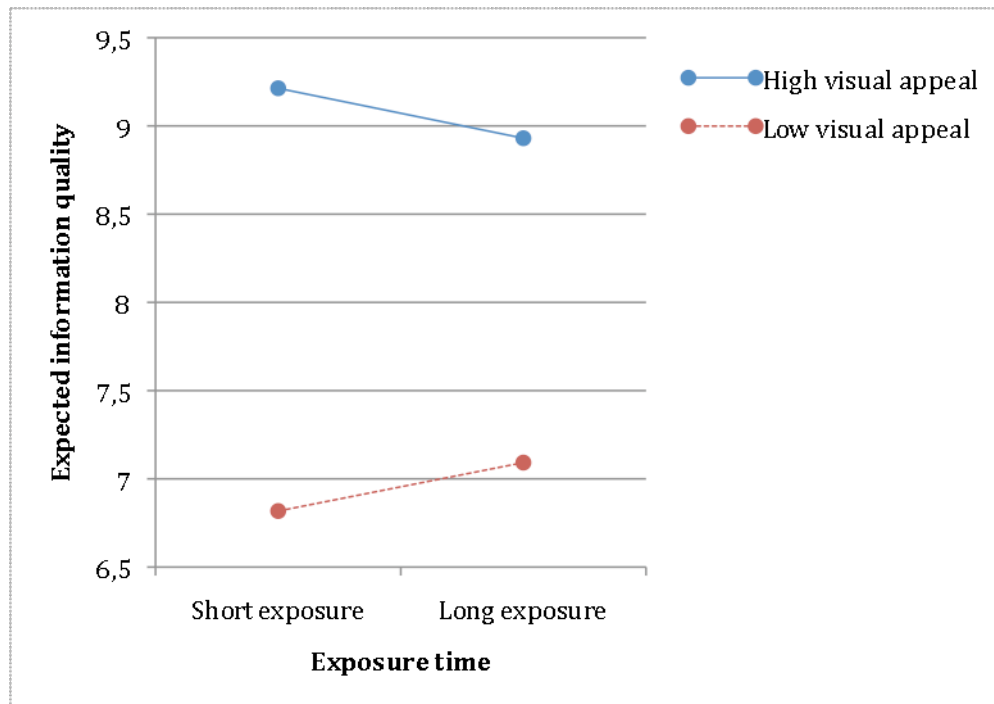
Note: Websites were rated on a scale of 1 (low quality) to 13 (high quality).

Figure 4 shows that the gap between the low and the high visual appeal versions decreases when the participants are exposed longer to the websites. This suggests that the influence of visual appeal on expected information quality decreases when exposure is longer.

To test the hypotheses, the results were further analyzed using a two-way analysis of variance (ANOVA). All judgements with a z-res of 2.5 or higher or -2.5 or lower were removed from the dataset.

The first question that has to be answered is whether visual appeal has an influence on the expectation of information quality. The results shows that this influence was significant, $F_{(1, 8345)} = 1368.74$, $p < .01$. This supports hypothesis H1 that states "Websites with a high visual appeal are rated higher on expected information quality than comparable websites with a low visual appeal."

The second question that has to be answered is how time influences the relation between expected information quality and visual appeal. The effect of visual appeal is influenced by time; and this interaction is statistically significant: $F_{(1, 8345)} = 23.754$, $p < .01$. To investigate this relation further, contrasts were analysed for high and low visual appeal. The contrasts show that websites with a high visual appeal received a lower rating in the long exposure condition: $F_{(1, 57035)} = 12.184$, $p < .01$.

Websites with a low visual appeal received a higher rating in the long exposure condition: $F_{(1, 57035)}=11.573, p<.01$. This means that the gap between the high and low visual appeal versions of the websites decreases when the participants were exposed longer to the websites. This supports hypothesis H_4, stating that "The influence of visual appeal on expected information quality decreases when a user spends more time on a page.". Visual appeal and exposure time explained 14% of the variance of expected information quality.

To see whether the setting had an influence on the relation between visual appeal and expected information quality, the setting was added to the model as a co-variate. Setting had a significant but very small effect, $F_{(1, 11034)} = 24.470, p<.01$. Setting explained .0.2% of the variance in expected information quality.

5.2.2   Dark side of beauty?

To investigate whether this study lends support to the notion of a dark side of beauty, the mean scores of the low and high visual appeal websites were compared. Figure 5 shows the mean ratings for the websites.

Figure 5

Mean expected information quality ratings for each of the websites versions.
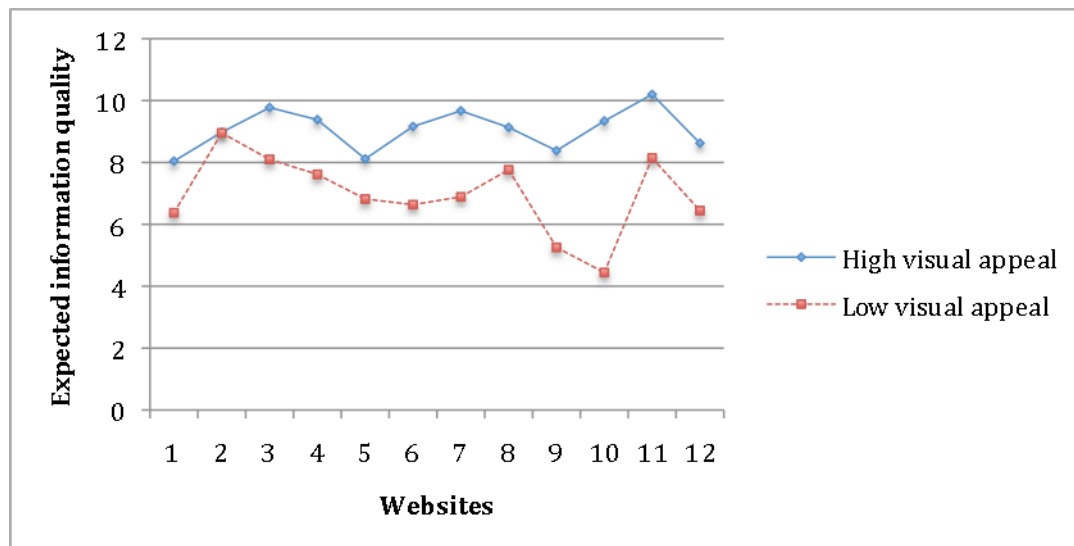
Note: The website numbers used in the figure correspond to the following websites: 1 = Formative years, 2 = Imagination, 3 = Early years, 4 = Einstein, 5 = Einfach, 6 = 100% German, 7 = Isle, 8 = Largest, 9 = Medline, 10 = Illustrated, 11 = Health, 12 = Netdoctor. The websites can be found in Appendix C,
All ratings were made on a scale from 1 (low quality) to 13 (high quality)

Figure 5 shows that for none of the websites the low visual appeal version received a higher mean rating for expected information quality.

This supports H3 that states that none of the websites versions with a higher visual appeal received a lower mean rating than its lower visual appeal counter part.

5.3 Results of the selection round

The selection procedure has the same goals as the rating procedure, that is, to see whether visual appeal influences the expected information quality. The participants were asked to select four websites from a list of eight based on the highest information quality. The main question for this procedure was whether the websites with a high visual appeal would be selected more often than the websites with a low visual appeal.

The participants selected a total of 7104 websites. The results show that websites with a high visual appeal were selected more than websites with a low visual appeal. Websites with a high visual appeal were selected 5051 times, websites with a low visual appeal 2053 times.

