NTNU
Norwegian University of
Science and Technology

Don`t Make It Too Complex

How surface credibility of websites is affected
by visual aesthetics.

Are Ingulfsen

Master in Interaction Design
Submission date: December 2017
Supervisor: Frode Volden, ID

Norwegian University of Science and Technology
Department of Design

NTNU
Norwegian University of
Science and Technology

Don`t Make It Too Complex

How surface credibility of websites is
affected by visual aesthetics.

Are Ingulfsen

Master in Interaction design
Delivered: Desember, 2017
Written by: Are Ingulfsen
Supervisor: Frode Volden

Norwegian University of Science and Technology
Department of Design

Abstract

With rise of the World Wide Web there has been a steady increase in use. People use internet for more and more services such as banking, news and shopping. This has created an increased competitive marked between businesses, but also increased the need for users themselves to validate the credibility of the source. This is caused by the lack of control, as most people has the opportunity to create website for both honest and dishonest purposes. The importance of credibility, especially surface credibility, is therefore increased as users constantly make judgements and decide in a short time if the page is worth staying on or not. Judgements of credibility has been found to be influenced by aesthetics.

The purpose of this study was to increase the understanding of how differences in visual aesthetics, particularly the property of complexity, influences the perception of credibility. A better understanding of how our perception might be influenced can benefit designers as they can design for increased surface credibility and enhance the user experience by creating a good first impression. Companies reduce the risk of losing customers based on a bad or wrongly impression of surface credibility. Users themselves can understand how their perception might be influenced which might cause them to stay long enough to cognitively judge if a website is credible or not.

Complexity is believed, according to Berlyne`s theory, to impact the aesthetic impression by causing changes in arousal and this study found that the degree of complexity is influencing perception of attractiveness and credibility in a short exposure. This indicates that complexity is an important property of visual aesthetics and should be considered when creating designs. It was found to be a linear relationship between degree of complexity and both perceptions of credibility and attractiveness indicating that users prefer simpler designs. Although not too simple as indications were made that the least complex webpages were found slightly boring. The study also looked into how differences between conventional and unconventional navigation and logo might influence credibility, but no significant difference was found. There were also found a correlation between attractiveness and credibility supporting previous findings of Fogg and Robins & Holmes. This study supports the notion that aesthetics is an important factor in credibility judgements and that particularly complexity is an important influencer of aesthetics especially in a short exposure.

Foreword

I would like to thank my family and friends for their support and distractions during the long process of this thesis, and all the years before. I also want to thank my supervisor Frode Volden for feedback along the way and particular expertise in regards of statistical analysis. Lastly, I would also give a thanks to my colleagues and workplace that have given me support and have made it possible to gain knowledge alongside working as a designer.

This thesis started from my work in the subject Research Project Planning at NTNU Gjøvik, and marks now the end of my time as a student at the school. It has been some good years which have granted me many great experiences and I have learned a lot. While this marks the end of my time as a student it does not end the everlasting search for knowledge. That will continue, but now with slightly more money in the pocket.

Hamar, 12.12.2017

Are Ingulfsen

Index of contents

1.0 Introduction

Our society has changed in many ways since the emergence of the World Wide Web which has, thanks to technical advances, become increasingly more available. One aspect of the increased availability and following increased usage (Anderson, 2015) is that we have both easier access to receiving and distributing information. This has changed the way we communicate and it has created a competitive market that today has no geographical limitations. The tools necessary to be able to run a company, service or distribute information is easily available to most people. Companies and organizations do not only have to compete with the company next door, but with thousands of web pages fighting for people's attention in the hope that users will select their products, services or information. This isn't limited to one business sector either, but is happening when people are selecting which schools to apply to, which doctor to choose, which news source to read, where to go to dinner etc. We, as users, are used to finding information, products and services online – and with the amount of choices available the new currency is our attention and trust (Elliott, 2017). Companies need to persuade users to use and trust them instead of their competitors.

Persuasion is a large field in itself that has been in focus of different academic fields for centuries, but if we look at how technology can be persuasive then one of main keys to persuasion is, according to B. J Fogg (2003), the perceived credibility of the website and the people behind it. However, unless an organization or company already have earned credibility through reputation or previous experience there isn't much time to convince users. Recent research has shown that our attention span has changed from 12 seconds in 2000 to 8 seconds now (Watson, 2015), and that the first 10 seconds are crucial to whether a user will leave a page or not (Nielsen, 2011). In the short time that users decide whether to stay on a web page it is likely that *surface credibility,* if the user have no previous experience or knowledge about it, is a strong factor in persuading users to stay.

There are several factors affecting surface credibility, but one of those factors that has been found to have a strong influence is visual aesthetics (Fogg, 2003; Robins & Holmes, 2008). Although there has been previous research on both visual aesthetics and credibility there aren`t many clear answers other than that attractive webpages are viewed as more credible than non-attractive ones. This does not give a lot for designers and businesses to

work on. There is little information on how differences in visual aesthetics affects judgements of credibility. Interaction design as a field is concerned with creating intuitive and good experiences between users and the products they use. If a user`s first impression judges the surface credibility to be low then the experience is already reduced, and that user might not even try the product. It is therefore important to gain an increased understanding of how surface credibility is affected by differences in visual aesthetics, such as complexity, to reduce the risk of losing users based on appearances. Complexity is one property which is theorised by Berlyne (1971) to impact aesthetic impression of a stimuli by either increasing or decreasing arousal. Berlyne (1971) believes that arousal causes positive emotions while it is increased up to an optimal level which is a considered to be a moderate amount, believed to be correlated to a moderate amount of complexity, while further increase in arousal causes a decline of the positive emotions eventually turning into negative emotion as the level of arousal gets to an uncomfortable level. This master thesis will investigate how complexity and common web page elements, navigation and logo, affect user first-impressions of websites in order to increase the understanding of the effects differences in visual aesthetics might have on our judgement.

1.1 Keywords

Interaction design, credibility, surface credibility, visual aesthetics, first-impression, complexity,

1.2 Research questions and hypoteses

The main objective of this study is to investigate how the relationship between visual aesthetics, more specifically the aesthetic property of complexity and two common domain elements in navigation and logo, and user's judgement of surface credibility of a web page based on a first impression. In addition, this thesis will investigate if the immediate psychological response of the visual aesthetics (first impression) is causing a confirmation bias on the credibility judgement even when cognitive resources join in on the evaluation. In order to reach those objectives, I want to address the following sub questions and hypotheses:

1. How do differences in visual complexity impact the judgement of surface credibility of webpages based on a short exposure?
    a. Webpages with a medium degree, on a three-point scale, of visual complexity will be rated higher than webpages with low or high degree.
    b. Webpages with a low degree, on a three-point scale, of visual complexity will be rated higher than webpages with a high degree of complexity.
    c. There is a correlation between users´ rating of webpage between expertise, trustworthiness and attractiveness.
2. How do common elements in web design, such as the navigation and logo of a web page, impacts surface credibility when making a physiological judgement during a short exposure?
    a. Users will fixate on a clear logo on a website if present.
    b. Webpages with a clear logo are rated as more professional (higher expertise) than webpages without a clear logo.
    c. Users will fixate on navigation elements if present and after conventions.
    d. Webpages with clear and visible navigation options are rated as more trustworthy than webpages with hidden navigation options.
3. How are judgements of surface credibility of the same content impacted when moving from a physiological judgement based on a short exposure to a cognitive judgement based on a longer exposure?
    a. Judgement of surface credibility is not subject to change from a short exposure to a longer exposure indicating a confirmation bias from the first impression.

1.3 Reasoning, motivation and planned contributions

Interaction design is about removing negative aspects and enhancing positive aspects of a user experience in order to make better systems, products and services for the end users (Rogers, Sharp & Preece, 2011). Traditionally there has been a strong focus on usability in the field of interaction design, but over the last decade or so the focus of research has increasingly turned to all aspects of the user's experiences, among them aesthetics (Lavie & Tractinsky, 2004). Aesthetics has been found to impact our perception of credibility (Fogg, 2003), but it is important to investigate further the effects of different aspects of aesthetics,

other than attractiveness, on credibility judgements. This is especially important regarding to surface credibility in the competitive and non-regulated environment of the World Wide Web. If a website can't persuade users to stay long enough, then a user might never figure out how easy a website is to use or how important the content is. Lazar, Feng & Hochheiser (2010) states that research in the field of Human-Computer Interaction should be of relevance to people, organizations or design, and in addition be able to influence practical applications within the field such as the design of interfaces. Uncovering more knowledge of the relationship between aesthetics and credibility can benefit several groups. Designers benefit as they gain increased knowledge of how one can design for increased surface credibility and enhance the user experience by creating a good first impression. Companies selling products or services can reduce the risk of losing customers based on a bad impression of surface credibility. Even users themselves may benefit by increased knowledge of how they judge websites causing them to stay for a longer period in order to cognitively judge if a website is worth using.

The purpose of this thesis is to build upon the results of Fogg (2003) and Robins & Holmes (2008) of how visual aesthetics impacts the judgement of credibility and create more specific insight into how visual aesthetics impacts credibility – in terms of complexity, over exposure time and specific elements. In addition, this thesis will test if Berlyne´s theory (1971) on the inverted u-relationship between arousal and complexity in regards to web pages. The result, if successful, will create a deeper insight into the effects that differences in aesthetics have on surface credibility. It will either support or refute Berlyne`s theory (1971) on arousal and complexity, as well as the findings of B. J Fogg (2003); Robins & Holmes (2008).

2.0 Background

This chapter covers the background for the topic of the master thesis, and will give insights into existing research done in the fields involved. The first section will cover the topic of credibility before going through some of the existing research on immediate responses, visual aesthetics and specifically complexity.

2.1 Credibility

The topic of credibility has been around in the academic fields, especially communication/marketing and psychology, for a while. Credibility is defined by the dictionary as the 'the quality of being trusted and believed in' and 'the quality of being convincing or believable' (Google Dictionary, 2017). It origins from the Latin word *credere* which means to believe (Fogg, 2003). In other words, a person or product that is deemed credible is a believable person or product. It is an important quality, as credible sources, be it people or webpages, have been found to have increased ability to influence those that listen in terms of opinions, attitudes and behaviours (Fogg, 2003). We are more likely to believe in what they have to communicate as the source is deemed believable. The following sections give an insight into credibility in society, dimensions of credibility and surface credibility.

2.1.1 Credibility in society and online

Credibility is in our society an important quality as there are plenty of situations or fields of work where credibility, or being believable, matters. When it comes to leadership, in a company or a sports team, the members of the organization want their leader to have qualities like honesty, competence and to be inspiring according to Kouzes & Posner (2011). These three qualities correspond with trustworthiness, expertise and dynamism – three proposed dimensions of source credibility. It displays the importance of credibility in leadership as people want to be able to believe in their leaders, and it caused Kouzes and Posner (2011) to state that credibility is the foundation of leadership. It is therefore easy to understand that credibility also is essential in politics and why scandals have played an important role in the outcome of elections or caused leaders/parties to lose their positions as their credibility is reduced. In Norway, in the 2017 election, there were several news cases around Jonas Gahr Støre (Arbeiderpartiet) that might have been a factor in the political party`s loss at the election (NTB, 2017; Svaar, 2017).

With credibility being an important quality in many fields this study mainly directs its focus on so called computer credibility – more precisely on the perception of credibility with regards to web pages. Since the internet was released in 1991 it has gain increased usage and implementation in our daily life due to increased availability and technical advances. In

Norway the number of people, between the age of 9-79, that use the internet daily has gone from 27 percent in 2000 to 89 percent in 2016, and in the same period traditional media such as radio, TV and newspapers has experienced a decrease in usage (SSB, 2017a). With this increase there are also an increase in the vast variety of services, stores and information provided online. Since 2006 there has been an increase in the use of the internet for reading news, using bank services, buying entertainment (music and movies) and buying clothes (SSB, 2017b).

Computers were originally perceived as infallible, but that perception has been eroded over time (Fogg & Tseng, 1999). The same is maybe even more true for web credibility. We use the web for a variety of purposes, and over time it has been clear that the quality from web page to web page varies just as much. There are plenty of great resources for services and information online, but also plenty of hoaxes and deceiving information. In a study by Flanagin & Metzger (2000) it was found that internet users are doing more content verification behaviours while being online which in turn is causing them to be more attentive and selective of web pages. This might be caused by the lack of gatekeeping online so that compared to traditional media such as newspapers, internet users have to do more verification themselves, rather than trusting a system of journalistic processes in traditional media (Stavrositu & Sundar, 2008). Considering users actively conduct evaluations of webpages, either cognitively or automatically, Fogg & Tseng (1999) has suggested a conceptual framework of four possible ways to evaluate credibility where two ways consist of correct judgements and two ways consist of incorrect judgements. One is an “Gullibility Error” where the user incorrectly perceives a product to be credible, the other but more important error for this study is “Incredulity Error” – when a is unwilling or unable to believe the product when it is credible (Fogg & Tseng, 1999). This error is a major concern for designers because if credibility is perceived to be low then users are likely to quickly leave the page and there is no chance of persuading users to use the site or be influenced by any message they wish to convey (Fogg, 2003; Robins & Holmes, 2008).

2.1.2 Dimensions of credibility

In 1953 Hovland et al. (1953) wrote that credibility is a perceived quality that doesn’t physically reside within something. It isn’t something we can touch or measure directly through physical attributes. Since Hoveland et al. (1953) defined credibility as a perceived

quality there has been a strong agreement among researchers and academics of this fact (Fogg, 2003; Robins & Holmes, 2008; Pornpitakpan, 2004). While credibility is a perceived quality and therefore up to individual judgement it is still like evaluating beauty, according to Fogg (2003), where people still seem to agree in most cases. It seems therefore that there are some common dimensions or factors that we use to evaluate the credibility of a source. Hoveland et al. (1953) and plenty of researchers following them have suggested different dimensions, although not always being in agreement on which ones are true (Fogg, 2003), that we base our judgement of credibility on. Berlo, Lemert and Mertz (1969) came for example up with three dimension´s receivers of a message used to evaluate the source: *Safety, Qualification* and *Dynamism***.** While Whitehead (1968) found there to be four dimensions to credibility which he considered to be: trustworthiness, objectivity, dynamism and professionalism (competence). McCroskey (1966), as cited by Whitehead (1968), found two other dimensions, which was authoritativeness and character.

Despite the fact that several research studies through history, as seen in examples above, have found several different dimensions there seem, over the last decades, to be a fairly strong agreement on two key dimensions: perceived *trustworthiness* and perceived *expertise* (Fogg, 2003). These are similar to *Safety* and *Qualification* (Berlo; Lemert & Mertz, 1969), and the dimensions of trustworthiness and professionalism from Whitehead (1968). It can be argued that Whitehead`s (1968) factor of objectivity, open-mindness and level of objectivity of the source, has been absorbed somewhat by the newer definition of trustworthiness. Both trustworthiness and expertise have been used quite commonly in research on credibility through history (Pornpitakpan, 2004). Trustworthiness covers how we perceive a person or objects to be good, fair, truthful and unbiased. It generally builds on our perception of the goodness and morality of the source that might be affected by several factors: whether the source appears to be unbiased and act fairly, if the source appears to act against its own interest and the similarity between the source and receiver (Fogg, 2003). If the source appears to be good, for example by doing charity work or giving away part of the profits of sales to charity, then the source is more likely to be considered credible. That is however as long as the expertise of the source is considered too be good as well or unknown to the receiver, and not considered to be poor (Fogg, 2003). Users judgement of these two key dimensions, trustworthiness and expertise, together is what equals perceived credibility (Fogg, 2003). Expertise as a dimension covers how we perceive the quality of

knowledge and, as the name suggests, expertise of a person or object. For example, a lecturer with a doctor's degree within a field is likely to be considered to have higher degree of expertise than a lecturer with a bachelor degree. Similar to trustworthiness there are several factors that might affect expertise such as a titles, evidence of praise and accomplishment etc. (Fogg, 2003).

The two dimensions together, as stated previously, equals perceived credibility. It are however some uncertainties as to the degree each of the two dimensions affects credibility. In a review of research done on credibility from the last five decades it can be seen that some experiments find trustworthiness to affect credibility judgement more than expertise, other experiments find it to be the opposite and some have found them to be equally important (Pornpitakpan, 2004). This gives an indication that the two dimensions might be weighted differently when users are assessing credibility. It is however no clear answer to this weighting of the two dimensions, but while the two dimensions can exist without each other an object or source won't be considered very credible without high levels of both trustworthiness and expertise (Fogg, 2003).

As mentioned earlier there are several factors that might affect credibility. In regards to especially source credibility the receiver, for example a web user, existing attitudes and opinions towards the topic or subject of the message being communicated might impact the effect of persuasion. Highly credible sources do not necessarily have the strongest persuasive power (Harmon & Coney, 1982). In an experiment it was found that a highly credible source was most effective in terms of persuasion when communicating towards receivers that had unfavourable opinions and attitudes towards the message, while moderate credibility sources were most effective towards receivers that had favourable attitudes and opinions (Harmon & Coney, 1982). Harmon and Coney (1982) attributed this to the high credibility source inhibiting activations of own-thoughts and increasing the acceptance of the message being communicated. While sources of moderate credibility better activated already favourable thoughts towards the message. This highlights an interesting and important point which is that previous attitudes might affect how users perceive new messages that are either favourable or unfavourable too the previous position. This study will mainly look at surface credibility, which will be covered in the next section, and therefore try to avoid the effects of previous experiences as much as possible to ensure that it is the first-impression that impacts the judgements. However previous experiences

and attitudes towards similar products, messages or content during the experiment cannon be excluded and may have some degree of impact.

2.1.3 Surface credibility

Credibility can also be divided into four different types of credibility according to Fogg (2003): that is *Presumed, Surface, Reputed* and *Earned*. This division is specifically suggested as a way to divide credibility when it comes to computing products which includes websites. Today there is limited research done to support this division of credibility, but it seems to have been excepted and there are several studies that have used this division in their research (Liu & Huang, 2013; Wathen & Burkell, 2001) to clarify credibility. Liu (2004) also suggests two additional types: *verifiable* credibility and *cost-effort* credibility.

From the six types there is only surface credibility that has an immediate impact on the user's perception of credibility that is controllable by the company or organization. Presumed credibility is based on general assumptions that users may have despite having never visited a site, reputed credibility is based others endorsement and lastly earned credibility is based on use over time (Fogg, 2003). Surface credibility is based on the first impression of surface traits and initial inspection (Fogg, 2003). These judgements are constantly being made throughout a day whenever we meet something new, and it allows us to make quick decisions. Both Fogg (2003) and Robins & Holmes (2008) argue that achieving high surface credibility it is especially important on the web as it is critical in convincing users to stay on a site and potentially become a regular user in the future.

Recent research seems to support that the initial meeting between users and a web page is important in determining whether a user stays or not. In research by Liu, White and Dumais (2010), as cited by Nielsen (2011), where they tracked time spent by users on 205 873 web pages, they found that the first 10 seconds of a visit is critical when deciding whether to leave or stay for a substantial period. Nielsen (2011) argues that this is because the users are aware of the variable quality of web pages and therefore decide quite quickly if it's worth sticking around. In another large research study with over 2,500 participants at Stanford by Fogg et al. (2003) they also argue that the short time spent on a website causes users to evaluate credibility quickly often using peripheral cues as a basis for their judgement. In the same study it was found that elements that increase credibility within the surface credibility category are highly related to the design and look of a web page – almost

50% of all comments where regarding what they labelled as “design look” (Fogg et al., 2003). The reasons for this might be many. It might come from the variable quality and the vast number of sites offering the same services and information that leads users to make quick decisions to stay or try their luck elsewhere. It might also stem from the research experiment context where the task of evaluating the websites isn’t necessarily very important to the participants, and they might therefore rely more on peripheral cues in their evaluation compared to someone very motivated to evaluate it (Fogg et al., 2003). However, design is undoubtedly important. Several research studies have found that credibility is affected by visual attractiveness (Fogg, 2003; Fogg et al., 2003; Robins & Holmes, 2008). In addition to the “design look” being an important part of a user’s credibility judgement, 28.5% mentioned the “structure of information” and 25.1% mentioned “information focus” – both of which are partly affected by the way a website is designed (Fogg et al., 2003). While designing a website so that it looks “professionally designed” increases credibility there are several elements that might reduce surface credibility: difficulty in distinguishing ads from content, pop up advertisements, more than one ad etc. (Fogg, 2003).

2.2 Automatic & visual judgements

Humans have automatic responses to the environment around them (Zajonic, 1980) and these responses might influence our subsequent judgements. The following sections cover research into automatic and visceral responses, as well as the speed at which we can make judgements based on visual cues.

2.2.1 Automatic responses

The research conducted by Fogg et al. (2003) give an indication of what people use to judge credibility, but as it is based on a questionnaire people are cognitively considering what they base their judgement on and not necessarily judging the webpage itself as we would do naturally. It does however get support from research done in other fields that find that both visual cues and design might be a strong factors in our judgement whether we are aware of it or not. Research has shown that we as humans have immediate and automatic responses to our environment, and that these responses (emotional) occur faster than rational and cognitive responses (Ekman, 1992; Zajonic 1980). Zajonic (1980) states that these responses “cannot always be voluntarily controlled” and may occur unintentionally. Because of the

speed at which these responses happen, as little as few milliseconds (Zajonic, 1980), they are likely to be based on visual or audible cues. This indicates that we have responses to what we see that we cannot control and since these are likely to be based on visual cues then obviously the design of websites might be of strong importance. These unintentional responses or judgements may influence our attitudes, opinions etc. before we experience and make intentional responses. In marketing and psychology there is an effect called the Halo effect which is defined as: "the influence of global evaluation on the evaluations of individual attributes of a person" (Nisbett & Wilson 1977). This is a general and unprecise definition which is pointed out by Nisbett & Wilson (1977). Humans subconsciously make judgements about people based on individual attributes such as appearance etc. If a person looks attractive, then we are more likely to think more positively about that person (Fogg, 2003). This is a stable finding in research, but it isn't only limited to people as the halo effect also occurs in for example digital interfaces such as a web pages. Tractinsky, Katz, & Ikar (2000) found that the perception of usability of a system is correlated to the aesthetics of the interface. This result has been confirmed by several other studies. It has been found that products that are considered to be aesthetically pleasing has caused users to perceive usability to be high (Sonderegger & Sauer, 2010; Ben-Bassat, Meyer, & Tractinsky, 2006). This raises the question if the perception of aesthetics might lead the user to confirmation bias in terms of credibility, we know that aesthetics influences the perception of credibility, but if a confirmation bias is present it might cause us to ignore elements that detract from the credibility.

2.2.2 Visual judgements

In more recent research by Lindgaard et al. (2006), they investigated how fast we form first-impressions of web pages based on visual aesthetics alone. Using three tests they showed users web pages of varying degrees of visual aesthetics, each web-site was shown twice in a randomized order, for 500 milliseconds (ms) (for two tests) and 50ms (for the last test). The users were asked to rate the pages on an unmarked line with very unattractive and very attractive on each end. The results showed that users were able to reliably rate a web page as unattractive or attractive in as little as 50ms, which supported their claim that "visual appeal could represent a mere exposure effect" (Lindgaard et al., 2006).

These judgements can to some degree be considered due to the short exposure time to be based on the visceral level, one of three levels of processing design (beauty) defined by Norman (2004) and referenced by several (Lindgaard et al., 2006; Robins & Holmes, 2008; Frohlich, 2004). The two other levels of processing are the behavioural level which happens, just as the visceral level, in a subconscious matter and the reflective level which happens when one consciously considers the beauty of an element (Norman, 2004). The visceral level is an immediate reaction to visual cues in design and is mostly a reaction to something being perceived as either good or bad, giving us a subconscious judgement of an element such as a website is attractive or not without considering specific elements.

This division into levels done by Norman (2004) is however put into question by Lindgaard et al. (2006) as using Norman´s definition of the visceral level (happening on a subconscious level) and making a judgement on how attractive it was on a scale like in the research done by Lindgaard et al. (2006) will be a response that is to some degree conscious and therefore happening on the reflective level. Seeing the stimuli material in as little as 50 milliseconds and making judgements based on that is however such a short time that it isn't likely to give much information to base cognitive judgements on, and while it is made consciously it should still be mostly based on an immediate feeling and reaction from seeing the stimuli. Although it is vulnerable to interpretation due to a somewhat conscious judgement, it could still be considered close to what Norman defines as a judgement on visceral level.

In a separate study Tractinsky et al. (2006) attempted to replicate the study done by Lindgaard et al. (2006), but they also found that the perception of attractiveness which was formed in 500ms was consistent with the same perception after 10 seconds. This indicates that our initial emotional response to visual aesthetics might stay fairly consistent even after cognitive resources sets in, strengthening the belief that first impressions plays a big part of our judgement.

2.3 Visual aesthetics

Aesthetics or the concept of beauty has a long tradition in philosophy (Pandir & Knight, 2006). Aesthetics has over recent years been recognized as a dimension that plays an important part in our judgement and perception of our surroundings, and is now considered

an important part of the user experience (Rogers, Sharp & Preece, 2011). The following sections covers existing research on aesthetics which creates the background for the study.

2.3.1 Classical approaches to aesthetics

It has been found, as previously mentioned, that the way a person or an interface looks is important. If something is considered beautiful or attractive it is found to increase our perception in a positive way, for example in terms of credibility (Fogg, 2003; Robins & Holms, 2008). While it is found that beauty has a strong effect on our perceptions it is however less clear what it is that makes us see something as beautiful instead of something else. The question of beauty has been part of discussions traced all the way back to ancient Greek philosophy and is still discussed today. One long lasting discussion within aesthetics is whether beauty resides in the object or if it is in the eye of the beholder (the subject). In ancient Greece it was proposed by Aristotle, as cited by Pandir & Knight (2006) from Hanfling (1992), that beauty in an object resided in the object itself and was shown through its properties. This is an objective view of aesthetics. One believes that the object through its properties or attributes decides whether it is beautiful. Later, according to Liu (2003), it was Kant (1790) that "established aesthetics as philosophical discipline". Kant defined beauty in 1790, as cited by Frohlic (2004), as not being present in the object, but as perceived by the receiver. Beauty is in the eye of the beholder. This view challenges the objective definition of aesthetics from Aristotle, and can be considered a subjective view on aesthetics. Subjectivism considers aesthetic judgement to be a result of "the pleasure or displeasure that perception of the object arouses in any spectator" (Ward, 1992; as cited in Pandir and Knight, 2006). An aesthetic impression can therefore differ from person to person despite any objective properties. This might be true to some extent – most people have experienced not agreeing on the beauty of an object just as we might disagree on the taste of a dish served in a restaurant. This is supported to some extent by Karvonen (2000) who has written that "aesthetics, if not emotions, have often been considered to be unique – everyone has a taste of their own, so universal beauty assessments do not hold". She does however also state that people`s taste might not be as independent, but is affected by what is considered beautiful in society around us through trends and conventions in our culture (Karvonen, 2000). That there are commonalties despite the view that aesthetics judgements are subjective is also found in an old essay by Hume from 1757. He suggested that because

people share common emotions and experiences we will also have commonalties in our aesthetic preferences (Pandir & Knight, 2006). It does seem to be true as we often do seem to agree on what`s beautiful as well (Fogg, 2003). In several of the research experiments previously mentioned by for example Robins & Holms (2008) and Lindgaard et al. (2006) the participants seem to have a general agreement on what stimuli they deem attractive and unattractive, with of course some exceptions.

Aesthetics impressions are however a lot more varied than just a judgement of beauty, but instead there are several aesthetic notions and an aesthetic impression is multi-dimensional as it is the results of several factors (Liu, 2003). An aesthetic impression is in other words not only based on visual cues, but other factors such as the situation and the intention behind the receiver can also affect the impression (Park, Choi & Kim, 2004). The impression of a website might be different if one is looking for a travel destination than if one is looking for jobs, or even if one is looking for a family holiday at the beach or an extreme sport adventure holiday. Liu (2003) discusses aesthetics from a multi-dimensional perspective where the impression of aesthetics is not only based on visual cues, but on the interplay between several senses such as the tactile, auditory, olfactory etc. where the visual is one of many. This is true for many products we use everyday, but when it comes to websites which is a medium we mainly experience visually (sometime with audio) – it is likely that visual cues are the major source of our impressions. As Fogg (2003) found, in relation to credibility, it does seem that design of and how we structure the information is a key part of the basis we make our judgement on. If we also consider the automatic responses we humans have to our environment, which due to the speed they happen at are likely to be based on visual and audio cues (Zajonic, 1980), then a first impressions of websites are likely to be mainly impacted by the visual system.

2.3.2 Aesthetics and the world wide web

There are many previous studies into the visual dimension of aesthetics in specifically regarding websites. One major dimension that has received a lot of focus is colour. According to Bonnardel, Piolat & Bigot (2011) colour has, been found to influence people's emotions, judgements and decisions in several different situations despite a lack of precise theoretical explanations. It has been found that cooler colours are often preferred over warmer colours, and that cool colours such as blue create more relaxation than warm

colours such as red (Bonnardel, Piolat & Bigot, 2011). In terms of web pages colour has also been found by Lindgaard (1999) to be a strong predicator of the appeal. Brady & Philipps (2003) found, in addition to colour, balance to be an important dimension in user's judgement of preference. Symmetry, especially vertical symmetry, has also been found by Tuch, Bargas-Avila & Opwis (2010) to have an effect on a straightforward judgement of beauty – interesting enough this effect was mainly found in male participants. This indicates that there might be gender difference in aesthetics without going into more explanation of the reason behind this. An experiment conducted by Moss, Gunn & Heller (2006) where they looked at websites created by female and male designers they found that there was a significant difference between genders in 13 of 23 factors which included colours, numbers of straight/rounded lines etc. This also supports the notion that there might be gender differences and it could be smart to consider such possible differences when conducting experiments on aesthetics.

2.3.3 Dimensions of aesthetics

There have been several attempts to look at aesthetics on a slightly higher level than specific elements and instead look at dimensions that together makes up an aesthetics perception. Looking at aesthetics can possible create an opportunity to judge aesthetics of products such as web site as a whole without considering elements on the page in isolation. Lavie & Tractinsky (2004) conducted a large study on user's perception of aesthetics of web pages, and proposed that user's perception was covered by two dimensions: Classical and expressive aesthetics. The former, classical aesthetics, was found to consist of what Lavie & Tractinsky (2004) considered to be classical notions of aesthetics which emphasize order and clear design. The dimension of classical aesthetics is also considered by them to be closely related to design guidelines or conventions, which is discussed later, and is advocated by directions of user-centred design. The dimensions of expressive aesthetics cover more of what Lavie & Tractinsky (2004) found to be connected to a designer's flair – how original and creative their work is. These dimensions have been examined by several studies since. Tuch, Bargas-Avila & Opwis (2010) found that vertical symmetry had a similar effect on both classical and expressive aesthetics, but the stronger effect was found on classical aesthetics. Which is natural with classical aesthetics being concerned with design being balanced, clear design etc.