The effect of visual appeal on selection was analysed with a one-way analysis of variance (ANOVA). This analysis shows a statistically significant effect of visual appeal on selection: $F_{(1, 14206)}=$ 3078.285, $p<.01$. Visual appeal explained 17.8% of the variance in selection. These results support hypothesis H2 "When presented with pages with a higher visual appeal and pages with a lower visual appeal, an information searcher is more likely to select pages with a higher visual appeal".

The results were further analysed to see whether the experimental setting had an influence on the selection patterns. A t-test shows that the participants in the lab setting selected the websites with a high visual appeal more often than participants in the Internet setting, $t_{(7102)}=-3.337$, $p<.01$. The difference between the two settings however is small. In the Internet setting 70% of the websites selected had a high visual appeal, in the lab setting 75% of the websites had a high visual appeal.

5.4 Response times

5.4.1 Overview

The response times were analysed for the participants who participated in the short exposure condition and in the long exposure condition. The response times for expected information goodness and expected information usefulness will not be combined in the analysis; this is done because the response time is linked to a specific variable. The mean response time for expected information goodness was 2449.27 ms (SD=1107.306); the mean response time for expected information

usefulness was 1264.07 ms (SD=813.517). Participants were significantly faster when answering the question about expected information usefulness; the Wilcoxon signed rank showed a statistically significant difference: T=-78.445, p <.01. This difference in reaction time was probably caused by the order in which the question about expected information goodness and expected information usefulness had to be answered. To reduce the skewness, a logarithmic transformation was performed on the response data. Before the analysis, transformed response times with a z-res higher than 3 were re-coded as missing. For expected information goodness 101 judgements were removed. For expected information usefulness 28 judgements were removed.

5.4.2 Reaction time by response category

Figure 6 shows the mean response times for the response categories of expected information goodness.

Figure 6

Mean response time for each of the seven possible rating for expected information goodness
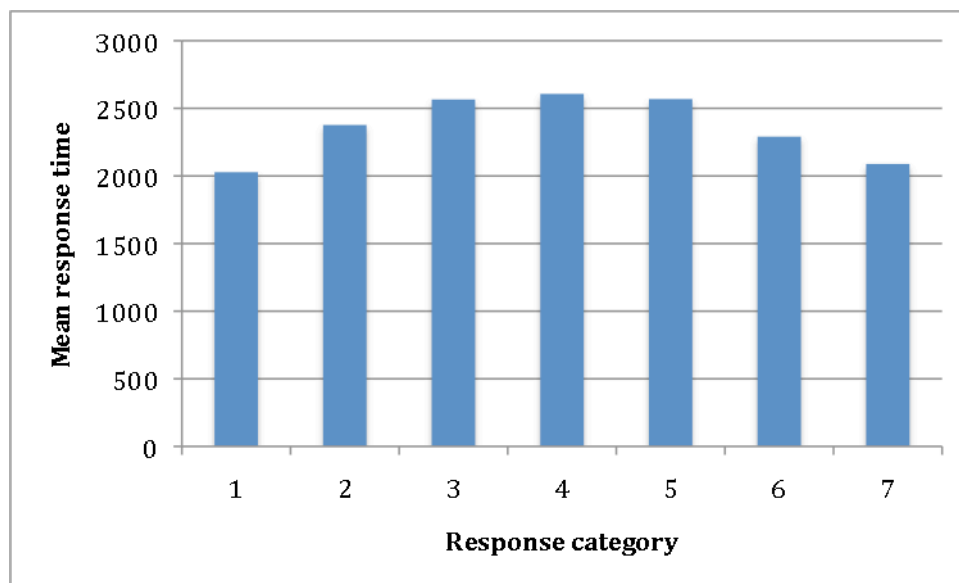
In Figure 6 the most neutral option (=4) has the highest mean reaction time.

The reaction time per answer category seems to follow a normal curve.

Figure 7 shows that mean response time for expected information usefulness follows almost the same pattern:

Figure 7

Mean response time for each seven possible rating for expected information usefulness
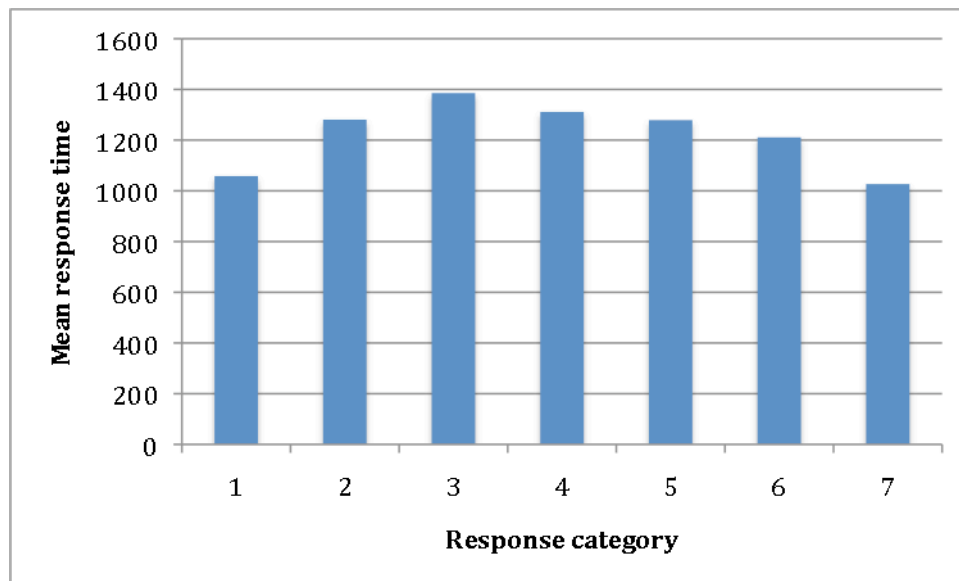
Figure 7 shows that the participants had the slowest reaction to option 3. Option 3 is slightly more negative than neutral.

5.4.3   Extremeness of the judgement and reaction time

Following the procedure used by Tractinsky et al. (2006), the response categories were recoded into extremeness categories. A judgement is seen as extreme when the distance to the middle of the scale is maximized. Options 1 and 7 are the most extreme. Option 4 is the least extreme. The scale of extremeness is divided into 3 categories. The category "high" comprises of options 1 and 7 on the scales of expected information usefulness and expected information goodness. Category "medium" comprises of options 2 and 6 on the scales of expected information usefulness and expected information goodness. The category "Moderate", finally, comprises of options 3, 4 and 5 on the scales of expected information usefulness and expected information goodness.

To reduce the skewness, a logarithmic transformation was performed on the response data. An ANOVA test was performed with the extremeness of the ratings as a random factor and the transformed response times as dependent variables. The results show that for goodness and usefulness a participant takes more time for more moderate judgements. The effect was statistically significant for the extremeness of goodness on the response time: $F_{(2, 11190)} = 62.007, p < .01$. Likewise, the effect was significant for usefulness: $F_{(2, 11095)} = 28.092, p < .01$.

To detect differences between the categories of extremeness a Bonferoni test was carried out. The results for expected information goodness can be found in Table 5

Table 5

Comparing the categories for the extremeness of expected information goodness

	Extreme	Medium	Moderate
Extreme (n=1248)	-		
Medium (n=3222)	-.063*	-	
Moderate (n=6726)	-.116*	-.053*	-

*$p<.01$

The comparison of categories in Table 5 shows that "extreme" judgements of expected information goodness are given faster than "medium" judgements, and that "medium" judgements are given faster than "moderate" judgements.

Table 6 shows the results for expected information usefulness.

Table 6

Comparing the categories of the extremeness of expected information usefulness

	Extreme	Medium	Moderate
Extreme (n=1217)	-		
Medium (n=3076)	-.090*	-	
Moderate (n=6809)	-.125*	-.035*	-

*$p<.01$

Table 6 shows that the difference between each of the categories is statistically significant. The most extreme category had a lower response time than the "medium" category, and the "medium" category had a lower response time than the "moderate" category.