The dimension of classical aesthetics shares similarities with simplicity, one of four interrelated facets of perceived visual aesthetics created as a measure of aesthetics by Moshagen and Thielsch (2010) through seven studies. Simplicity covers the perception of clarity, orderliness, grouping, homogeneity and balance. It is also considered by Moshagen and Thielsch (2010), like classical aesthetics by Lavie and Tractinsky (2004), to be based on classic parameters that have been used from a historic viewpoint to measure the aesthetic value of objects. The other facets suggested by Moshagen and Thielsch (2010) is diversity which covers dynamics, novelty and creativity. This is deemed an important facet by Moshagen and Thielsch (2010) because of Berlyne`s (1971) theory, which we will come back to, that states that a stimuli material need to be of a certain level of complexity to cause a certain level of arousal. If a stimulus has low level of arousal then it will be perceived as boring and therefore not aesthetically pleasing, but too high levels of arousal will also have a negative effect as Berlyne (1971) found it to be an inverted U-curve relationship to preference of arousel. The two other facets from Moshagen and Thielsch (2010) is colourfulness, as colours have been found to have a strong effect on perceived aesthetics, and craftsmanship which covers the implementation of the design in a good way. In an earlier study done by Lindgaard et al. (2006), in an attempt to find out which design dimension that affects visual appeal, found that 5 out of 7 design characteristics was strongly correlated with visual appeal of websites. These where: interesting-boring, good design-bad design, good colour-bad colour, good layout-bad layout and imaginative-unimaginative. In addition, they had simple-complex and clear-confusing that did not show a strong correlation. Several of these overlap to some degree with the facets of Moshagen and Thielsch (2010) and the dimensions of Lavie and Tractinsky (2004). Interestingly enough the design characteristics of simple-complex and clear-confusing, which can be considered to contain parts of the facets of diversity and simplicity, was not found to affect visual appeal. Lindgaard et al. (2006) isn't offering any explanation to this, but their participants viewed the stimuli material first in two timed phases and then a third time for as long as they wanted when they were asked to rate the web pages in accordance to the design characteristics. The length of time and number of repetitions stimuli is presented to users might impact the weighting of design characteristics, and the selection of pages used in the experiment was selected based on visual appeal and not selected to necessarily represent each design characteristic.

2.3.4 Perception and processing of visual information

If we look at the findings by Fogg et al. (2003) in regards to credibility we see that, while design look was mentioned as most important to their judgement of credibility, the two next categories where: structure of information and information focus. These are not directly connected to aesthetics, but are likely to be affected by it through the way information is structured and the levels of focus it receives through the design of the web page. The psychologist behind the gestalt principals, and others after them, have found that humans tend to perceive objects around as whole and this have implications for the way web pages should be designed (Johnson, 2014). The way we structure information affects, according to Johnson (2014), our ability to scan and understand information. If it is presented in a structured way, with good visual hierarchy and following guidelines of graphic design such as the gestalt principals, it is increases users ease of consuming the information.

The field of interaction design is focused on "creating user experiences that enhance and augment the way people work, communicate, and interact" (Rogers, Sharp & Preece, 2011). This is a large task as user experiences can be quite complex, but one aid used by interaction designers are design principals (Rogers, Sharp & Preece, 2011). They are guidelines to help designers think about aspects of their designs that impacts the users experience. There are several principals, but in general they are concerned about allowing user to easily find information and understanding their options for further action while reducing the risk of the users making mistakes. If we consider the research by Fogg et al (2003) it seems that both the aesthetics (design look) and communication of information is important for credibility. The latter, communicating information and option clearly to users, is heavily focused in several of the design principals. Findability is a design principle that has two levels, which is to the degree an object is easily to discover/locate and to what degree the system supports finding objects (Morville, 2005). Visibility as a principal puts emphasis on making actions/functions/information clearly visible to increase the chance that users will know what to do next (Rogers, Sharp & Preece, 2011). Consistency refers to creating designs that act in a consistent way in terms of its operation so that users can easily remember and expect the outcome of their actions (Rogers, Sharp & Preece, 2011). Consistency when it comes to web pages is often connected to conventions. Conventions is created if enough people use an idea/pattern over time such as having the navigation system as a bar on top of the page (Krug, 2006). This helps users to faster scan the webpage and figure out how the

page work (Krug, 2006). These principals in general puts emphasis on making information and functions clear, natural and as simple for the users as possible. Nielsen (1999) have suggested another principal not exclusively, but with emphasis on web design. Simplicity was suggested as a design principle where Nielsen (1999) want designers to make sure that users get what they are looking for on a site by reducing any obstructions and reducing complexity. Simplicity has reasoning in human nature. It has been accepted that users don't read web pages carefully, but instead scans bits that resemble what they are seeking. However instead of finding the optimal choice, users will settle for the first reasonable option they come over (Krug, 2006). This happens often when looking for information online because the error of choosing incorrectly is low, we might not find a better option and perhaps most importantly it takes both more time and effort (Krug, 2006). We, as humans, have limited capacity of attention and cognitive abilities (Johnson, 2014). Humans are very goal oriented and we look for what can bring us to complete our goals. The more information, options and content that is presented the more time and effort it takes to make a decision. (Loranger, 2015). Since our capacities are limited as mentioned previously we choose the path of least resistance which puts pressure on the system to help users make the right decisions by reducing the mental strain. One way to do that is to follow the principle (Nielsen, 1999) or facet of simplicity (Moshagen and Thielsch, 2010). It is proposed that the easier perceivers, or users of a website, can fluently process an object, the more positive aesthetics impression they will have of the object (Reber, Schwarz & Winkielman, 2004). This suggest that the more fluently a webpage can be processed the more positively it will be perceived. It is easier to fluently processes a simple layout (Moshagen and Thielsch, 2010) which follows features that we judge aesthetics from an objective view such as: balance, symmetry, goodness of form, figure-ground contrast etc. (Reber, Schwarz & Winkielman, 2004). This theory is based on processing fluency which is how easy information can be processed, but Reber, Schwarz & Winkielman (2004) use processing fluency as an umbrella term as they also include perceptual fluency and conceptual fluency in their definition. Simpler designs can, if we follow the approach of Reber, Schwarz & Winkielman (2004), increase the aesthetic impression of web pages. Simplicity has also been found to be highly important in brand names because of two cognitive processed, limited attentional capacity causes humans to value simple information positively because less attention is needed to process it, and secondly when users aren't highly involved in a task

the simpler information is easier to memories (Robertson, 1989). In regards to simplicity as a design principle, Rogers, Sharp & Preece (2011) questions effectiveness of the processes of stripping a page of everything to the bare minimum in regards to aesthetics. Removing to much might have the opposite effect and the site might be perceived as less appealing, instead Rogers, Sharp & Preece (2011) states that "good interaction design involves getting the right balance between aesthetic appeal and the optimal amount and kind of information per page". Karvonen (2000) also raises the question of what degree of simplicity is optimal and thinks the simplicity users wants is a designed simplicity which not only have less elements, but that still is perceived as clean and beautiful.

2.3.5 Complexity

If theory mentioned in the previous section around processing fluency is correct then there is likely to be significant difference of the aesthetic impression of a simple and a complex site. Complexity was seen by Berlyne (1971), together with novelty, surprisingness, ambiguity, and puzzlingness as collative properties that have an influence on arousal experienced by a person. All humans have at any time some level of arousal and changes in arousal is found to potentially cause several psychophysiological changes (Berlyne 1971). Berlyne (1971) theorized that properties of the object, mentioned earlier, affects the aesthetic perception through either increasing or decreasing the arousal experienced from the stimuli.

In 1960 Berlyne (1960) defined complexity loosely as determined by the variety or diversity in a stimulus. Berlyne refered to stimuli patterns, but his definition for complexity should be as actual today as it was back then. Complexity is, according to Berlyne (1960), affected by the three following properties:

1. The number of distinguishable elements. If everything else is the same, then complexity will be increased as one adds more distinguishable elements.
2. Dissimilarity between elements. Complexity can, if the number of elements is the same, be increased by having dissimilarity between elements while it can be decreased by creating similarity.
3. The perception of separate elements responding as a unit. If elements are perceived as one group, then complexity is reduced compared to them being perceived as separate elements.

These are three properties that affect or make up complexity in a stimulus, and the third point is particular interesting as it partly refers to the Gestalt principals of visual perception, and he (Berlyne, 1960) specifically mentions proximity in regards to the third point. The Gestalt principals builds on human's tendency to see things in our surroundings as whole object such as when they are placed closely together (proximity) or look similar (similarity)(Johnson, 2014), which seem natural to be influencing the perceived complexity. Berlyne (1960) viewed the perception of complexity as being partly made up of physical properties and partly based on habit structures that varies from individual too individual. In other words, there might also be some individual differences in perception of the degree of complexity, but still similarities through the physical properties.

Complexity, being one of those collative properties affecting arousal, are quite interesting if we consider simplicity and processing fluency which advocates as little complexity as possible. Berlyne (1971) theory however suggest that moderate levels of arousal are optimal as he found there to be an inverted U-curve relationship between arousal and preference. Too high arousal gives displeasure and too low doesn't give the right amount of pleasure. Complexity is believed to affect this in a similar pattern with a moderate level of complexity creating the most positive impression by reaching optimal pleasure (Berlyne, 1971). These findings have been supported by Geissler, Zinkhan & Watson (2006) in regards to levels of complexity, who found that complexity on home pages influenced users in terms of attention, attitudes and purchase intent. The amount of complexity they found to be most effective was a moderate amount which supports the inverted u-relationship suggested by Berlyne (1971). They did also find that elements, in addition to the number of elements, that affected the perceived complexity was the number of screens, graphics and links (Geissler, Zinkhan & Watson, 2006). In another experiment investigating the effects of complexity, Pandir and Knight (2006) found that their subjects significantly agreed on what stimuli that was perceived as complex and which were not. In addition, they found correlations between high scores for pleasure and low/medium levels of complexity. It was however not found any data to support Berlyne`s inverted u-relationship between arousal and preference (Pandir and Knight, 2006). Interesting enough, the words that their test participants used to describe the most pleasing websites included balanced, clear, stylish and minimalist. The least pleasing websites were described with words such as stressful, complex, overloaded and aggressive. This support to a certain

degree the thought that simplicity and classical aesthetics as dimensions of aesthetics might be important to create aesthetically pleasing websites. As well as complexity being an important property to our perception of aesthetics that to some extent is considered objectively (Pandir and Knight 2006) or at least that we subjectively have a general agreement on.

2.3.6 Logo and navigation

A logo is considered a key component of a company`s brand identity and are considered an important element for recognition, and to evoke emotions connected to the brand (Pittard, Ewing & Jevons, 2007). Using a logo have also become a common element in web design and it can be considered a convention to have it placed in the top left corner of the page as it's the first element users see (Whitenton, 2016). This is partly because of normal scanning patterns on the World Wide Web which in western country`s is from left to right. But we don't always scan or read all the way to the right as its been discovered that users spend 80 percent of their time looking on the left side of the page (Fessenden, 2017). This might explain partly why the placement of a logo has been found to have strong impact on both navigation, as its often used as a navigation itself, and brand recall. According to Whitenton (2016) companies that have their logo in the left top corner are 89 percent more likely to remember the brand of the page after their visit then if its placed on the top right corner. This might have an impact on credibility as brand recognition is an aspect of credibility, and Fogg et al. (2003) found that 14,1% of comments of how users judged credibility on websites were in the category *Name Recognition and Reputation* and 8,8% were in *Identity of Site Operator*.

Logos has also been found to evoke aesthetic responses which may yield positive or negative emotions (Pittard, Ewing & Jevons, 2007), which in turn might influence our perception. In addition to placement, the proportion of the logo (Pittard, Ewing & Jevons, 2007), complexity and exposure time has been found to have a strong influence on preference and recall (Grinsven & Das, 2016). The complexity of a logo, from simple to complex, has been found by Grinsven & Das (2016) to influence how fast a brand is recognized. The faster a logo is seen might also have a positive impact because of processing fluency, and the speed is affected by both placement and complexity. Judgement of surface

credibility might therefore be positively influenced by a simple logo placed after conventions.

Navigation is a key element on websites that is necessary for users to move around on a page. Humans are often goal-oriented and focus a lot of their attention on a task at hand (Johnson, 2014). On the web people are often looking for something specific which makes the navigation highly important, but it also gives users a sense of what else the website has to offer and how to get there (Krug, 2006). Conventions allows users to understand how a webpage works and it can also give a reassuring sense of familiarity (Krug, 2006). Familiarity has positive effects online as users that are familiar with a website`s terminology, content or how its structured will feel less lost and ease the decision making process (Chen et al., 2011). In the study of Fogg et al. (2003) the *structure of information* was the second most (28,5%) mentioned category when users evaluated credibility, and in that category they included among other things how hard it was to navigate the website. This indicates that navigation might have an influence on user's judgement of credibility. Considering that navigation systems that follows conventions are more likely to evoke familiarity and to be easier processed they are more likely to cause a positive effect on credibility.

3.0 Methods

This study aims to investigate the effects that visual aesthetics, more specifically the property of complexity, have on surface credibility judgements of web pages. The following section will describe the methods that have been used to answer the research question and hypothesis presented in chapter 1, including how they have been used and why they have been selected.

3.1 Research design

Research is, at its basic core, the collection, analysing and interpretation of data in a systematic process with the aim of gaining insight into a problem or question of interests (Leedy & Ormrod, 2014). We often use the process of collection, analysing and interpreting data (information) in our everyday lives when we encounter question, but what sets research in this context apart is the formal approach. In order to approach the problem at

hand a general strategy is created, which in the field of research often contains several common elements across research experiments, and that strategy is what's also known as research design (Leedy & Ormrod, 2014). Research design creates the structure of the process the research attempt follows in order to find answers. The choice of research design should be based on the purpose of the research and what questions seeks answers (Lazar, Feng & Hochheiser, 2010). There is however no perfect way to approach a research question as all approaches has some flaws and one effort alone is not enough to produce definitive answers to the question (Lazar, Feng & Hochheiser, 2010).

While research design often follows a common approach there are more differences between the different academic fields when it comes to research methodology. Research methods are used to gain meaning from data gathered in a research experiment, and the methods often depends on the data as one often needs specific methods to gain meaning from specific types of data (Leedy & Ormrod, 2014). In research there are two major categories of data: Quantitative data & Qualitative data (Rogers, Sharp & Preece, 2011). Quantitative data is data in forms of numbers and qualitative data is not usually expressed in numbers, but instead in descriptions, quotes etc. (Rogers, Sharp & Preece, 2011). These two categories of data have divided research into two main approaches which focuses on each type of data: quantitative approach and qualitative approach (Leedy & Ormrod, 2014). Both approaches have similarities in terms of identifying a problem, considering existing literature, collecting and analysing data in order to get closer to an answer to the problem (Leedy & Ormrod, 2014). Both approaches have their strengths and weaknesses, but they are not exclusive and can be used together in a mixed-methods design (Leedy & Ormrod, 2014).

In order to answer the research question and hypotheses this study has used an explanatory research design process. Explanatory research design consists of two phases where the first is quantitative (which will receive the most focus in this study) and the second phase is a qualitative phase (Leedy & Ormrod, 2014). The reason for choosing a mixed-method approach like this is that looking into aesthetics and the effects on credibility is a complex topic. Quantitative methods have their strengths in establishing or confirming relationships between variables and testing theories (Leedy & Ormrod, 2014). These methods often use bigger samples of test participants and are considered to be effective at creating findings that can be generalised to a bigger population (Lazar, Feng & Hochheiser,

2010). Qualitative methods have their strengths in gathering data that often isn't generalizable, but helps to describe and explain complex situations (Leedy & Ormrod, 2014). Mixing both quantitative and qualitative methods have several strengths. One is that each method may compensate for weaknesses in the other method and thereby strengthen the overall research (Leedy & Ormrod, 2014). An example of this is that quantitative methods find data in numbers, but combining those numbers with qualitative methods can give greater substance and meaning to the numbers creating a more nuanced result than one method alone can (Leedy & Ormrod, 2014). Other strengths of mixed-methods approaches are that it can allow research to fully address a research problem, and qualitative data can also help the researcher understand possible inconsistencies or contradictions that might appear in quantitative data (Leedy & Ormrod, 2014). This research study looks to see if certain aesthetic variables, such as complexity, navigation and logo, have effect on surface credibility judgements. Using quantitative methods this study will be able to investigate the relationship between credibility and complexity, navigation and logo, and using qualitative methods will allow the study to attempt to give more insight into the data collected in Experiment 1.

3.2 Experiment One

In order to answer the first research question which focuses on how differences of visual complexity affect the judgement of surface credibility made in short exposures of webpages, and second research question which visual elements users view when making their judgement of surface credibility, a research experiment was planned and conducted. The research experiment was set up as a within-group design which exposes participants to multiple variables and thereby removes the need for several groups of participants (Lazar, Feng & Hochheiser, 2010). This type of experiment design also has a positive advantage as it reduces the impact of individual differences, but has some disadvantages such as possible learning effects and fatigue (Lazar, Feng & Hochheiser, 2010). Some choices have been made in order to reduce the likelihood of the two advantages and can be read about in the sections below. The following section describes the choices that were made for the experiment in regards to operational definitions, methods, stimuli material, participants, procedure and experiment set up. The experiment guide (Attachment A) and stimuli material (Attachment F) used for the experiment can be found in the appendix.

3.2.1 Operational definitions

In order to conduct an experiment to investigate research question one and two some operational definitions had to be made. Operation definition are how a researcher has defined characteristics and variables in a study which is considered useful if the elements being defined doesn't have a clear or globally accepted definition (Leedy & Ormrod, 2014). The following section describes the operational definitions that have been used in the study.

Complexity is, according to Berlyne (1960), determined by the variation/diversity in a stimulus which is affected by the number of elements, dissimilarity between elements and how elements is perceived as a whole. Pandir and Knight (2006) found, as previously mentioned, that their subject agreed on the level of complexity in a stimulus when they created three levels of complexity based on the number of elements on a web page. To investigate how differences in complexity affects credibility judgements this study also divides complexity into three levels: low, medium and high. This division was based on the number of elements on a web page as seen in the table below:

Levels:	Low	Medium	High
Number of elements	<=10	11-19	>=20

Table 1: Number of elements determining each level of complexity

All distinctive elements where considered as one element. Exception to this was made in regards to the website navigation if the navigation follows gestalts principals of proximity and similarity so that it is considered to be one element, and also text paragraphs where considered to be one element (not counting each word).

Credibility also needed a definition in order to be measured and analysed. It has been widely accepted that credibility is considered to be the sum of perceived expertise and perceived trustworthiness (Fogg, 2003). This study accepts this definition and measure both in order to create a score of the perception of credibility.

Two specific visual elements, navigation and logo, were selected in order to measure how much they were seen by users and if this affected credibility judgements. Two operational definitions for each was made: Conventional logo, unconventional logo, conventional navigation and unconventional navigation. A selection of stimuli material, 4 stimuli for each category, where selected based on how they fit in the category. Conventional logo and navigation is in this study defined as navigation/logo that is placed

according to and look visually similar to conventions online on web pages. Conventions might have a positive impact on credibility if we follow the fluency processing theory by Reber, Schwarz & Winkielman (2004) which states that the more fluently user can process a stimulus, the more positive it will be perceived. Conventions allows users to recognize and understand how a webpage works faster as it follows similar patterns as they are used too (Krug, 2006). Unconventional navigation and logo is defined as logo/navigation that either are non-existing or differ from conventions in terms of placement or look. One exception is made for the hamburger navigation which can be considered a convention, but which grant the users very little information and is therefore here considered unconventional.

3.2.2 Methods and tools

The following section describes methods and tools that is used in the experiment in order to gather data to be analysed for answer to research question one and two.

In order to measure how different levels of complexity affects the perceptions of credibility this study uses rating scales. Rating scales were created to capture the perception of both perceived expertise and trustworthiness which in analysis is added together to create a score for the perception of credibility for each web page. In addition to expertise and trustworthiness, this study also attempts to measure the attractiveness of each web page. Attractiveness have been found to be correlated to credibility judgements (Fogg, 2003; Robins & Holms, 2008; Pornpitakpan, 2004) and can to some degree be used as a control for the ratings of expertise and trustworthiness. Rating scales are considered to be good at capturing attitudes, opinions or when something needs to be evaluated on a continuum (Leedy & Ormrod, 2014). Originally nine statement, 3 for each variable, on a 7-point Likert scale were created to measure the participant's perceptions. This was reduced to 1 for each variable after pilot testing due to concern of fatigue effect among the test participants. The following three statements were presented to the participants:

On a scale from 1 to 7, where 1 represents strongly disagree and 7 represents strongly agree, to what degree do you agree with the following statements:

1. *The website appears professional*
2. *The website is visually appealing*
3. *The website appears to be trustworthy*

These statements where read to the participants and they were asked to give their position on the statements which were manually recorded by the moderator.

In addition to the method above measuring the perception of credibility there is also the need of a method to limit the time participants got to see the web pages as the purpose was to investigate the effects in a short exposure, and a method that could measure how the participants viewed the web pages. The tools selected for this was eye tracking solutions from SMI. Eye tracking has given new opportunities in the field of human-computer interaction as it allows researchers to gain insight into where users attention is focused on a screen and not only where they move a cursor. The eye tracking tool allowed the study to limit the time the participants got to see the web page to 1500ms to increase the likeliness of capturing the effect of a first-impression. Previous research done by Robins & Holmes (2008) found that participants used an average time of 3.2 seconds when judging credibility, but Lindgaard et al. (2006) found that attractiveness could be judged in as little 50ms. Based on the pilot studies the time in this experiment seemed sufficient to make judgements on the statements presented previously. The main purpose of using eye tracking is to be able to answer research question two. Using AOI (Area of Interest) points on navigation elements and logo the study will be able to capture how much time the participants fixate on those elements, how many times they visit and in which sequence in their scanning pattern.

3.2.3 Stimuli material

This study needed a number of web pages with varying degree of complexity, and variation within navigation and logo elements. Each webpage should be represented by its homepage. The homepage is the main portal of the website to the rest of the content internally on a site and is the most likely starting place for the majority of users. Websites homepages has been both argued to and found to be important to convince users to stay on the page (Geissler, Zinkhan & Watson, 2001; Pandir & Knight, 2006). In addition, it was decided that the web pages should be split into two categories in order to see if there were differences in judgement based on category. The two selected categories were *business* and *information*. These categories were chosen because they are common and it is evident that the amount we are shopping and consuming information online is increasing (SSB, 2017b). It is also two categories of web pages where credibility is likely to play an important part when users

decide to stay or not (Fogg et al., 2003). In total 24 homepages were selected, 4 in each level of complexity and 12 in each category of web pages as in the table below:

	Low	Medium	High
Information	4	4	4
Business	4	4	4

Table 2: Number of stimuli divided for each degree of complexity and types of websites.

The webpages were selected from existing pages online through google search and webpages such as SiteInspire (https://www.siteinspire.com/). Existing pages were selected in order to increase the content validity, as the variance in the different pages we meet online is large and in order for this experiment to find an effect that is likely to be present outside an experiment setting this was deemed necessary. The web pages were selected by the researcher and attempts were made to find homepages of somewhat equal level of design quality, although there is no objective characteristic to judge the design quality of each web page on – so a subjective judgement were made. In addition, the web pages were not selected if they were likely to be used by the demographic test-population in order to limit any recognition effects when seeing a web page that the user has seen before, which will reduce the impact of a first-impression. Therefore, no Norwegian pages were selected. No webpages were edited with the expectation that all the url`s were removed as it has been found to have an effect on credibility judgements (Fogg, 2003). The level of complexity was judged based on the number of elements as presented earlier. A screenshot was made of each page to be used in the experiment as a still image, each image was shown in 1366x768 resolution. The resolution is the most common screen resolution (W3Schools, 2017).

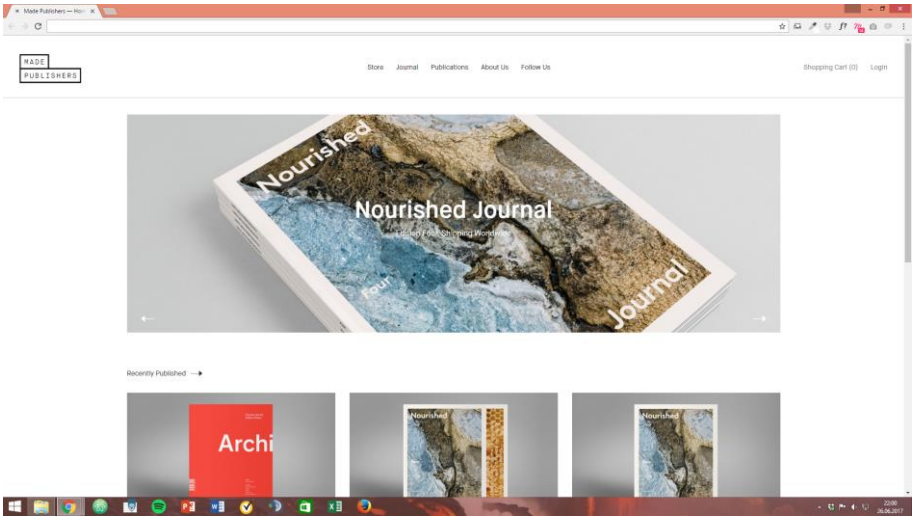


Figure 1: Example of one of the websites which is considered to be of low complexity.

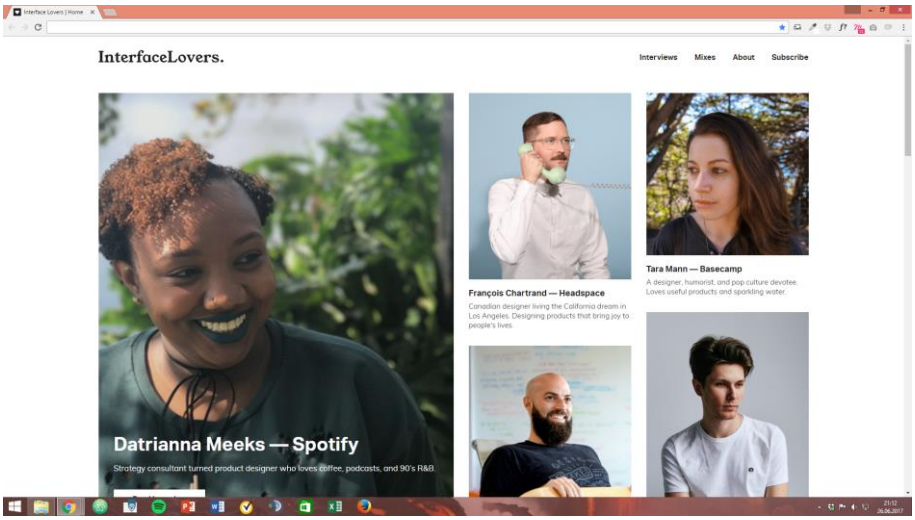


Figure 2: Example of one of the websites which is considered to be of medium complexity.

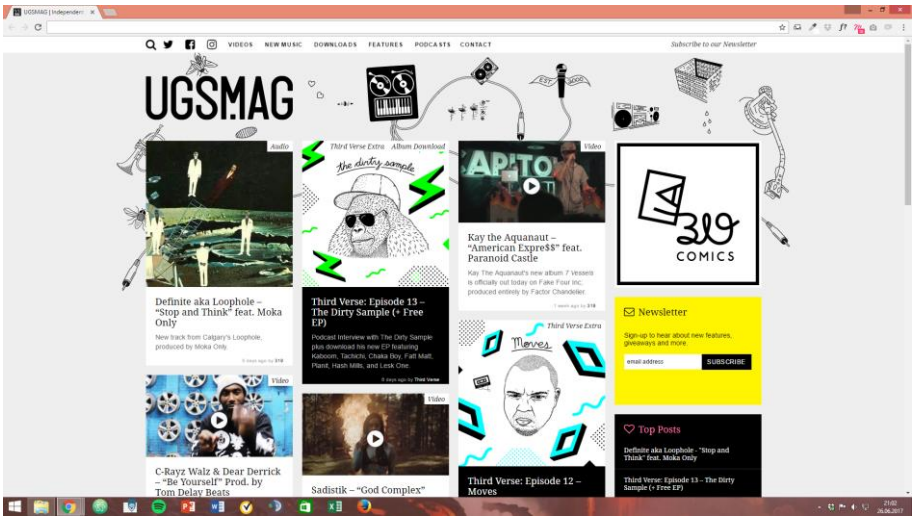


Figure 3: Example of one of the websites which is considered to be of high complexity.

Participants of the experiment was divided into two groups and the sequence of stimuli material where randomized differently for each group. Each set was given a randomized sequence by using a generator (https://www.random.org/sequences/). Each participant did not get their own individualized randomization because noting down answers during the experiment would be error-prone and time-consuming. In the start of each set of stimuli material presented there where added four extra websites which each participants viewed and responded too in the same procedure as in the rest of the set. This was done in order to reduce the likelihood that learning effect will have an impact on the study. Letting users getting used to a system, procedures or tasks will according to Lazar, Feng & Hochheiser (2010) reduce the learning effect during the actual tasks that is measured.

3.2.4 Participants

In total 35 participants took part in this study. The result of 4 participants were removed due to lack of understanding the task. This left the study with the result from 31 participants (11 of which were female and 20 which were male) which is a small number for an quantitative study, but because of the within-group design of the experiment the sample size required is smaller. This is because all participants judge the same material and are exposed to all the variables giving a decent sample size of judgements to do statistical analysis on. The participants were selected at convenience and was recruited from NTNU and a major company (more than 300 employees). While the background of the participants is varied, this study used a convenience sampling method which does makes it unlikely that the pool of participants is a representative sample of the population (Leedy & Ormrod, 2014). The majority of the participants were within the age span of 20-30 (21 participants) and the majority of the 31 participants were of Norwegian nationality (26 participants).

3.2.5 Procedure

The experiment set up consisted of a computer with eye-tracking equipment, a table were the computer was situated and a chair for the participant. Before each participant entered the experiment room they were randomly assigned to one of two experiments set. Once a participant entered the room they were was asked to sit approximately 40 cm away from screen in accordance to the eye tracking sensor. Another chair for the moderator was placed on an approximately 45-degree angle on the side of the participants according to Rubin &

Chisnell (2008). They were then read out load an introduction text (experiment guide in attachment 1), and asked to read and sign an informed consent (attachment 2). The experiment followed the same loop of procedures for each stimulus for the duration of the experiment. The procedure was as following:

1. A cross was shown in the middle of an otherwise blank screen. The participant had to focus his or her gaze there for 2 seconds in order for the next imagine to show up.
2. After step 1 the stimuli appeared for 1500ms before disappearing.
3. A blank screen appeared. The moderated presented the statements presented earlier, one at the time, and the participant was asked to state to which degree they agreed on the statements on a scale from 1-7. Moderator noted each answer in a form (Attachment B).
4. The participant was asked to continue by pressing space which started the process from step 1 again with new stimuli until the participant had been through 28 (The first four is where not included in the experiment as they were used to reduce potential learning effect).

After they had finished the experiment participants were given a short debrief, and the option to ask question about the experiment and purpose.

3.2.6 Pilot study

Two pilot studies of the experiment was conducted in order to check the viability of the method selected. This is recommended to do (Leedy & Ormrod, 2014; Rogers, Sharp & Preece, 2011; Rubin & Chisnell, 2008) as it gives the opportunity to check how the equipment, questions, procedures etc. work in order to reduce risk of the experiment being wrongly designed, and check that it is possible to gather the right data.

3.3 Interview and experiment

The third research question which aims to if participant's physiological judgement of complexity based on a short exposure changes to a cognitive judgement based on a longer exposure – indicating a confirmation bias of the first impression. In order to answer this question, and give additional insights into research question one and two, an interview and

small experiment was conducted. The interview/experiment guide can be found in the appendix (Attachment D). The interview evolves around stimuli material from the main experiment and how the participants perceives it. The following section describes the choices that were made in regards to methods, stimuli material, participants and procedure.

3.3.1 Methods and tools

The following section describes methods and tools used in the interview in order to gather data to build upon for research question three, and research questions one and two.

Interview as a method to collect data has its strengths when it comes to explore and interpret (Leedy & Ormrod, 2014) as it gives good opportunities to understanding people and their preferences, attitudes, concerns etc. This study selected to use a semi-structured interview form which consists of open and closed questions, and an interview script that makes sure all topics is covered (Rogers, Sharp & Preece, 2011). This allows the study to ask the same questions across all the participants, but still allow open questions which allows the users to answer as they see fit and gives the possibility to ask for clarification or follow up with additional questions (Lazar, Feng & Hochheiser, 2010).