These results lend support to hypothesis H5, which states: "more extreme reactions have a shorter response time".

6 Conclusion & Discussion

6.1 Discussion of the results

Research has shown that visual appeal can influence the way people perceive other attributes. The main objective of this study was to investigate the influence of visual appeal on the expectation users have of the information quality of a website.

This study clearly shows that visual appeal is an important contributor to the expectation of the user. This is an important finding because it helps to shed some light on the process of selecting websites on the Internet. Furthermore it shows that users rapidly form an opinion about the information quality of websites. The results show that users have higher expectations of the information quality when the website has a higher visual appeal. This study shows the effect of visual appeal in three different situations: where a user is exposed shortly to the website, where a user is exposed longer to the website, and where the user sees only a thumbnail of the website.

The speed with which the judgement of the user is coloured by visual appeal, is astounding. Lindgaard et al (2006) already showed that the visual appeal of a website can be judged rapidly. This study shows that such rapid judgement immediately has consequences for the rating of other dimensions. The judgement of the user is affected most when he is exposed shortly to the website. Not only is the effect of visual appeal immediate, the judgement of the user remains coloured by the visual appeal. These findings are in line with the model of the judgement of websites by Briggs et al (2002), who state that the user starts the evaluation of a website by evaluating its look and feel. Briggs assumed that visual appeal is important during this first process (Briggs et al., 2002). This study provides empirical support. Furthermore they state that factors other than visual appeal begin to play a role after the first process. This study also supports this notion: Whereas the perception of visual appeal is relatively stable over time (Lindgaard et al., 2006), the influence of visual appeal on expected information quality decreases over time. It is likely that other attributes of the website begin to play a role. Tractinsky (2004) suggested that time is a moderator of the relation between visual appeal and other attributes. This study provides an empirical foundation for this suggestion. Although time dampens the effect of visual appeal, there is still an important effect.

Following the suggestion of Tractinsky et al (2006), response times were used to check the construct validity of expected information quality. It was hypothesized that the participants would use less time

for expressing more extreme judgements. The results were in line with the expectations, lending support to the strength of the constructs.

The effect of visual appeal was strongest in the selection procedure. In this procedure the participants were asked to select the websites with the highest information quality. Websites with a high visual appeal were selected more often than websites with a low visual appeal. Again, as in the rating procedure, the *ability* of the participants to make a fully reasoned judgement was manipulated. In the rating procedure the participants had limited time, whereas in the selection procedure the participants had limited information. The effect of visual appeal seems to be the strongest when the user is not *able* to perceive other attributes.

This study did not find any support for a dark side of beauty. All of the websites with high visual appeal received a higher mean rating than the other websites. This is in line with the expectations of this study and with other research (Hassenzahl, 2007). This casts some serious doubts on the existence of a dark side of beauty. The amount of HCI research that has found *no* support shows how hard it is to find negative effects for beauty. If a dark side of beauty exists, then it is more likely to be a dark spot than a dark side.

The study was performed in a lab setting and an Internet setting. The results for the two setting were comparable. For this type of study there does not seem to be an obvious best setting. The biggest advantage of the lab setting is that problems can be easily detected. The biggest advantage of the Internet setting in this study was the ability to reach a large and diverse audience. This study has benefited from both settings.

6.2 Contributions

After e-mail, using search engines is the most popular activity on the Internet (Hargittai, 2007). Billions of queries are performed each month by search engines (ComScore, 2007). This means that billions of decisions are made as to whether or not to invest effort in websites. This staggering figure shows how important it is to know more about factors contributing to website selection. This study contributes to the understanding of the role of visual appeal when a user is searching for information on the Internet. It shows that the immediacy of visual appeal also immediately colours the expectation of the information quality, from which it follows that the perception of visual appeal has consequences immediately. This study also clearly shows that time is a moderator of the relation between visual appeal and the expectation of information quality.

6.3   Limitations and suggestions for future research

Response times were used to investigate the construct validity of expected information goodness and expected information usefulness. This was done by categorizing the extremeness of the reactions. The effect of the extremeness categories on response times were then analysed using an ANOVA. However, ANOVA assumes that the groups are of similar size. This is not the case for the extremeness. Because the extremeness categories are only used to illustrate the validity of the constructs I believe that this does not affect the outcome of the study. It is likely that in studies of this type the more moderate judgements will be made with greater frequency than extreme judgements; and different methods of comparison must be used.

The participants of the study rated the usefulness of the websites more rapidly than they rated the expected information goodness. Because the rating of usefulness always had to be done directly after the rating of goodness of the website, the rating of goodness might have influenced the rating of the usefulness. In future research it might be preferable to change the order in which the questions have to be answered.

I have set participants a limited number of tasks. It is unclear as to which degree the results can be generalized to other information scenarios. Rieh suggests that the information task that a person is working on, influences how each website is evaluated (Rieh, 2002). For medical sites, for example, participants might evaluate information more thorough. In a future study the relation between visual appeal and information task needs to be examined. Is the influence of visual appeal on the expectation of information quality the same for all types of task?

It is not clear what a user will do when he expects a low information quality of a website. One might think that users will then immediately leave the website. Research on the relation between attitudes and behaviour has shown that people do not always act as they feel (Fazio & Roskos-Woldsen, 2005). To create a study that better predicts what the users will do after their first impressions, should be directed towards the intention of the participant to stay on a website.

One of the participants in this study commented that he rated some of the websites low because "they did not look the way they ought to". This suggests that participants have a mental image of what the website should look like. If the website does not match his mental image, it is dismissed. Hassenzahl calls this the discounted average strategy (Hassenzahl, 2007). The discounted average strategy might provide an alternative explanation of the results. Therefore, the discounted average strategy should be compared to the "visually appealing is good" heuristic.

Do users choose what is familiar? Or are they more inclined to appreciate websites with high visual appeal?

6.4   Suggestions for the work field

Web designers already know that visual appeal is important. This study provides insight into the reasons as to why this is so. The most important conclusion for the work field is that visual appeal rapidly influences whether users think a website offers the answer to their problem.

Users of the Internet are provided with ever more tools to obtain a visual impression of a website before actually visiting the website. For example, the website "snap.com" offers a view that is not unlike the selection procedure used in this study. Thus, a website may be rejected before it is even visited.

I feel that it is important that web designers test the expectations users have of a website. The expectations of users could be tested by asking potential users to perform "use cases". The test user is presented with a website and asked to indicate whether or not he expects to find the answer to his question there. In this manner, the designer is given feedback on how his design impacts the expectations of the users.

An important consideration is and should be the visual appeal of the website. Because what is beautiful is good *and* useful.

7 References

Briggs, P., Burford, B., De Angeli, A., & Lynch, P. (2002). Trust in online advice. *Social science computer review*, 20(3), 321-332.

Coleman, J. (1986). *Individual Interests and Collective Action*: Cambridge: Cambridge University Press.

ComScore. (2007). comScore Releases August U.S. Search Engine Rankings. Retrieved 3-12-2007, from http://www.comscore.com/press/release.asp?press=1745

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of personality and social psychology*, 24, 285-290.

Fazio, R. H., & Roskos-Woldsen, D. R. (2005). Acting as we feel – When and how attitudes guide behaviour. In T. C. Brock & M. C. Green (Eds.), *Persuasion Psychological insights and perspectives*. New York: Sage.

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). *How do users evaluate the credibility of Web Sites? A study with over 2,500 participants.* Paper presented at the Designing for user experiences.

Frolich, D. M. (2004). Beauty as a design prize. *Human computer interaction*, 19, 359-366.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Hargittai, E. (2007). The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction. *Journal of Computer-Mediated Communication*, 12(3).