In the start of the interview, after a set of general questions around background and internet usage, a small experiment was conducted using the same rating scales as in the main experiment and 9 of the stimuli material. The first three web pages was added in order to allow users to get used to the process and thereby reduce the risk of learning effect. The participants were asked to perform the same task as in the main experiments, but this time twice. First time they viewed the material for 1500ms, and the second time they were allowed to view it for 10 seconds.

3.3.2 Stimuli material

Stimuli material was used for multiple purposes in the interview. First it was used to measure perceived credibility of the presented material to see if their judgement were consistent with those of the main experiment and if they changed or not, based on exposure time. The second purpose was to ask question about the selected web pages themselves. Prints of the stimuli was used as props to help users remember what they looked at and make it easier to discuss the web page. Using props can enrich the experience of an interview by giving more context and allowing users to give fuller answers (Rogers, Sharp &

Preece, 2011). The selection of stimuli material consisted of nine homepages from the main experiment. Three were selected and shown to get participants used to the structure and process of the experiment at the start of the interview. They were only rated and not discussed in detail. The six main homepages that were discussed in detail were selected from each category of page and each level of complexity.

	Low	Medium	High
Information	1	1	1
Business	1	1	1

Table 3: Number of stimuli divided for each degree of complexity and types of websites – for the interviews.

The pages were selected based on which received the lowest variance score based on judgements in experiment one within each category and level of complexity. This was done in order to select stimuli material which was in general viewed as quite stable across all participants, and was therefore more likely to be producing clearer results also in the interviews – possible allowing a pattern to appear clearer. The sequence which the stimuli material appeared to the users was randomized using a random number generator for both the first set (1500ms) and the second set (10000ms). The sequence was however not randomized individually for each participant.

3.3.3 Participants

In total 6 participants took part in the interview. Convenience sampling was also used this time to select participants. The majority of the participants were within the age span of 20-30 (5 of 6 participants) and the entire selection were of Norwegian nationality.

3.3.4 Procedure

Participants were greeted, given an introduction, and asked to read and sign an informed consent. A set of general question was asked with two-sided purpose. One is to let the user get warm and used to answering questions. Second is to gather some background information about their experience around design and usage on the internet.

Next they were asked, as in the main experiment, to view a homepage for 1500ms and then answer to what degree they agreed with three statements. Read the section about

the main experiment to see the statements and the exact steps as the same procedure was used on this part.

The next step worked the exact same way as above with a different sequence of the stimuli material, but this time they were given 10 seconds to study each homepage before they were given the three statements.

The last step involved giving the user a print out of one homepage and ask a set of questions around their perception of the web page. This also included checking if their judgement of the material had changed based on the exposure time, with more than one step, and in that case ask the additional questions around why they felt different the second time. Questions are repeated for each stimuli material.

3.4 Ethical considerations

With any research study there is important to take into account possible ethical issues the study might bring. It is suggested by Leedy & Ormrod (2014) that most ethical issues "fall into one of four categories: protection from harm, voluntary and informed participation, right to privacy, and honesty with professional colleagues". The amount of potential issues does depend on the purpose and way a study is made, and potential issues can also be reduced by making ethical considerations.

The purpose of the study and the methods used do not have any particular strong ethical issues connected to them. The only aspect is that credibility is an important element of persuasion and persuasion can be unethical depending on the purpose behind (Fogg, 2003). The knowledge from this study can be used to increase persuasion of web sites, unrelated to the purpose, which can be used on unethical websites. The knowledge from this study is however not alone enough to cause harm, and it is available for everyone which can help users as well as content providers. It is not deemed to be a strong ethical issue of the study.

This study did not put any participants in risk of harm. The material used in the study did not contain any sensitive content that are likely to provoke or cause distress for the participants. A debrief and potential for questions from participants were also available in order to alleviate any uncomfortable reactions or uncertainties that might have appeared during the study.

Informed consents are a common practise and often required when working with human subjects (Leedy & Ormrod, 2014). The purpose of informed consent is two-folded were the first is to secure that users are well informed about the study in terms of its purpose, procedures, potential risks and where the participants can get more information about it (Lazar, Feng & Hochheiser, 2010). The second part is to make sure that the participants are aware and are making a voluntary decision to partake in the study, and that they can without any consequences withdraw from the study (Lazar, Feng & Hochheiser, 2010). This study created informed consent forms both for experiment one (Attachment C) and one for the interviews (Attachment E).

The privacy of the participant is also taken into account as no name are collected and each participant are kept anonymous with their data only connected to a participant number. No sensitive data were collected and raw information will only be kept for the duration of the study before its deleted. This study did not use any video or sound recording which presents several challenges regarding privacy (Lazar, Feng & Hochheiser, 2010).

4.0 Results

Two experiments, an eye tracking experiment and an interview session, have been conducted in this study to find answers to the research questions in section 1. Through data gathering a lot of raw data is available, but the data itself is meaningless unless they are analysed and interpreted (Leedy & Ormrod, 2014). That is true regardless of method of data gathering. There are generally different strategies to analyse quantitative data and qualitative data, but combinations of the two can also be made (Rogers, Sharp & Preece, 2011). The method of analysis that can be selected is however dependent on the kind of data that is gathered, and the goals of the study (Rogers, Sharp & Preece, 2011). After data is collected and before its analysed it often has to be prepared as there might be errors, need more coding or need to be structured differently to be analysed through a specific method or software (Lazar, Feng & Hochheiser, 2010).

In order to analyse data from this study the data from Experiment One was structured in an excel file. The data from both test groups where added to one file and information on the different variables such as category of websites, information & business, were coded into the document as numerical values in order for statistical analyses to be

possible. The same was done for degree of complexity, type of logo, type of navigation and gender. In addition, each two unique set of keys(ID) were made, one for each participant and one for each stimuli. This allows the study to connect data to specific stimuli, and analyse within-group effects of the research questions. During the preparation of the data the result of the four participants were removed due to errors in term of not understanding the task properly.

The interviews also had to be analysed, as with quantitative data we also looked for patterns in qualitative data (Rogers, Sharp & Preece, 2011). The goal of conducting the interviews was to investigate possible bias effects of the first-impression and to get more insight into what people think of the webpages to get a nuanced look into why they are rated as they were.

The following sections describes how the experiment and the interviews have been analysed and the results from them.

4.1 Experiment One

Due to the quantitative approach to Experiment One the data collected is numerical and therefore statistical analyses is the best way to analyse the data. Statistical analyses allow the study to investigate the relationship between variables by statistically comparing variables (Lazar, Feng & Hochheiser, 2010). Experiment One resulted in data from 31 participants who responded to 24 stimuli materials which gives 744 isolated judgements in total. Analyses were run for these isolated judgements, but due to the within-group design of the experiment it is most natural to analyse the data for each participant and not for each stimulus. Therefore, the result here covers the analyses for a within-group design which results in fewer judgements, the number depends on the variables involved in the different analyses. Within-group design does however isolate individual differences better and one can according to Lazar, Feng & Hochheiser (2010) find the expected result with a lower number of participants as we compare performances of each participant under different conditions. An example is that since this study has three different levels of complexity each participant has been exposed to three conditions giving a sample of 93 when comparing complexity to credibility. This allows the study to still analyse quantitative data and find results. Both rounds of analyses found similar results.

In order to analyse the data, it was first structured in excel before it was taken into SPSS, a statistical analysis program, where some additional structuring of data was done and analysis were conducted. The following sections presents results from the analysis of data from the experiment.

4.1.1 Complexity & Credibility

In order to investigate the relationship between differences in complexity and its effect on credibility rating, scales were used to measure credibility and attractiveness. All the stimuli material were divided into three levels of complexity. This allowed the study to measure credibility judgements for each stimuli by combining scores for perceived expertise and trustworthiness which together equals perceived credibility (Fogg, 2003). The credibility judgement collected for each stimuli allowed us to make three variables for each participant, which consisted of the mean of their individual ratings of all the stimuli in each level of complexity. This left the study with a sample size of 93 (31 for each level of complexity). Using mean is a common way to measure central tendency in a dataset and together with statistical significance test they can show differences between groups of data (Lazar, Feng & Hochheiser, 2010). Statistical significance test checks if the probability of the observed differences in mean, in this case, between the groups appear due to chance (Lazar, Feng & Hochheiser, 2010). In table 4 one can see the mean for each of the three levels, each consisting of 31 means, and the standard deviation. There are several ways of doing statistical tests to compare mean values between groups, but since we have more than two conditions for complexity in this study then a repeated measures ANOVA was conducted.

Complexity	Mean	N	Std. Deviation
Low	9,8387	31	1,23246
Medium	9,4516	31	1,22376
High	7,7177	31	1,62017

Table 4: Estimated mean value for each level of complexity including standard deviation.

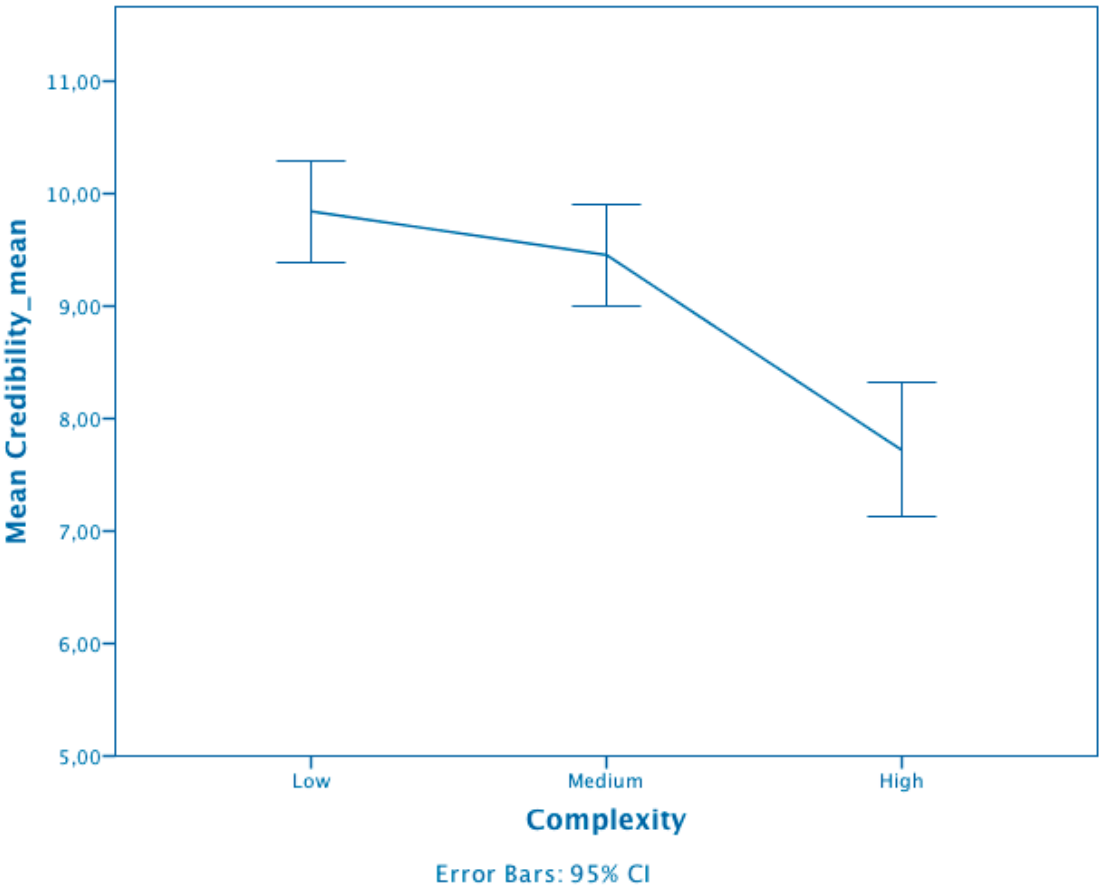


Figure 4: Estimated mean values for credibility based on low, medium and high degree of complexity shows a linear relationship between credibility and complexity.

The repeated measures ANOVA test (illustrated in figure 2) showed clear significance ($p <$ .000, $F(2, 90) = 21{,}031$) between the three levels of complexity indicating that differences in complexity is a strong influencer of user's perceptions of credibility. This results indicates that high degree of complexity, 20 or more elements on a web page, has a negative effect on perception of credibility in web pages based on a short exposure. Low levels of complexity, 10 or less elements on a web page, has the most positive effect on perceived credibility. From the estimated mean values for each level of complexity in table 4 and illustrated by figure 4 we can see that there is a clear difference between high and medium/low levels of complexity. Between low (9,8387) and medium (9,4516) levels of complexity it is only 0,4 in difference in mean scores of perceived credibility, despite some overlap it was still found to be significant in the ANOVA test. This indicates that there is a linear relationship between the degree of complexity and perception of credibility, in a limited exposure time(1500ms) of webpages. The linear relationship is also supported from the fact that 30 of 31 participants judged, on average, the group of stimuli with low degree of credibility more

positively than the group with high degree of complexity. The one participant that judge high degree of complexity favourable did so with as small margin as -0,75 (the data is the result of subtracting the sum of each participant judgement of high degree of complexity to their degree of low complexity).

4.1.2 Credibility & attractiveness

There has previously been found that attractiveness affects perceptions of credibility (Fogg, 2003; Robins & Holms, 2008). In this study the perception of attractiveness was also measured as a way to control user's perceptions of credibility. The perception of credibility and attractiveness should according to theory be correlated. For there to be a correlation between different conditions in an experiment there has to be a significant relationship between them which is commonly tested by using Pearson`s product moment correlation coefficient test (Lazar, Feng & Hochheiser, 2010).

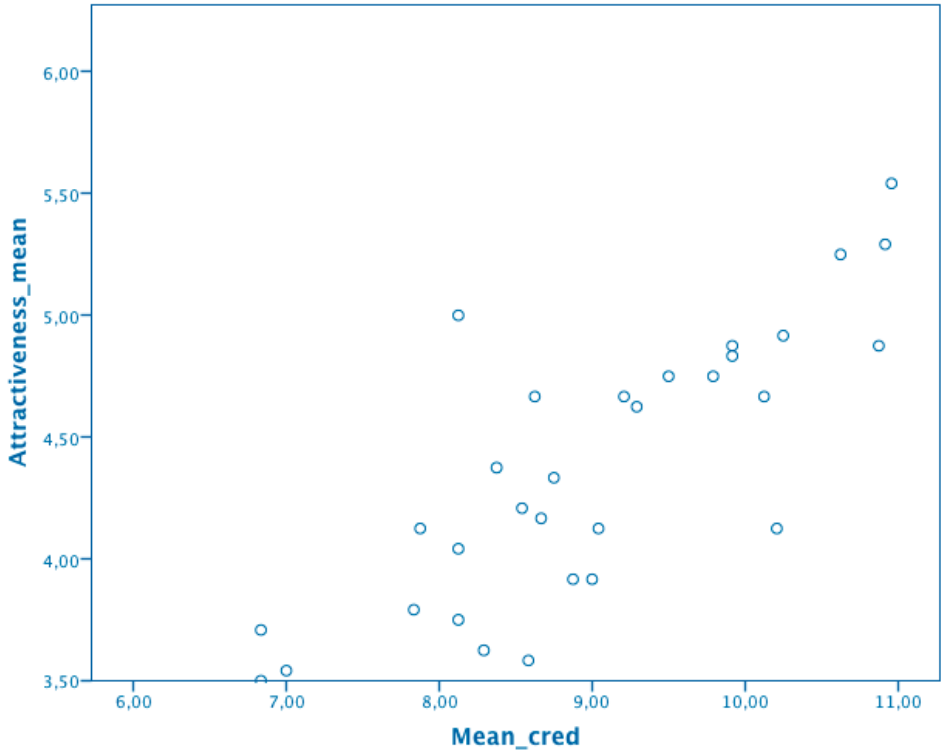


Figure 5: Correlation between Attractiveness and credibility was found to be significant (p < .000).

The result of the Pearson`s product moment correlation coefficient test showed a clear significance relationship ($p < .000$, $r = .803$ ) between the conditions: attractiveness and credibility. This was true not only for the Pearson`s correlation test ($r = .803$), but also Spearman`s rho ($r_S = .767$) and Kendall`s tau ($T_b = .627$). As figure. 5 illustrates there are, just

as with credibility and complexity, a linear relationship between the two conditions which indicates that the judgements of credibility are consistent based on the theory of correlation between attractiveness and credibility. If a website is perceived as attractive it is also likely to be perceived as credible (Robins & Holms, 2008). In Figure 6 below we can also see that, just like credibility, the perception of attractiveness is also affected in a linear relationship by the degree of complexity.

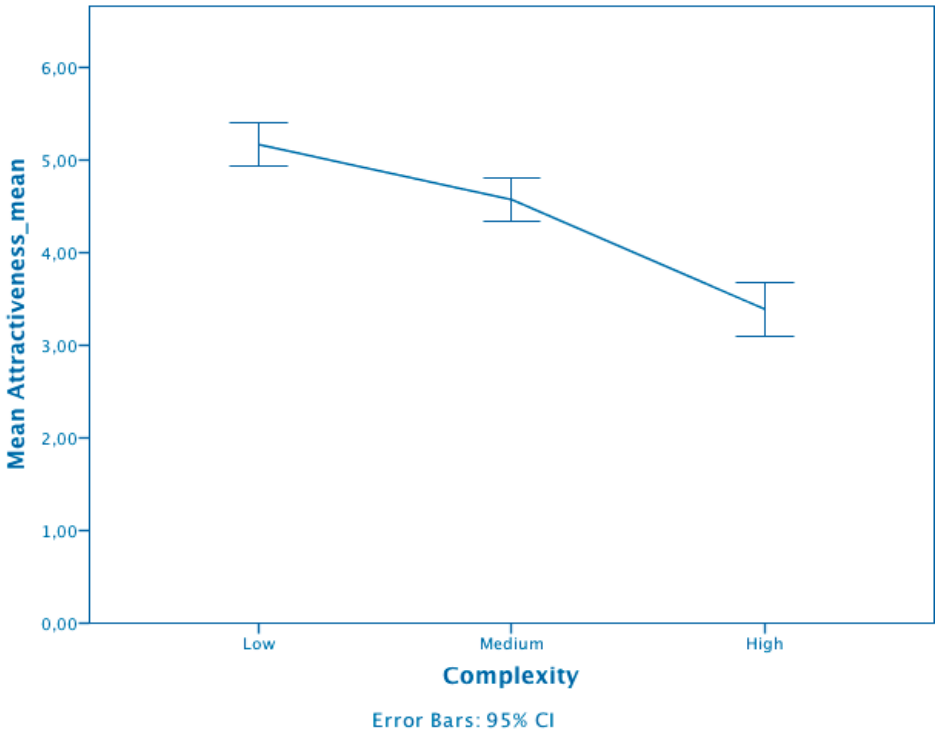


Figure 6: There is a clear and signifcant linear relationship between attractivness av degree of complexity.

4.1.3 Credibility in relation to business and information websites

All home pages that were selected for this study were divided into two categories: business or information. As part of the analyses for experiment one we also investigated if the two categories, in terms of the measurement of credibility and attractiveness, are affected differently by complexity. The mean of attractiveness and credibility were calculated for each level of complexity. This is based on the mean from the 31 participant's judgements for each group of homepages. An ANOVA test was done and the result showed that differences in perceived attractiveness for both business and information was found to be significant (Business: $p < .000$, $F(2, 90) = 69{,}447$ / Information: $p < .000$, $F(2, 90) = 23{,}863$). Indicating that differences of complexity is affecting perceived attractiveness across both types of web pages presented in the experiment. The difference in perceived credibility was found to be

significant for business website ($p < .000$, $F(2, 90) = 53{,}593$), but was not found to be significant for information websites ($p < .221$, $F(2, 90) = 1{,}534$). The result indicates that the degree of complexity cause bigger differences on business websites than information websites. Table. 5 shows the differences in estimated mean value for attractiveness and credibility between the two categories for each level of complexity. The estimated mean value of credibility for business website with high degree of complexity is considerably lower (6,3871) than for information websites with the same degree of complexity (9,0484). Although both shows a linear relationship, based on the mean for credibility across the levels of complexity, it is only business that is considered significant.

	Business		Information	
Complexity	Attractiveness	Credibility	Attractiveness	Credibility
Low	5,1532	10,0000	5,1774	9,6774
Medium	4,7097	9,7177	4,4274	9,1855
High	2,8710	6,3871	3,8952	9,0484

Table 5: Estimated mean values for credibility & attractiveness for the two categories of web pages in the study.

4.1.4 Logo & Navigation

In order to investigate how differences in navigation and logo impacts perceived surface credibility based on short exposure, the navigation and logo of a 10 of the selected stimuli were divided into two groups. Unconventional and conventional for navigation, and unclear and clear for logo. This was done in order to see if following or breaking with conventions for those elements would influence the judgement of credibility. Additionally it was used eye tracking to measure dwell time, entry time and fixations among others to see if they impacted the attention of the participant differently during the short exposure. The data were collected by using AOI´s for the elements in the selected stimulus and exported from the eye tracking software in order to be analysed together with the existing data in SPSS. The sample was however smaller than for complexity. The result of the analysis indicates that navigation or logo, regardless of following conventions, is affecting the perception of credibility made in a short exposure.

Looking at the entry time for logos it showed that unclear logos was seen earlier, 597ms compared to 651ms, than clear logos. It was however seen fewer times. The difference in dwell time was however large, where clear logos following conventions, was in

average viewed for 205ms while unclear logos where viewed for 91ms. In an independent t-test, that was used to compare the mean dwell time between the two groups, this difference was found to be significant ($p < .000$, $t(308) = 5,096$). However, when an independent t-test was run for the expertise score compared for the unclear and clear logos it did not find any significant results. It also didn't come back as significant for credibility, attractiveness or trustworthiness. These results indicate that the logo had little to no effect on the participant's judgement of credibility, or specifically expertise as hypothesised, of webpages based on a short exposure.

In terms of conventional and unconventional navigation it was found no significant difference in mean between trustworthiness, expertise or credibility. Indicating that navigation as well did not have an impact on the participant's judgement of credibility in a short exposure. It did come up as significant ($p > .000$, $t(60) = -3,806$) for attractiveness, with a difference in mean between the groups of 0.4 in favour of unconventional navigation.

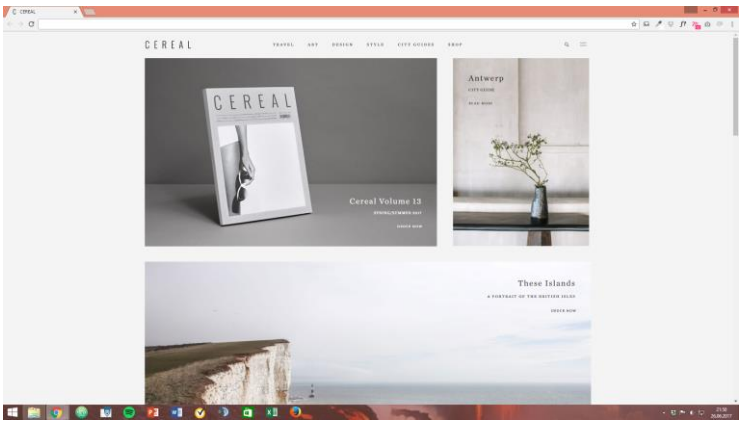


Figure 7: Example of one website with a clear logo place on the top left part of the website after conventions.

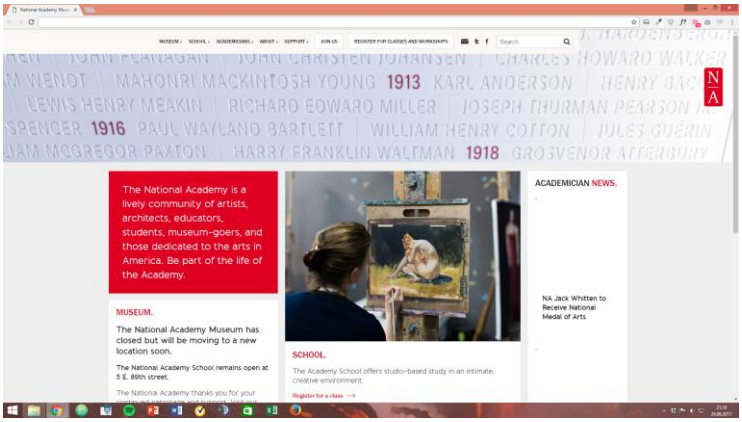


Figure 8: Example of one website with a unclear logo placed on the left side, slightly below the header.

4.2 Interviews

The approach to the interviews had two parts, one where the participants used rating scales as in the main experiment and the second contained questions about six specific home pages. In total six participants took part in the interviews, which due to the low number did not produce enough data to make any proper statistical analysis on their judgements of credibility and attractiveness of the six web pages. The mean of each participant's judgement on credibility of the web pages sorted by complexity and exposure time was still calculated and can be seen in table 6.

	Low	Medium	High
Short exposure	10,91	9,91	6,25
Long exposure	10,00	11,00	6,75

Table 6: Mean credibility based on short (1500ms) and long (10 000ms) exposure.

The difference, or lack of it, in mean credibility scores between short and long exposure was as expected not enough to produce any significant result in statistical analyses. This gives us no statistical indication to how judgements of credibility change from a physiological judgement based on a short exposure to a cognitive judgement based on a longer exposure.

Instead, the focus of the interview was mainly on content analysis and trying to look for patterns and insights through the verbal feedback to questions in the interview. The interviews were first structured together in an excel sheet. In order to find patterns in the peoples answers an affinity diagram was made, a common technique used in qualitative analysis (Rogers, Sharp & Preece, 2011). Comments the participants made during the interview about each web page were noted on post-it notes and marks were made for which complexity the specific website and if the comment was of a positive or negative nature. After all comments had been written done they were sorted into groups based on the nature of the comments. The groups were considered to be: *Structure of Information, Information Focus, Craftsmanship, Navigation, Colorfullness, Diversity, Multimedia* & *Advertising.* In addition, there were 9 comments made that did not fit together or within other categories, these was disregarded. The spread of comments can be seen in figure 9.

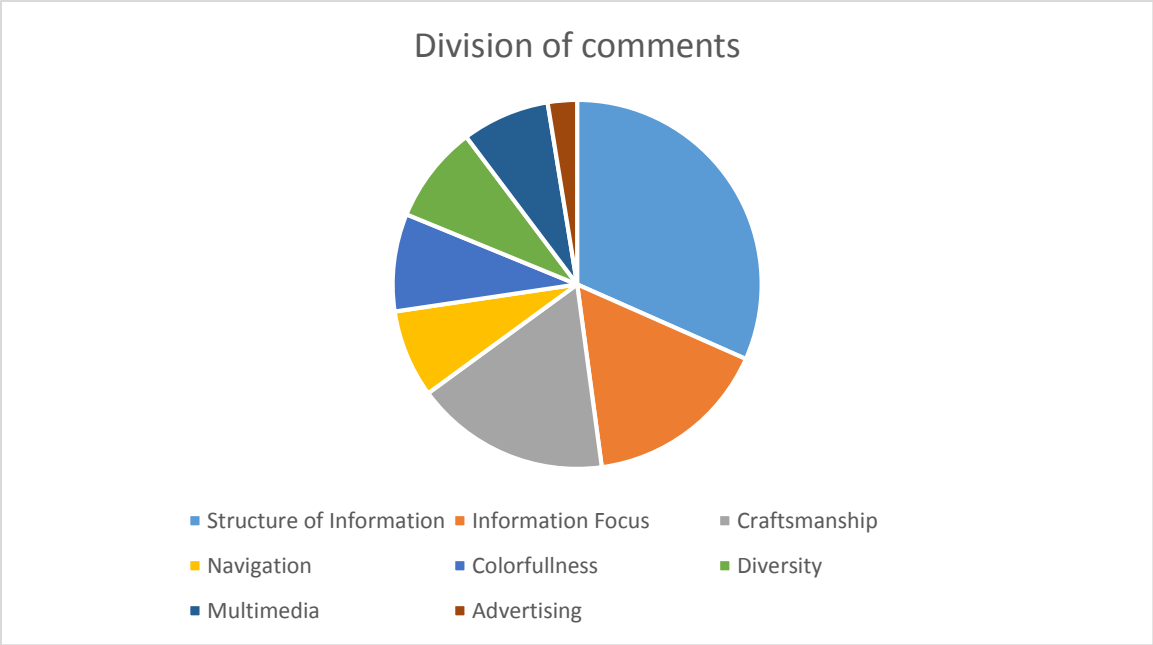


Figure 9: Division of amount of comments between the eight groups identified through comments.

Most comments were made about the *Structure of information (37 comments).* The category is based on one found in research by Fogg et al. (2003) but unlike the category used by them, *navigation* is here not considered part of *Structure of information* but as its own category. The category covers in this study how the participants perceived the structure of information and content on the page both, positively and negatively. It included comments such as:

- It feels a bit unorganized. – M, 25
- It is a terribly messy site. It doesn't exactly evoke trust. – F, 56
- It isn't a lot on the web site, which makes it easier to notice everything at once. – F, 24

The second largest category is *Craftsmanship* (20) and the third was *Information Focus* (19). Then its followed by *Diversity* (10), *Colourfulness* (10), *Navigation* (9), *Multimedia* (9) and *Advertising* (3). The amount of comments in the category of *Structure of information* (37) indicates that the way information put together is important to users. The comments in the category, is generally in a positive manor in regards to web sites of low and medium degree of complexity, and in a negative manor in regards to websites with high degree of complexity. This can be seen in figure 10. This indicates that the more complex a website is, the less likely it is to be perceived well by users. Words use to describe the site negatively was often "messy" and "noisy", while the words used in a positive matter was often "clean"

and “simple”. Supporting notions from Nielsen (1999) and Moshagen and Thielsch (2010) that simplicity is an important quality.

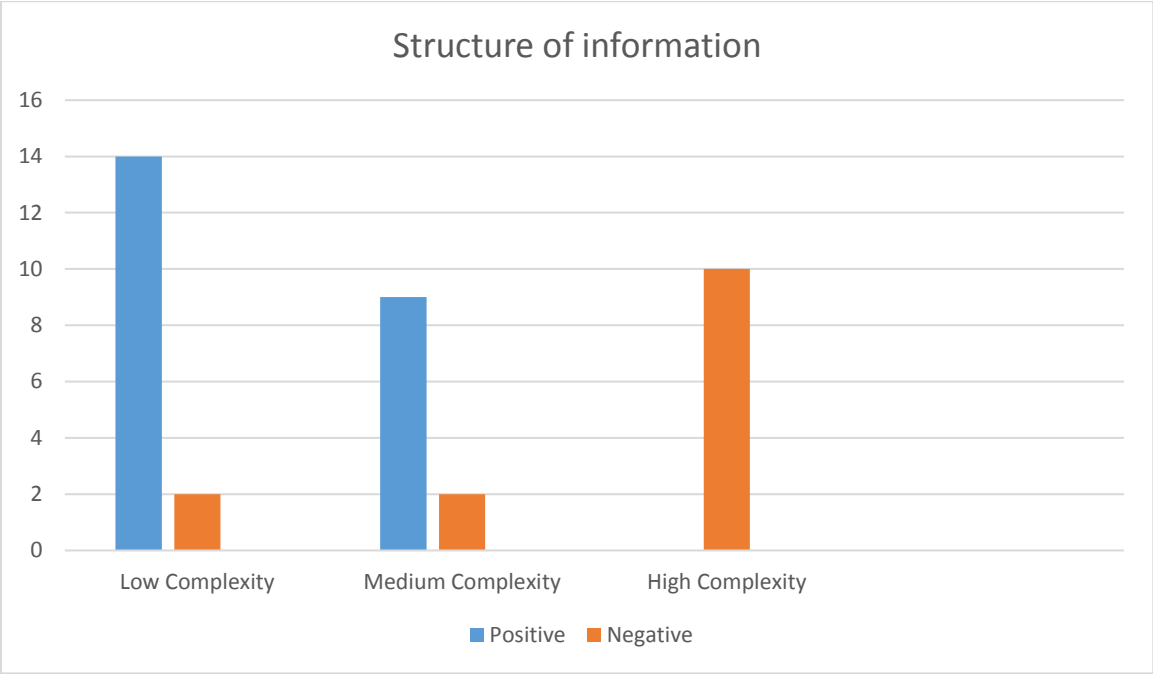


Figure 10: Division of amount of positive and negative comments regarding structure of information divided by complexity.