Hassenzahl, M. (2004a). Beautiful objects as an extension of the Self. A reply. *Human computer interaction*, 19, 377-386.

Hassenzahl, M. (2004b). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-computer interaction*, 19, 319-349.

Hassenzahl, M. (2007). Aesthetics in interactive products: Correlates and consequences of beauty. In H. Schifferstein & P. Hekkert (Eds.), *Product experience*: Elsevier

Imram, N. (1999). The role of visual cues in consumer perception and acceptance of a food product. *Nutrition & Food science*, 5, 224-228.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics American Economic Review, 93(5), 1449-1475.

Karvonen, K. (2000). *The beauty of simplicity*. Paper presented at the Proceedings on the 2000 conference on Universal Usability.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., HaUam, M., & Smoot, M. (2000). Maxims or Myths of Beauty? *A Meta-Analytic and Theoretical Review Psychological Bulletin*, 126, 390–423.

Lavie, T., & Tractinsky, N. (2004). Assessing Dimensions of Perceived Visual Aesthetics of Web Sites *International Journal of Human–Computer Studies*, 60, 269–298.

Lin, J. C.-C., & Lu, H. (2000). Towards an understanding of the behavioural intention to use a web site *International Journal of Information Management*, 20, 197-208.

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.

Liu, C., & Arnett, K. P. (2000). Exploring the factors associated with Web site success in the context of electronic commerce. *Information & Management*, 23-33

Marchionini, G. N. (1995). *Information seeking in electronic environments*. Cambridge: Cambridge University Press.

Mill, J. S. (1844). On the Definition of Political Economy; and on the Method of Investigation Proper to It. In *Essays on Some Unsettled Questions of Political Economy*.

Norman, D. (2004a). Emotional design. New York: Basic books.

Norman, D. (2004b). Introduction to This Special Section on Beauty, Goodness, and Usability. *Human-Computer Interaction*, 19(4), 311-318.

Pham, M. T., Cohen, J. B., Pracejus, J. W., & Hughes, D. G. (2001). Affect Monitoring and the Primacy of Feelings in *Judgement Journal of consumer research*, 28(167-188).

Reber, R., Schwartz, N., & Winkielman, P. (2004). Processing Fluency and Aesthetic pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review*, 8(4), 364-382.

Reips, U.-D. (2002). Standards for Internet-Based Experimenting *Experimental Psychology*, 49(4), 243-256.

Rieh, S. Y. (2002). Judgement of information quality and cogntive authority in the web. *Journal of the American society for information science and technology*, 53(2), 145-161.

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A Multidisciplinary Framework. In B. Cronin (Ed.), *Annual review of information science and technology* (Vol. 41, pp. 307-464).

Robins, D., & Holmes, J. (2008). Aesthetics and credibility in web site design. *Information processing and management* (1), 386-399

Schenkman, B. N., & Jönsson, F. U. (2000). Aesthetics and preferences of webpages *Behaviour & Information technology*, 19(5), 367-377.

Simon, H. (1955). A behavioral model of rational choice. *Quarterly journal of economics*, 99-118.

Simon, H. (1990). Invariants of human behavior. *Annual review of psychology*, 41, 1-19.

Tractinsky, N. (2004). *Toward the study of Aesthetics in information technology*. Paper presented at the Twenty-fifth international conference on information systems.

Tractinsky, N., Cokhavi, A., Kirschenbaum, M., & Sharfi, T. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. *International journal of human-computer studies*, 64, 1071-1083.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with computers*, 13, 127-145.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and Biases *Science*, 185(4157), 1124-1131.

Wirth, W., Böcking, T., & von Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of computer-mediated communication*, 12(3).

Appendix A: Design of the experimental websites

The goal of this study is to prepare a set of websites that will be used in the main study. A set of twelve websites is required. For each website two versions are needed; a version with high visual appeal and a version with low visual appeal. The difference in visual appeal between the two versions should be statistically significant; on other dimensions, the versions should be equal. This study will provide these websites.

The set will be created by selecting twelve websites from the Internet. Of each of these websites a version will be created by with a higher or a lower visual appeal. Following this design phase, the 24 versions of will be validated on visual appeal by a large number of students.

The manipulation of visual appeal

In this document the versions of a website that are designed to look better than the other version of the website are called higher visual appeal versions. The other version is called the lower visual appeal version. Important to note is that a higher visual appeal version does not have to be beautiful, and a lower visual appeal version does not have to be ugly. It only describes the relationship to the other version of the website.

One of the versions has to look better than the other. All other properties should remain stable. The properties that should remain stable are those properties that influence the "expected information quality". If a graphical designer chooses to improve the visual appeal of a website by restructuring the content of the website, this also changes the actual information quality of the website. Where these versions to be are used it would be unclear what causes the reaction of the user: the difference in visual appeal or the improved structure of the content. Therefore it is very important that the designers are guided as to how the manipulations are made. In this study the designers are given a list of guidelines based on the work of Briggs (2002), Fogg (2002) and Rieh (2002). The work of these authors provides recommendations to web designers on how to improve the information quality of a website. The factors mentioned by these authors are the factors that designers must keep stable. The designers in this study are instructed to not change any factors concerning information quality, and to remove all references that influence the authority of a website.

The factors that must remain stable are:

- The *type* of information on the website
- The *organisation* of the information
- The *presentation* of the information
- The actual content
- The *functionality* of the website

Not only characteristics of the website influence the perceived information quality, information about the source of the information does the same. Information issued on a website by government officials is more likely to be of high quality, and users are more inclined to trust the information. This is what Rieh calls cognitive authority (Rieh, 2002). For this study the authority should have no influence. Therefore all information about the origin of the information will be removed. Information indicating that the issuer of the information is trustworthy (as is implied for example by a certificate) is also removed.

Judgement of visual appeal

In this study the participants have to rate 24 websites (12 websites with both a high visual appeal and a low visual appeal version). It is important that they do not rate the actual quality of the website. If the users were given unlimited time they might rate the website based on the quality of the text, or the usability. This might influence their opinion about the visual appeal of the website.

This study will give the participants a short amount of time to view the website they have to rate. Lindgaard et al. have shown that users can rate a website fast and that this judgement is consistent (Lindgaard et al., 2006).

Method

7.1.1 Overall design

In this study websites are selected and manipulated to create two versions of each website. The two versions of each website must have a significant different rating for visual appeal. The following steps were carried out to accomplish this:

- Twelve websites were selected from the Internet.
- Source information and certificates were removed.
- Of each website, a second version was created with a different visual appeal
- A large group of participants rated all the versions of websites on visual appeal. The websites for which the original version had a significant different rating than the manipulated version were approved. For the other websites, the manipulated version was redesigned.

After two rounds all the websites displayed the required gap in visual appeal. The next sections will discuss the steps in more detail.

7.1.2 Design procedure

The websites were selected for three categories: the first category contains websites about Einstein; the second category websites about the isle of Rugen; and the final category contains websites about the subject Headaches. The topics were chosen for their diversity. Each of the websites was cleaned

up to make sure the website did no longer contain advertisements and information about the website owners. This information was removed to make sure it did not play a role when the participants of later studies are asked to judge information quality. Two graphical designers carried out the manipulations on the websites. The designers were both trained in graphical design and had worked in the field of graphical design. They were paid for their work.

After cleaning the website the designers created a new version of each website that was designed to have a different visual appeal than the original version. The test leader decided whether the designer created a version with high visual appeal or a version with low visual appeal. The designers were free to do whatever they thought was necessary to manipulate the visual appeal, within the scope of a list of guidelines. This list of guidelines was created to prevent manipulations of the perceived information quality of the website. The work of the graphical designers was validated using this list. Several re-designs took place before the websites were presented to the participants.

Websites that were not validated in the validation procedure were re-designed and re-validated on a new group of participants. Two validation cycles were needed.

7.1.3   Validation procedure

The websites were tested on their visual appeal in an experimental setting. For this study, undergraduate students were selected with no background in visual design. The participants could win tickets to the cinema. The procedure was carried out in a classroom equipped with 10 identical computers. On each of the computers a web-based program was running, designed to control the test and ensure that all the participants were in an identical situation. Before the start of the test, the test leader gave a brief verbal explanation. The participant was instructed to give his "first impression" of each page. The software showed explanations of each step in the test. Each participant started with a short questionnaire about his web usage, followed by the rating procedure. Two practice screens familiarized the participant with the user interface and the pace of the rating procedure. After the practice screens, the screens under consideration were shown in random order, to avoid order problems. The reactions were recorded on a 7-point semantic differential scale for visual appeal, ranging from unattractive to attractive. To ensure that actual content did not play a role, the pages were shown for 750 milliseconds. This period is long enough to get an impression, but too short to read any content. Each version of a website was shown at a resolution of 1024 * 768 pixels. The two versions of a website were compared using a students t-test. A level of $p < .05$ is considered to be significant.

7.1.4   Participants

In this study a total of 78 individuals participated (58 male and 20 female). All participants were students at the NHL University, enrolled in two different programs. Students were selected on the

basis of their regularly using the Internet to obtain knowledge about subjects they are interested in.

As such, they can be used to make generalizations about information searchers.

Students were asked to take part in the experiment at the beginning of a lecture. The median age of the participant was 20, minimum 17, maximum 53.

Results

7.1.5   Judgement of visual appeal

On average the high visual appeal versions of the websites had a higher rating than the low visual appeal versions. The mean rating for a high visual appeal version was moderately positive at 4.16 (SD=1.626). The low visual appeal versions on average received a negative rating (M=2.51, SD=2.51). Using the student t-test this difference is significant: $t(2026)=23.972, p<.05$.

In the following paragraphs, the results for each of the websites are discussed. The results are shown grouped by information task. Information tasks are only used as a way to group the results.

7.1.6   Validating the websites about Einstein

The websites that show information about Einstein on average received a low rating (M=2.78, SD=1.600) on the seven-point scale of visual appeal. Only the high visual appeal version of the page "Albert Einstein" received a positive rating of 4.96 (SD=1.391). Figure 8 shows the mean ratings for each of the websites.

Figure 8

Ratings of the visual appeal of the website about Einstein
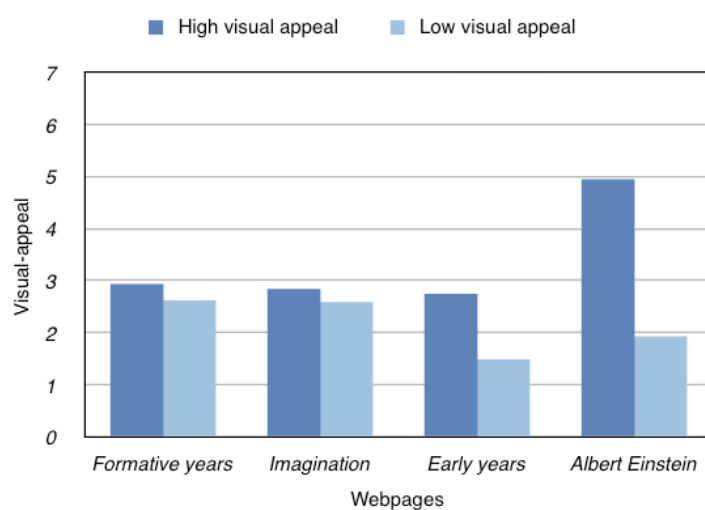
Note: Visual appeal is a semantic differential scale starting at unattractive (=1) and ending at attractive (=7). A total of 78 judgements were made for each website.

Figure 8 shows that the websites "formative years" and "imagination" have a small difference between the ratings of the high visual appeal and the low visual appeal versions of the website. The difference between these versions is not significant (see Table 7).

Table 7 shows that designs for "Early years" and "Albert Einstein" had the demanded gap in visual appeal: The high visual appeal versions were rated significantly higher than the low visual appeal versions.

Table 7

Examining the differences in rating for visual appeal for the websites (n=78)

Website	Version	Mean rating	SD	Visual appeal gap
Formative years	Low visual appeal	2.64	1.195	$t(154)=1.449$
	High visual appeal	2.94	1.342	
Imagination	Low visual appeal	2.60	1.272	$t(154)=1.223$
	High visual appeal	2.86	1.346	
Early years	Low visual appeal	1.50	1.214	$t(154)=6.111$ [*]
	High visual appeal	2.76	1.350	
Albert Einstein	Low visual appeal	1.96	1.253	$t(154)=14.155$ [*]
	High visual appeal	4.96	1.391	

*p<.01

The participants successfully validated two of the four websites. The two other websites had to be re-designed.

7.1.7 Validating the websites about the isle of Rugen

The websites about the isle of Rugen were on average rated neutral ($M=4.04$, $SD=1,760$). For the website "Largest" there were two alternatives for the high visual appeal version. Both alternatives were shown to the participants to be able to select one of the alternatives for the main study. Figure 9 shows the mean rating for all of the websites. The difference between the high visual appeal and the low visual appeal version of "Einfach" is small, but significant (see Table 8).

Figure 9

Ratings of the visual appeal of the websites about the isle of Rugen
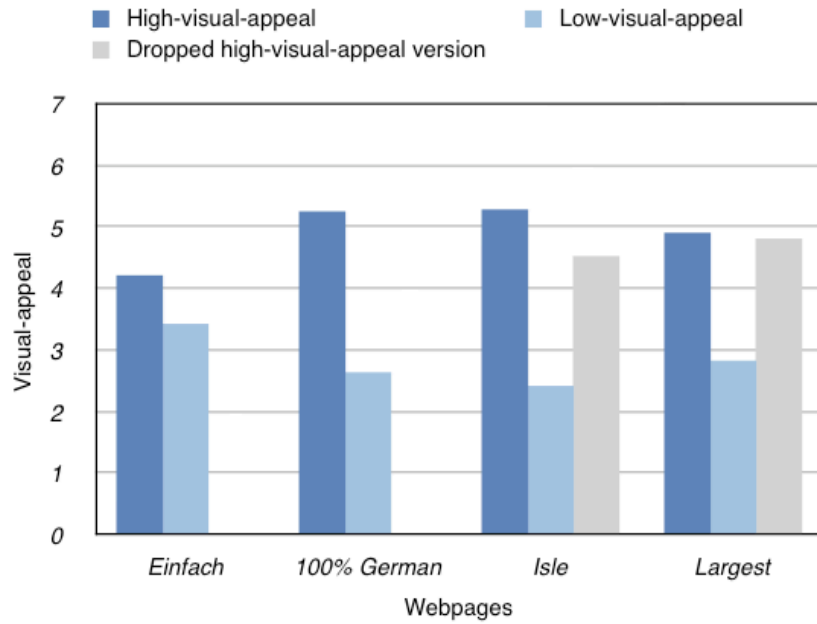
Note: Visual appeal is a semantic differential scale starting at unattractive (=1) and ending at attractive (=7). A total of 78 judgements were made for each website.

All of the websites about the isle of Rugen have the demanded gap in visual appeal (see **Table 8**). For the website "Isle" the high visual appeal version with the highest mean rating was selected. The two high visual appeal versions of "Largest" did not have a significantly different rating ($t(154)= .351, p >.05$). Therefore the version was chosen that was considered the most successful manipulation of the two.