Craftsmanship (20) as a category was the one containing the second most comments. It is here considered, in a similar way as Moshagen and Thielsch (2010) facet named the same, to cover the perceived level of craftsmanship that has gone into the design and the result of that. It included comments such as:

- It works well as a whole. – M, 24
- It feels very professional. – M, 25
- It is simple and easy, but I could have made it myself. – M, 26

Information focus (19) is the third big category before a steep drop off in comments. It is focused on information, not on the structure of it, but on the information itself. In terms of how good information is presented to users, do they understand it or not, and the amount of information presented. It included comments such as:

- I’m unsure what it is about. – M, 25
- It is difficult to find out what it is about. – F, 56
- It isn’t packed with information. – M, 24

Those were the three categories with most comments. They can therefore be considered to be important or on top of mind of the participants when viewing websites. More from the interviews will be discussed in the next chapter.

5.0 Discussion

The following chapter covers the discussion of the results and the method used in the study in regards to existing theory and own thoughts.

5.1 Complexity and Credibility

The theory by Berlyne (1971), as previously mentioned, suggest that a medium level of complexity is the preferred amount as too little complexity doesn't give enough arousal and too much complexity creates to much arousal. This has been both supported (Geissler, Zinkhan & Watson, 2006) and not found to be true (Pandir & Knight, 2006) in regards to websites. In this study focus has mainly been on the relationship between complexity and credibility, but attractiveness has also been investigated and the results shows a clear preference on level of complexity for both complexity and attractiveness. The optimal degree of complexity in this study was found to be low, 10 or less elements on a page. Followed by medium degree and lastly high degree.

Complexity was in this study defined by the number of elements, in a similar fashion as done by Geissler, Zinkhan & Watson (2006), although here the elements were not limited to only number of screens, graphics and links. Berlyne (1960) has previously defined credibility as not only the number of elements, but also the dissimilarity between the elements and whether or not they are perceived as one. This study did not focus on the dissimilarity between element or if they were perceived as one with one expectation being the main navigation on each side if it followed the gestalt principals of similarity and proximity (Johnson, 2014). This creates a less precise, but more manageable, definition of complexity as users might see parts of the web page as one elements while its counted as several in this study. This has its positive aspect in that it makes less time consuming to gather stimuli material, but using the gestalt principals to analyse stimuli material when creating an operational definition of complexity might create a more precise view of

complexity. This is likely have an significantly effect on the results as the definitions of complexity used in the study would have only been skewed towards less elements for each level of complexity, but its recommended for future research to take the gestalt principals more into consideration when defining complexity.

It was found a clear linear relationship between the degree of complexity, as defined in this study, and both perceived credibility and attractiveness when exposed to a website in a short exposure. This conflict with the proposed inverted u-relationship for complexity theorized by Berlyne (1971) and supports the findings of Pandir & Knight (2006). The reasons for this might be several. While found to be significant, the difference between low and medium complexity in terms of rated credibility is quite low. The mean for low level of complexity was rated to 9,8387 while medium level was rated to 9,4516, and high degree is rated to 7,7177. Based on the interviews there were one negative comment that were repeated among several subjects in regards to the two low complexity home pages – that comment revolved around the pages being a bit boring. While the low complexity pages were positively received for being “clean” and “simple”, they were also perceived to be a “bit boring”. This might be an indication that there might be a minimal range of complexity which websites need to exceed, as found by Geissler, Zinkhan & Watson (2006), in order to be interesting enough for the users by evoking enough arousal as found by Berlyne (1971). In their work on aesthetics Moshagen & Thielsch (2010) advocates simplicity, but they also advocate diversity as a facet of visual aesthetics because a web site that is only simple will be perceived as boring. Diversity act as a counterweight to simplicity by creating interest and tension through adding visual complexity or richness (Moshagen & Thielsch, 2010). This experiment defined low complexity by ten or less elements. Most stimuli in the low complexity category that was used in the study contained close to 10 elements, while the lowest range in Geissler, Zinkhan & Watson (2006) had 2 (1 link and 1 graphic, and number of screens was 1). This is a lower number of elements than used in this study and might have created less visual interest resulting in less arousal for the users which might influence their impression negatively. The level of low complexity might have, with less elements than in the current study, scored lower and created an inverted u-relationship as found by Berlyne (1974). This study as it is did not find this to be true, and it is difficult to find many real pages with as few elements as in the study by Geissler, Zinkhan & Watson (2006) without manipulate the pages.

Another factor in this study, in addition to number of elements in the definition of low complexity, that might have caused a linear relationship is time. This study did not find a minimal range of complexity that needs to be exceeded in order to create positive judgements of credibility in a short exposure. But through the interviews the linear relationship was found to be subject to change as the exposure time went from 1500ms to 10000ms. After the short exposure the linear relationship, based on a small sample, was still there as found in the main experiment. After the long exposure the linear relationship had changed to a more inverted u-relationship as stimuli in the low complexity category were rated lower in regards to credibility then stimuli in the medium complexity category which were rated highest. The sample for the interviews is very limited and no statistical judgement can be made in order to reach significant results so this needs more research. It does however give a small indication that time might influence credibility judgements affected by complexity and the inverted u-relationship found by Berlyne (1971) and Geissler, Zinkhan & Watson (2006). The effect caused by exposure time might be explained by processing fluency effect proposed by Reber, Schwarz & Winkielman (2004) which states that the more fluently something can be processed the more positively the aesthetic response will be. It is also connected to the limits of human capacity of attention which causes simple information to be positively valued as it uses less of an already limited capacity of attention (Robertson, 1989). The websites in the low complexity category should be processed more fluently and combined with the short exposure time this might cause the low complexity websites to be rated highest in terms of credibility and attractiveness. The effect of processing fluency might then be reduced when user have more time to view the content during the long exposure as there is more time to process the page and make a judgement. Exposure has been found to have a similar effect on logo design where simple logo was recognized faster, but more exposure to logos led attitudes and recognition to be increased for complex logo`s (Grinsven & Das, 2016). Providing what Grinsven & Das (2016) described as short-term benefits for simple logos and long-term benefits for complex logos. Signs of a similar effect may be interpreted by the finding in this study, but it needs more research. The effect the exposure time is also supported by one of the comments made when a participant was asked why he changed his judgement by two or more points from a short exposure to a long exposure:

- “It probably stood out as messy when I first saw it for a short time. It was a lot of elements, but when I got to see it over time it was better as I could see more of what it was about even though it is a lot of information” – M, 24

Exposure time seems to be an important aspect that might influence the relationship between complexity and preference. The result of the experiment shows a linear relationship between complexity and perceived credibility of web pages in a short exposure. While the relationship might look differently, as indicated by the interviews, if the exposure time where longer or definitions of complexity had been different it still shows that complexity has a strong effect on the perception of credibility (and attractiveness) with high levels of complexity being negative.

5.2 Difference in credibility judgements between information and business web sites

The study selected web sites are based on two different categories: business and information. Only the business category found the perception of credibility to be significantly impacted by complexity, while the perceived attractiveness for both categories were significantly impacted by differences in complexity. This goes a bit against what has been found in research previously where it has been found that credibility is correlated with attractiveness (Fogg, 2003; Robins & Holms, 2008). Still while not being perceived at particularly attractive the high complexity web sites in the information category were perceived to be so credible that the difference was not enough to be found significantly different from medium and low degree of complexity. A reason for this might be either or both, because of familiarity or difference in quality of stimuli material.

Finding stimuli material with enough elements to put it in the high complexity category was difficult and within the information category all of the high complexity websites are news sources. This might have an impact on the judgement of credibility as users are likely to already be using similar web pages both in terms of layout and content. The previous experience users are likely to create perceptual patterns or frames to be present which in turn can bias the user’s perception as they might have expectations to how the site work and what it contains (Johnson, 2014). That the web sites are following conventions in terms of layout and style also helps users faster to recognize and understand

the web pages (Krug, 2006). Both conventions and perceptual patterns might play a role in increasing the processing fluency of those pages reducing the effect of complexity and that, combined with the possibility of the web sites being experienced as familiar might be causing users to rate credibility higher. While judgements of credibility online are shifting from authority to reliability (Lankes, 2008) it might be that the information pages, as they both resembles and are news outlets, gain additional credibility unlike business pages because of a presumed journalistic process.

The second possible explanation or second additional factor is that the quality of the high complexity web sites in the information category is considered to be of higher quality than those in business category. This is supported to some degree by the ratings of attractiveness which can be seen in table. 5. The mean for attractiveness in high complexity business sites are (2,87) and information sites (3,89), while medium and low was fairly similar with maximum 0.3 in difference. The reason for a potential quality difference is difficult to explain, but it might come from the difficulty to find high complexity in the business category compared to the information category with several news outlets available. Another potential reason is also the familiarity with news outlets as mentioned in the previous section.

5.3 The effect of attractiveness on surface credibility and the impact of complexity

The findings in this study has found that attractiveness seem to be correlated to surface credibility. Attractiveness is likely to be an important determining factor for the perception of credibility, especially surface credibility, which have been found in previous research (Fogg, 2003; Robins & Holms, 2008). It is therefore important to continue research into what makes something attractive and if the different aspects of aesthetics impacts credibility differently. One thing that was made clear in this study is that complexity has a strong impact on perceived attractiveness ($p < .000$, $r = -,724$) and the level of complexity should therefore be taken into consideration when creating design. The study also found a linear relationship which supports the notion of simplicity being important to aesthetics perception at least in a first impression. The interviews also support this notion with most comments being concerned with the category of structure of information and common words being used were “simple” and “clean”. These comments might also give an indication of support to the facet of simplicity (Moshagen & Thielsch, 2010) and the dimension classical aesthetics

(Lavie & Tractinsky, 2003) is important to an aesthetic impression, especially in short exposures. Lavie & Tractinsky (2003) had a second dimension, expressive aesthetics, which covers creativity and originality. While Moshagen & Thielsch (2010) suggested a facet named diversity which consisted of novelty, dynamics and creativity. Both of these suggests that there isn't enough to just make a clean and balanced design. This was also echoed in the interviews as a negative aspect that evolved around the low complexity pages were that they were perceived to be a bit boring. There seems therefore that just making a web site simple doesn't necessarily mean it will be perceived as attractive, although the lower complexity in study was perceived as most attractive, but instead creating simple design with a low degree of complexity while still maintaining a visual flair.

5.4 Lack of impact of navigation and logo on credibility judgements in a short exposure

This study did not find differences in navigation and logo to have an effect on perception of credibility in a short exposure. This was not expected. It indicates that those elements do not play an important role in judging credibility in a first-impression, at least in a short exposure. That might be true based on two factors. Navigation might not be prioritized at first by the users, and perhaps even less in this study where they do not get to use the website. Nielsen (2000) has found through several years of studies that users often skip navigation and go straight for the content. This doesn't mean navigation isn't important, but it indicates that it might not be what users are interested too look at first. Overtime he has also found users to rarely look at logos (Nielsen, 2000). These elements may also play a very small role, at least in a short exposure, in judgements of credibility as they have little to do with the content and with limited time it might be even less prioritized. Users are also very goal oriented (Johnson, 2014) and in this study the participants did not have a clear goal they wanted to achieve which may also have reduced the importance of the navigation as it did not play any role in finding something they would look for.

The result for navigation and logo might also have been impacted by the fairly small sample of stimuli, and because normal scanning patterns might have been interrupted as a result of the way eye tracking was set up. Both of these elements is discussed in the discussion of selected methods.

5.5 Discussion of selected methods

The methods for this study have been both qualitative and quantitative. A mix of methods have its strengths which are mentioned earlier, but it also has challenges. The perhaps largest challenge is to combine the two methodological traditions (Leedy, Ellis & Ormrod, 2014), and that have been a challenge in this study. Selecting a good quantitative method to follow up on the data collected in the experiment in order to get more substance to the numbers in an explanatory research design was a challenge. This study did not only aim to get more substance to the existing data by using interviews as a qualitative method, but it was also attempted to find answers to an additional research question (#3). The success of finding answers to research question 3 was limited because interviews in general, as a qualitative method, is good at investigating a topic more in depth to get nuanced insights not easily found in statistical data (Leedy, Ellis & Ormrod, 2014), but data collected from interviews are often a smaller sample and therefore less generalizable to a larger population. This hurts the external validity of the answers collected in the interviews, and future studies is recommended to look at research question 3 in a pure quantitative study.

This study attempts to answer several hypothesises, and because of time and resources only one experiment has been done. Number of hypothesis investigated in one experiment is recommended to not be too many as it leads to several variables to control for which makes the experiment more complex (Lazar, Feng & Hochheiser, 2010). This has allowed the study to investigate several factors at once but it has also led to a slight skewed focus, where complexity has been the main focus and navigation/logo has become secondary. These will be discussed more below as the two following paragraphs discusses some of the advantages, disadvantages and challenges that the study has encountered in relation to the experiment and interviews.

5.5.1 Experiment one

The main experiment was conducted as a within subject experiment design. Within subject design has advantages and disadvantages that have been mentioned earlier. The main reason for choosing such a method is for ease of recruiting as a lower sample is necessary than with between subject's design (Lazar, Feng & Hochheiser, 2010) and it reduces potential individual effect between-subjects. The downside with within-subject design is that its susceptible to potential learning effect, but this was considered as unlikely to affect the

experiment as no measurement are made that can really be enhanced by experience. Potential learning effects were also limited by having four stimuli at the start that were used as practise so the participants would learn and get used to the experiment process/tasks. Thereby reduce any potential effect of a learning curve when they started on the actual stimuli material. It didn't seem to be any learning effect affecting the participant's answers and it is not considered to be an issue in the study.

Stimuli in the experiment was selected by existing websites online instead of create the stimuli material from scratch. This choice was made as the study wanted a realistic selection of websites that were more likely to be similar to what participants would meet on the World Wide Web. This is positive for the external validity of the experiment as the findings of the study are likely to also be true for more webpages, and different categories, as the selected set should reflect a realistic variance of webpages that exist. However, the variance between the selected stimuli also means there is several variables that is difficult or impossible to control for. Aesthetics dimensions such as colour were for example not considered when selecting the stimuli. This reduces the internal validity of the experiment, but the choice was made for increasing the realism of the experiment because if complexity was to be found as an influencing variable of credibility it should also be able to affect credibility despite the possibility of confounding variables. Still it wasn't possible to get a completely realistic experiment as the participants were still put in front of another computer then they are used to and in another setting than they are used to. The websites were also still images that gave no opportunity for scroll or display animations which might be a factor influencing the complexity. This reduces the external validity of the experiment to some degree, but it was a choice that was made due to the short exposure time which would cause scrolling to be unnatural to do. While animations, because of the movement is likely to drag attention (Johnson, 2014), it isn't that common to start animations right away but instead wait for the participant to scroll.

This study used two randomized sequences of the stimuli in the experiment which was done because of the difficulty handling the data if it was completely randomized. This does increase the chance that the sequence might have influenced to some degree the answers of the participants, and the experiments internal validity could be improved by either running a completely randomized sequence or do additional experiments with new sequences.