Table 8

Examining the differences in rating for visual appeal for the websites (n=78)

Website	Version	Mean rating	SD	Visual appeal gap
Einfach	Low visual appeal	3.44	1.740	$t(154)=3.014$ *
	High visual appeal	4.22	1.492	
100% German	Low visual appeal	2.64	1.423	$t(146.784)=12.748$ *
	High visual appeal	5.27	1.136	
Isle	Low visual appeal	2.44	1.305	$t(154)=13.450$ *
	High visual appeal	5.31	1.361	
Largest	Low visual appeal	2.85	1.460	$t(154)=8.569$ *
	High visual appeal	4.83	1.436	

*$p<.01$

The websites about the isle of Rugen were successfully manipulated and can be used in further studies.

7.1.8 Validating the "Headache" websites

The "headache" websites were on average not rated as visually appealing (M=3.21, SD=1.597). Two of the high visual appeal versions are rated positively ("Illustrated" and "Health" with means of 4.60 and 4.64, respectively). All the other versions of websites are rated negatively (see Table 9). Figure 10 shows the average ratings for all the websites.

Figure 10

Ratings of the visual appeal of the websites about "Headaches"
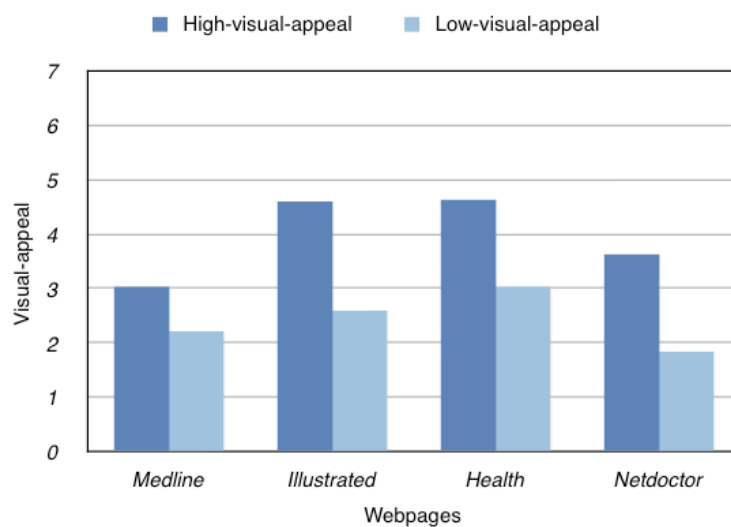
Note: Visual appeal is a semantic differential scale starting at unattractive (=1) and ending at attractive (=7). A total of 78 judgements were made for each website.

Table 9 shows that all of the websites have the required gap in visual appeal. The low visual appeal version of "Netdoctor" receives a very negative rating at 1.86; the low standard deviation shows that there is much agreement between participants regarding this website.

Table 9

Examining the differences in rating for visual appeal for the websites about Headaches (n=78)

Website	Version	Mean rating	SD	Visual appeal gap
Medline	Low visual appeal	2.24	1.175	$t(154)$=3.877 *
	High visual appeal	3.03	1.338	

Illustrated	Low visual appeal	2.60	1.515	$t(154)=8.701$ *
	High visual appeal	4.60	1.352	
Health	Low visual appeal	3.06	1.303	$t(154)=7.911$ *
	High visual appeal	4.64	1.184	
Netdoctor	Low visual appeal	1.86	1.016	$t(144.021)=9.337$*
	High visual appeal	3.63	1.330	

*$p<.01$

Table 9 shows that all of the websites had the required gap in visual appeal. All of the versions can be used in further studies.

Discussion of the first round

The first round of validation sessions resulted in ten of the twelve websites being validated. Two pages about Einstein did not have the desired gap in visual appeal. For these two websites, two new high visual appeal versions were designed. The results of the validation of these two websites are discussed in the next paragraphs.

Validation of the re-designed websites

Method

One of the designers who also worked on the websites in the first round, re-designed the high visual appeal versions of the websites "Formative years" and "Imagination". The results replaced the two rejected versions in the test set. A group of 20 students participated in the second round of this study (12 male, 8 female). None of the students had participated in the first round. The participants were all students at the NHL University, enrolled in two different programs.

The procedure discussed in Design procedure on page 42 was followed.

Results

The participants consider neither of the redesigned high visual appeal versions visually appealing. Figure 11 shows the mean ratings for both websites.

Figure 11

Ratings of the visual appeal of the redesigned websites
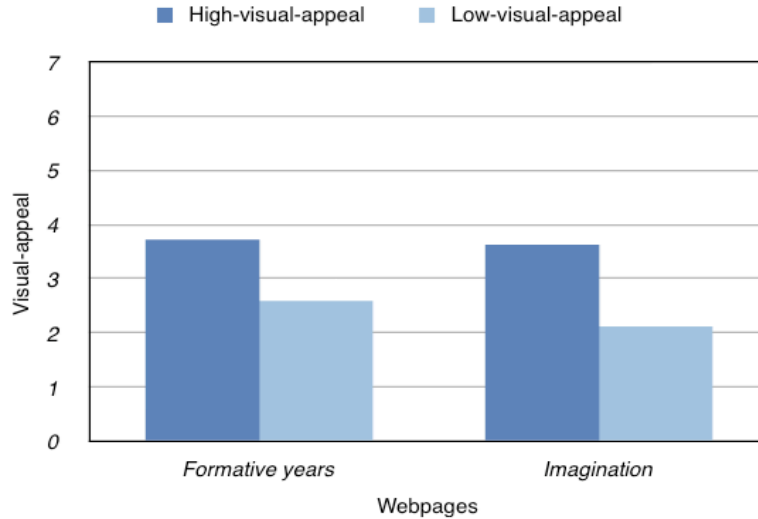
Note: Visual appeal is a semantic differential scale starting at unattractive (=1) and ending at attractive (=7). A total of 20 judgements were made for each version of a website.

Both high visual appeal versions of the websites are rated significantly higher than their low visual appeal counterparts. Table 10 shows the results for the websites.

Table 10

Website	Version	Mean rating	SD	Visual appeal gap
Formative years	Low visual appeal	2.60	.821	$t(30.087)=3.092$ *
	High visual appeal	3.75	1.446	
Imagination	Low visual appeal	2.15	.875	$t(31.080)=3.939$ *
	High visual appeal	3.65	1.461	

*p<.01

The second round validated the remaining websites.

Discussion and limitations

This study has validated the designs of the twelve websites. Of each website a version with low visual appeal and a version with high visual appeal have been created. It is important to note that a high visual appeal version of a website does not have to be beautiful. If we consider a website version with a mean rating of 4 or higher beautiful, 7 of the high visual appeal version can be considered beautiful, the other 5 cannot be considered beautiful. This has implications on how the websites can be used: The websites can be used to investigate the relation of visual appeal, rather than the role of beauty. Male and female participants on average rate the websites similarly.

There are some limitations to the results:

In this study students participated. Although they fit the profile of "information searchers", the conclusions may not generalize to all types of information searchers. A second limitation is the limited time period during which the participants were shown the screens. Although a rating after a brief exposure is highly correlated with the rating after a longer exposure (Lindgaard et al., 2006; Tractinsky et al., 2006), research also suggests that participants are more positive when the exposure time is longer (Tractinsky et al., 2006).

Appendix B: Manipulation guidelines

To prevent accidental manipulations of the information quality of websites used in the main study, a list of guidelines has been created for the designers. The list is based on the advice given by Briggs (2002), Fogg (2002) and Rieh (2002) on how to improve the perceived information quality (or credibility) of a website. The positive guidelines presented by these authors were turned around and used to provide the designers with a list of elements that should *not* be manipulated. Table 11 discusses the guidelines.

Table 11

Guidelines for the manipulation of the visual appeal of the websites used in the main study

Element	Guideline
Content	
	G1: The type of information must remain the same between the original website and the manipulated website. A website must remain a website and must not be replaced by for example a flash application with the same functionality. Rationale: The type of information influences the information quality (Rieh, 2002).
	G2: The actual textual content of the manipulated screen will not be altered. Spelling errors will not be corrected. Rationale: Textual content and small errors in content influence perceived information quality (Briggs et al., 2002).
	G3: The designer is not allowed to add headings, or to restructure the content. G4: The place of the information must remain largely the same. Rationale: The structuring of the content has an interaction with the perceived information quality (Rieh, 2002).
	G5: All labels indicating trustworthiness (such as trustee) must be removed in the original screen. No labels will be added to the manipulated websites.

	Rationale: Briggs recommends using seals of approval (Briggs et al., 2002).
	G6: The titles of the manipulated website and original website are the same. Rationale: Rieh (2002) suggests an influence of the title of a website on perceived information quality.
Source of information	
	G7: The source of the information (site name, URL) will not be changed. Rationale: People have trust in some organisations. This trust must remain stable. According to Rieh, the source of the information has an interaction with cognitive authority; which influences perceived information quality (Rieh, 2002).
	G8: No references to brands will be added. Rationale: Brands can function as a heuristic that indicates quality (Briggs et al., 2002; Chen & Chaiken, 1999).
Functionality	
	G9: The original website and the manipulated website clearly show the same functionality. Rationale: As indicated in earlier in this document, some experiments on visual appeal accidentally removed behaviour from the stimuli. This has implications on the ease of use and usefulness of the website and thus on the credibility of the website (Fogg, 2002).
	G10: All ads will be removed. Rationale: Advertisements negatively influence the perceived information quality of the website (Fogg, 2002; Rieh, 2002).
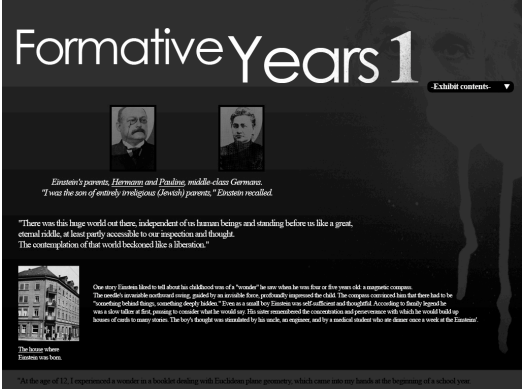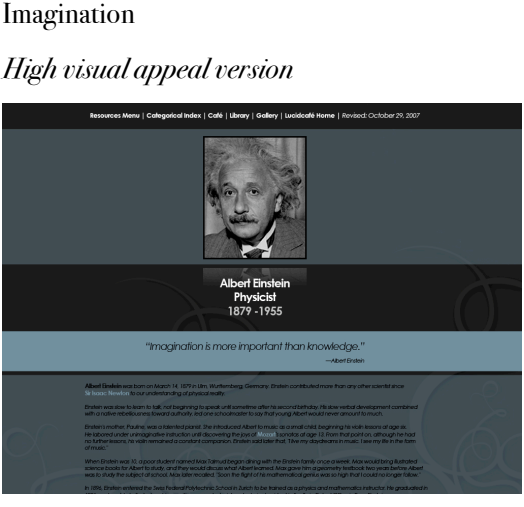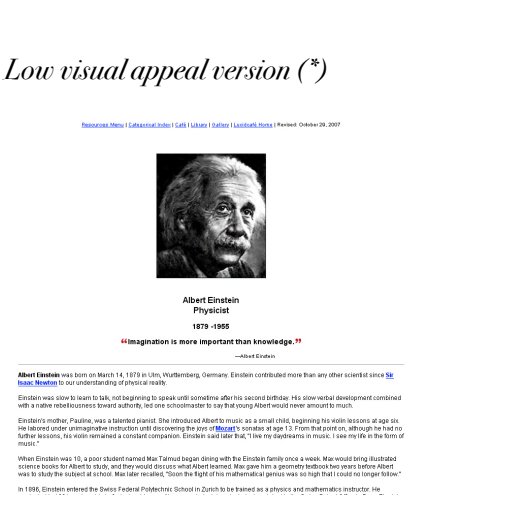
Appendix C: The websites

This appendix shows screenshots of the websites used in the experiment.

(*) Website was directly taken from the Internet

(**) Website was directly taken from the Internet, but received a small manipulation.

Einstein task

Formative years	
High visual appeal version	*Low visual appeal version (**)*
Imagination	
High visual appeal version	*Low visual appeal version (*)*
Early years	
High visual appeal version	*Low visual appeal version (*)*

Einstein

High visual appeal version

Low visual appeal (*)

Rugen task

Einfach

High visual appeal version (**)

Low visual appeal

100% German

High visual appeal version (*)

Low visual appeal

Isle	
High visual appeal version	*Low visual appeal (*)*

Largest	
High visual appeal version	*Low visual appeal (*)*

Headache task

Medline	
High visual appeal version (*)	*Low visual appeal*

Illustrated

High visual appeal version (*)

Low visual appeal

Health

High visual appeal version (*)

Low visual appeal

Netdoctor

High visual appeal version(*)

Low visual appeal

Rejected websites

The websites below were not included in the main experiment. They were used in the pre-study to determine whether they could be included in the main study.

Formative years

Rejected high visual appeal version

Rejected Low visual appeal version (*)

Imagination

High visual appeal version

Low visual appeal version (*)

Rejected websites for the Rugen task

Largest

High–visual appeal version

Isle

High visual appeal version

Appendix D: Information search tasks

This appendix presents the instructional texts (in Dutch) for the three information search tasks.

Einstein task

Find information about the private life of Einstein for a presentation.

"Je bent gevraagd een korte presentatie te geven over het privé leven van Albert Einstein. Voor deze presentatie zoek je verdiepende informatie. In de zaal zullen een aantal mensen aanwezig zijn die al enige kennis hebben van het onderwerp."

"Je krijgt nu kort vier pagina's te zien die allemaal informatie bevatten over het privé leven van Einstein. Elke pagina wordt kort getoond, waarna je je eerste indruk moet geven over de kwaliteit die je *verwacht* dat de informatie op de pagina heeft. Probeer de antwoorden op de vragen zo snel mogelijk te geven en snel op de "volgende" knop te drukken."

Rugen task

The second task is to look for information on a holiday resort on the German isle Rugen.

"Je oriënteert je op een vakantie voor de volgende zomer. Je hebt gehoord van het eiland Rugen dat net boven de Noordoost-kust van Duitsland ligt. Je bent nu op zoek naar meer informatie over het eiland. Wat is er te doen? Is het er toeristisch? Is het een leuk eiland?"

"Je krijgt nu kort vier pagina's te zien die allemaal informatie bevatten over het eiland Rugen. Elke pagina wordt kort getoond, waarna je je eerste indruk moet geven over de kwaliteit die je *verwacht* dat de informatie op pagina heeft. Probeer de antwoorden op de vragen zo snel mogelijk te geven en snel op de "volgende" knop te drukken."

Headache task

Because of regular headaches the participant is looking for medical information on the Internet.

"Omdat je de laatste tijd regelmatig hoofdpijn hebt, ben je op zoek naar meer informatie over de achtergronden hiervan."

Je krijgt nu kort vier pagina's te zien die allemaal medische informatie bevatten. Elke pagina wordt kort getoond, waarna je je eerste indruk moet geven over de kwaliteit die je *verwacht* dat de informatie op pagina heeft. Probeer de antwoorden op de vragen zo snel mogelijk te geven en snel op de "volgende" knop te drukken."

Appendix E: Test tool

Introduction

The results for this study were gathered using a custom-made test tool. This appendix briefly discusses the design of the tool. Some of the technical decisions are explained, and a short overview is given of the data that the tool recorded.

Design

The tool was designed to provide a maximum exposure to the group of potential participants. The following principles guided the design:

- The tool should be easy to distribute on the Internet.
  The participant should be able to participate in the experiment without installing any additional software.
- Recorded data should be easy to import in SPSS.
- It is essential that all images of websites are shown without a delay.
- Mistakes of the user should be prevented.
  Using Internet browsers to gather data for experiments is likely to lead to several errors. The participant uses an interface of the experiment within the interface of the Internet browser. Participants might use the "back" button; but many Ajax/ Flash applications will end up in a wrong internal state when the user does this. In the case of software for experiments, incorrect data may then be recorded. Another problem is that participants may enter the experiment multiple times, by visiting the URL of the experiment more than once.
- Exclusive use of the experiment.
  When the user is doing the experiment, he should not be distracted by other software programs that are running.
- The participant can be invited to take part in the second phase of the experiment.
  A week after the first phase has taken place, the participants should be invited to take part in the second phase. To be able to save the new information without overwriting any information gathered in the first round, the participant must be identified.

Technical choices

The following technical choices were made to make sure the design requirements were met:

- The tool uses full screen mode to make sure the participant is not distracted during the test.
  When the participant exits the full screen mode, the experiment stops.

- During the experiment the keyboard does not work, to prevent mistakes by the participant.
- Every participant receives a unique e-mail with a unique link to the test tool. When the participant clicks on the link, he is immediately identified.
- The test tool utilizes the Adobe Flash plug-in to create an interface. This environment has a high installed base. Furthermore it is the only environment that allows for creating a full-screen interface.
- All images are pre-loaded.
- The test tool remembers the status of each participant. It is not possible to return to a previous round.
- Data from the ldap server of the two universities is used to send every potential participant a unique e-mail. Data from the ldap is automatically used to fill in the background of the participant, to relieve the participant of the burden to fill in any data about his background.

Evaluation of the test tool

The test tool worked the way it should, but it had some shortcomings when used in the Internet setting. The solutions to some of the shortcoming are presented here as recommendations for future Internet-based research:

- Register the reason why participants leave the experiment. For example, people may leave the experiment website when a pop-up window is opened. Make sure that the participant only has one question to answer, but provide an opportunity to be verbose.
- Record as much information as possible about the environment in which the participant is working. Characteristics such as screen resolution, browser version, Flash version and operating system may al be relevant. In this study a particular combination of Internet Explorer with a particular version of Flash created problems that did not show up in any other configurations.
- Send each participant a unique link to the online experiment. This allows for tracking who visited the experiment website, and who did not.
- Provide an opportunity for participants to leave the experiment and come back to it later to finish the experiment.
- Make sure that participants can provide feedback on the experiment through channels other than the experiment website itself; for example, by leaving an e-mail address.

Data

The test tool recorded the following data:

About potential participants:

- What e-mail addresses received an invitation to the test?

- When was an e-mail send?
- What is the background (the study program) of the potential participant?
- Was a reminder sent when the potential participant did not respond?

About participants:
- Information about the computer of the participant: screen resolution, browser version, operating system, Flash version, and IP address.
- The frequency with which the participant uses the Internet.
- Start and end times for both the short exposure and the long exposure conditions.
- When did the participant start using the Internet?
- E-mail address.
- Start and end time.
- Status of the participant (finished questionnaire, rated some screens, rated all of the screens under short exposure, rated some screens in long exposure condition, rated all of the screens for both conditions, something went wrong).
- Study program of the participant.
- Gender and age.
- Lab setting or Internet setting.

Ratings:
- Website version ID.
- Time stamp.
- Order of the task, order of the website.
- Expected information goodness, expected information usefulness.
- Reaction time for rating of expected information goodness, reaction time for rating of expected information usefulness.

Selection task:
- Order of each website.
- Selected websites.

Screens

To give an impression of the test tool, the main screens are presented here:

Introduction screen:

Identificatie (versie 2.1)

Universiteit Twente
de ondernemende universiteit

Welkom bij dit onderzoek naar de eerste indruk van webpagina's

Het onderzoek wordt uitgevoerd voor de Universiteit Twente, Toegepaste Communicatie Wetenschappen.

Tijdens dit onderzoek geef je je "eerste indruk" van 12 webpagina's.

Dit onderzoek bestaat uit twee delen. Beide delen duren ongeveer tien minuten. Voor het tweede deel van het onderzoek wordt je een week na afloop van het eerste onderzoek uitgenodigd.

Sluit MSN, en zet je telefoon uit.

Start

The two screens of the questionnaire about web usage

Start

Wat is je geslacht:

⊙ Man

○ Vrouw

Wat is je leeftijd:

[   ] jaar

Vorige   Volgende

Start

Hoevaak gebruik je het internet:

○ Dagelijks

○ Niet dagelijks maar minstens een keer per week

○ Niet wekelijks maar minstens een keer per maand

○ Minder dan een keer per maand

Hoelang maak je al gebruik van het internet:

○ Minder dan een jaar

○ Langer dan een jaar, maar minder dan vijf jaar

○ Langer dan vijf jaar

Vorige   Start schermen

Introduction of the rating tasks

Voorbereiding

Uitleg beoordeling schermen

Je krijgt achtereenvolgens drie taken te zien. Tijdens een taak stel je
voor dat je informatie zoekt over een bepaald onderwerp.
Bij elke taak beoordeel je vier schermen. Je geeft aan of je verwacht dat
je "goede" informatie over het onderwerp kunt vinden op de website en of
je verwacht dat de informatie die je vindt goed "bruikbaar" is.
Het onderzoek draait om je verwachting, het tempo is te hoog om een
afgewogen oordeel te krijgen.

Omdat het onderzoek draait om je eerste indruk, moet je zo snel
mogelijk antwoord geven.

Elke pagina wordt kort getoond, dus let goed op!

Om aan het tempo te wennen krijg je eerst twee oefenschermen.

Start

Introduction of the Headache task

Taak 1 van 3

Hoofdpijn

Omdat je de laatste tijd regelmatig hoofdpijn hebt, ben je
op zoek naar meer informatie over de achtergronden
hiervan.

Je krijgt nu kort vier pagina's te zien die allemaal medische
informatie bevatten. Elke pagina wordt kort getoond,
waarna je je eerste indruk moet geven over de kwaliteit die
je verwacht dat de informatie op pagina heeft. Probeer de
antwoorden op de vragen zo snel mogelijk te geven.

Start taak

Rating screens for each website

64

Wat verwacht je van de kwaliteit van de informatie op de pagina?

Slecht ○○○○○○ Goed   ○ Geen idee

Hoe bruikbaar verwacht je dat de informatie op de pagina is?

Onbruikbaar ○○○○○○ Bruikbaar   ○ Geen idee

Progress and speed control screen

Voortgang

Als je klaar bent voor het volgende scherm, druk dan op "volgend scherm".

Volgend scherm

Selection screen for the headache task

Selecteren websites

Selecteer de vier schermen met de hoogste informatie kwaliteit:

☐ Selecteer   ☐ Selecteer   ☐ Selecteer   ☐ Selecteer

☐ Selecteer   ☐ Selecteer   ☐ Selecteer   ☐ Selecteer

End screens of the experiment

Je bent aan het einde gekomen van de vragen. Bedankt voor je medewerking!

☐ Ik heb technische problemen ondervonden tijdens de test

Rond af

Bedankt voor je medewerking!!

Over een aantal weken ontvang je een email waarin staat of je de bioscoopkaartjes hebt gewonnen.
Deze email ontvang je op het adres waar je ook de uitnodiging voor het onderzoek hebt gekregen.