Only one quantitative experiment was conducted in this study which caused there to be several hypotheses that were investigated at once. This gave challenges as there also were more variables study than if only focusing on one research question or one hypothesis. Because possible fatigue, which changes were done to reduce the likelihood of, it was deemed not wise to have separate stimuli investigating complexity, navigation and logo. Instead a decision was made to select the stimuli based on complexity, and then select stimuli to be used to answer research question 2 based on the selected stimuli which fitted the criteria for navigation and logo. This wasn't optimal as none of the websites in the high complexity had a unconvential logo or navigation. Limiting the investigation of navigation and logo to only two degree of complexity. It was still deemed worth it to go forward with it as the data through eye-tracking could still give an indication of answers to the research question, although it is important to realize that the answers are limited to webpages divided into two degrees of complexity and thereby a lower sample.

By using eye tracking equipment in the experiment it was possible to make sure each user has the same starting point when viewing the website by adding a cross with an AOI which requires the participant to look at it for a certain amount of time before showing the stimuli. This was positive in regards to giving all participants an equal starting point. The AOI was placed at the centre of the screen which may have had an impact on the effectiveness of logo and navigation as elements because it disrupts users normal scanning pattern. Normally users in western countries start scanning a page from top left – which coincides with were logos are normally placed (Whitenton, 2016). Not having users start at the top left corner might therefore have given less focus on conventional logos then it normally would have which might have influenced the lack of difference between unconventional and conventional.

The study found clear results in terms of the influence of complexity on credibility in a short exposure, and the author believes it would be possible to replicate the result by doing new experiments both following the context of this experiment and to find the result in different contexts. However, for the impact of navigation and logo, and the effect of exposure time on credibility – new research efforts have to be made which might potentially find differences.

5.5.2 *Interviews*

Interviews were used in the study for two purposes. One was to see if answer to research question 3 could be found which was limited due to the small sampling size as mentioned earlier which makes it hard to make any conclusions. It did however give some indication, based on the judgements the participants made and the answers they gave, to possible effects of exposure time. The second purpose was to get more nuances and insight into the result of the data. New participants were used, but they followed the same pattern as found in experiment one. This was considered successful as interviews gave more insights into the thoughts behind their judgements and what they based them on.

Semi-structured interviews as a method allowed for following up on answers to get clarification, and ask both open and closed questions. One major drawback for using interviews is that it often is based on memories that the subject has of previous experiences and memories are error-prone (Leedy, Ellis & Ormrod, 2014). This is a potential risk for the study. An attempt was made to help the participants by letting them see the stimuli while they were answering the question. Using props can be helpful in an interview setting (Rogers, Sharp & Preece, 2011). It might also cause participants to not base their answer on what they remember from their judgement as it happened previously, but instead see and answer based on what they think when viewing the stimuli at that moment. Analysing the answers was also a challenge as it would be combined with a quantitative experiment, but using affinity diagram for analysis gave more insight into the user's judgements and added, together with actual comments made by participants, insight and nuance to data found in the experiment.

6.0 Conclusion

This study aimed at investigate surface credibility in the domain of World Wide Web and how it is affected by aesthetic property of complexity and common elements such navigation and logo. Credibility is seen as an important aspect of persuasion (Fogg, 2003). On the web where users spend a short amount of time on webpages before they either leave or stay, and for persuading users to stay, then surface credibility is likely to be an important factor. Aesthetic has been found previously to impacted perception of credibility (Fogg, 2003; Fogg et al., 2003; Robins & Holmes, 2008) and this study echoes those findings. It was found that there is a significant correlation ($p > .001$) between attractiveness and

perceived credibility which has a linear relationship where high attractiveness equals high degree of credibility in short exposure. Common elements such as navigation and logo were not found to have an impact on credibility.

Complexity was seen as a potential property that could influence credibility and which based on Berlyne`s theories (1971) has a strong influence on perceived aesthetics through increasing or decreasing arousal, where moderate degree of arousal is believed to have the strongest effect. This study found complexity to have significantly impact, in short exposures, on credibility ($p < .000$, $F(2, 90) = 21{,}031$) in the experiment on across all the stimuli showing a linear relationship as with attractiveness. It indicates that moderate degree of complexity might not be the optimal amount in short exposures of web pages, but over a longer exposures signs were found indicating that the theory of Berlyne (1971) and finding of Geissler, Zinkhan & Watson (2006) might be correct without this study being able to make a clear conclusion. Complexity was however not found to be significant for the information category of websites which might be due to more tolerance for complexity on those types of websites because of common exposure to similar pages, but this requires more research. Still complexity, or lack thereof, is believed to be an important influencing property of user's perception of aesthetic quality based on the results and should be considered when designing for increased surface credibility. Designers is therefore recommended to keep the complexity of a web page as low as possible while still maintaining some degree of visual interest so it will not be perceived as boring. These finding do also indicate that there might be some aesthetic dimensions, properties or elements that impacts credibility differently and that more research should investigate further the impact of complexity and other potential influencers.

6.1 Future research

This thesis has touched upon several interesting points, but there is particularly one point that might be of particular interest for future studies. One of the goals of this study was to investigate how credibility judgements change over time, but due to lack of time and constraint of the methodological approach a clear statistically significant answer was not found. Indications were found, but further studies are needed to be able to understand how a first-impression of credibility is made and how it stands over time. This can give more

insight into how important a first-impression is and potential pitfalls that designers, users and businesses needs to be aware of.

7.0 Bibliography

Anderson, E. (2015) *Teenagers spend 27 hours a week online: how internet use has ballooned in the last decade.* The Telegraph. Available from: http://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/digital-media/11597743/Teenagers-spend-27-hours-a-week-online-how-internet-use-has-ballooned-in-the-last-decade.html [Accessed 25th August 2017].

Ben-Bassat, T; Meyer, J & Tractinsky, N. (2006) *Economic and subjective measures of the perceived value of aesthetics and usability.* Transactions on Computer-Human Interaction. 13 (2), p. 210-234.

Berlo, D K; Lemert, J B & Mertz, R J. (1969) *Dimensions for Evaluating the Acceptability of Message Sources*. The public Opinion Quarterly. 33 (4), p. 563-576.

Berlyne, D E. (1960) *Conflict, arousal, and curiosity.* New York, USA, McGraw-Hill Book Company.

Berlyne, D E. (1971) *Aesthetics and psychobiology.* New York, USA, Meredith Corportation.

Bonnardel, N; Piolat, A & Bigot, L L. (2011) *The impact of colour on Website appeal and users cognitive processes.* Displays. 32 (2), p. 69-80.

Brady, L & Philipps, C. (2003) *Aesthetics and Usability: A Look at Color and Balance*. SURL – Software Usability Research Laboratory Witchita State University. Available from: http://usabilitynews.org/aesthetics-and-usability-a-look-at-color-and-balance/ [Accessed 29th October 2017].

Chen, J V; Lin, C; Yen, D C & Linn, K. *The interaction effects of familiarity, breadth and media usage on web browsing experience.* Computers in Human Behavior. 27 (6), p. 2141-2152.

Ekman, P. (1992) *An argument for Basic Emotions*. Cognition and Emotion. 6 (3/4), p. 169-200.

Elliott, B. (2017) *Seth Godin: the new currency of trust and attention*. Behind the Brand with Bryan Elliott. Available from: http://behindthebrand.tv/2017/05/08/seth-godin-the-new-currency-of-trust-and-attention/ [Accessed 12th October 2017].

Fessenden, T. (2017) *Horizontal Attention Leans Left*. NNgroup. Available from: https://www.nngroup.com/articles/horizontal-attention-leans-left/ [Accessed 19th November 2017].

Flanagin, A J & Metzger, M J. (2000) *Perceptions of Internet Information Credibility.* Journalism and Mass Communication Quarterly. 77 (3), p. 515-40.

Fogg, B J. (2003) *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufman.

Fogg, B J; Soohoo, C; Danielson, D; Marable, L; Stanford, J & Tauber, E. (2003) *How Do Users Evaluate the Credibility of Web Sites? A study with Over 2,500 Participants*. Available from: https://pdfs.semanticscholar.org/4c8f/e677cddc84a058491f51701c2edbbaa5aefa.pdf [Accessed 02th September 2017].

Fogg, B J & Tseng, H. (1999) *The elements of Computer Credibility.*Proceedings of the SIGCHI conference on human factors in computing systems, 15-20 May 1999. Pittsburgh, USA, ACM. p. 80-87.

Frohlich, D. (2004) *Beauty as a Design Prize*. Human-Computer Interaction 19 (4), p. 359-366.

Geissler, G; Zinkhan, G & Watson, R T. (2001) *Web Home Page Complexity and Communication Effectiveness.* Journal of the Association for Information Systems. 2 (2), p. 1-46.

Geissler, G; Zinkhan, G & Watson, R T. (2006) *The Influence of Home Page Complexity on Consumer Attention, Attitudes, and Purchase Intent.* Journal of Advertising. 35 (2), p. 69-80.

Google. (n.d.) *Credibility.* Available from: https://www.google.no/search?q=credibility&oq=credibility&aqs=chrome..69i57j69i60j0j69i65j69i60j0.5913j0j1&sourceid=chrome&ie=UTF-8 [Accessed 05th Semptember 2017]

Grinsven, B V & Das, E. (2016) *Logo design in marketing communications: Brand logo complexity moderates exposure effects on brand recognition and brand attitude.* Journal of Marketing Communications. 22 (3), p. 256-270.

Harmon, R R & Coney, K A. (1982) *The Persuasive Effects of Source Credibility in Buy and Lease Situations.* Journal of Marketing Research. 19 (2), p. 255-260.

Hovland, C I; Janis, I & Kelley, H. (1953) *Communication and Persuasion: Psychological Studies of Opinion Change*. Yale University Press.

Johnson, J. (2014) *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines*. 2nd edition. Waltham, USA, Morgan Kaufmann.

Karvonen, K. (2000) *The beauty of Simplicity.* Proceeding on the 2000 Conference on Universal Usability, 16-17 November 2000. Arlington, USA, p.85-90.

Kouzes, J M & Posner, B Z. (2011) *Credibility: How leaders gain or lose it, why people demand it.* 2nd edition. San Fransisco, USA, John Wiley & Sons Ltd.

Krug, S. (2006) *Don`t make me think: A common sense approach to web usability.* 2nd edition. Berkely, USA, New Riders Publishing.

Lavie, Talia & Tractinsky, Noam. (2004) *Assessing dimensions of perceived visual aesthetics of web sites*. International Journal of Human-Computer Studies. 60 (3), p. 269-298.

Lazar, Jonathan; Feng, Heidi Jinjuan & Hochheiser, Harry. (2010) *Research methods in Human-Computer Interaction*. West Sussex, United Kingdom, John Wiley & Sons Ltd.

Leedy, P D & Ormrod, E J. (2014) *Practical research, planning and design*. Edinburgh, United Kingdom, Pearson Education.

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), p. 115-126.

Liu, Y. (2003) Engineering aesthetics and aesthetic ergonomics: theoretical foundations and a dual-process research methodology. Ergonomics. 46 (13-14), p. 1293-1292.

Liu, Z. (2004) *Perceptions of credibility of scholarly information on the web*; Information Processing & Management. 40 (6), p. 1027-1038.

Liu, Z & Huang, X. (2013) *Evaluating the credibility of scholarly information on the web: A cross cultural study*. International Information & Library Review. 37 (2), p. 99-106.

Loranger, H. (2015) *Simplicity Wins over Abundance of Choice.* NNgroup. Available from: https://www.nngroup.com/articles/simplicity-vs-choice/ [Accessed 16th November 2017].

Morville, Peter. (2005) *Ambient Findability.* Sebastopol, Canada, O`Reilly Media.

Moshagen, M & Thielsch, M T. (2010) *Facets of visual aesthetics.* International Journal of Human-Computer Studies. 68 (10), p. 689-709.

Moss, G; Gunn, R & Heller, J. (2006) *Some men like it black, some women like it pink: consumer implications of differences in male and female website design.* Journal of Consumer Behaviour. 5 (4), p. 328-341.

Nielsen, J. (1999) *The practice of Simplicity.* USA, New Riders Publishing.

Nielsen, J. (2000) *Is Navigation Useful?.* NNgroup. Available from: https://www.nngroup.com/articles/is-navigation-useful/ [Accessed 4th December 2017].

Nielsen, J. (2011) *How long do users stay on web pages.* NNgroup. Available from: https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages/ [Accessed 27th August 2017].

Nisbett, R & Wilson, T D. (1977) *The Halo Effect: Evidence for Unconscious Alteration of Judgment.* Journal of Personality and Social Psychology. 35 (4), p. 250-256.

Norman, D. (2004) *Introduction to this special section on beauty, goodness, and usability.* Human-Computer Interaction. 19 (4), p. 311-318.

NTB. (2017) *Støre: – Skulle ha avklart privatøkonomien før valget.* E24. Available from: https://e24.no/makro-og-politikk/jonas-gahr-stoere/stoere-skulle-ha-avklart-privatoekonomien-foer-valget/24154651 [Accessed 24th September 2017].

Pandir, M & Knight, J. (2006) *Homepage aesthetics: The search for preference factors and the challenges of subjectivity.* Interacting with Computers. 18 (6), p. 1351-1370.

Park, S; Choi, D & Kim, J. (2004) *Critical factors for the aesthetics fidelity of web pages: empirical studies with professional web designers and users.* Interacting with Computers. 16 (2), p. 351-376.

Pittard, N; Ewing, M & Jevons, C. (2007) *Aesthetic theory and logo design: examining consumer response to proportion across cultures.* International Marketing Review. 24 (4), p. 457-473.

Pornpitakpan, C. (2004) *The Persuasiveness of Source Credibility: A Critical Review of Five Decades Evidence.* Journal of Applied Social Psychology. 34 (2), p. 243-281.

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

Robertson, Kim. (1989) *Strategically Desirable Brand Name Characteristics.* The Journal of Consumer Markerting. 6 (4), p. 61-71.

Robins, D & Holmes, J. (2008) *Aesthetics and credibility in web site design*. Information Processing & Management. 44 (1), p. 386-399.

Rogers, T; Sharp, H & Preece, J. (2011) *Interaction design: beyond human-computer interaction.* 3rd edition. West Sussex, United Kingdom, John Wiley & Sons Ltd.

Rubin, J & Chisnell, D. (2008) *Handbook of Usability testing: How to plan, design, and conduct effective tests*. Indianapolis, USA, Wiley Publishing.

Sonderegger, A & Sauer, J. (2010) *The influence of design aesthetics in usability testing: Effects on user performance and perceived usability.* Applied Ergonomics. 41 (3), p. 403-410.

Statistisk Sentralbyrå (SSB). (2017a) *Norsk Mediebarometer.* Available from: https://www.ssb.no/kultur-og-fritid/statistikker/medie/aar [Accessed 12th October 2017].

Statistisk Sentralbyrå (SSB). (2017b) *ICT usage in households.* Available from: https://www.ssb.no/en/teknologi-og-innovasjon/statistikker/ikthus/aar [Accessed 12th October 2017].

Stavrositu, C & Sundar, S S. (2008) *If Internet Credibility Is So Iffy, Why the Heavy Use? The Relationship between Medium Use and Credibility.* CyberPsychology & Behavior. 11 (1), p. 65-68.

Svaar, P. (2017) *Partifelle: – Støres privatøkonomi ødela partiets troverdighet.* NRK. Available from: https://www.nrk.no/norge/partifelle_-_-stores-privatokonomi-odela-partiets-troverdighet-1.13688068 [Accessed 24th September 2017].

Tractinsky, N; Katz, A. S & Ikar, D. (2000) *What is beautiful is usable*. Interacting with Computers. 13 (2), p. 127-145.

Tractinsky, N; Cokhavi, A; Kirschenbaum, M & Sharfi, T. (2006) *Evaluating the consistency of immediate aesthetic perceptions of web pages*. International Journal of Human-Computer Studies. 64 (11), p. 1071-1083.

Tuch, A N; Bargas-Avila, J A & Opwis, K. (2010) *Symmetry and aesthetics in website deisgn: It`s a mans`s business.* Computers in Human Behavior. 26 (6), p. 1831-1837.

Wathen, N & Burkell, J. (2002) *Believe it or not: Factors influencing credibility on the Web*. Journal of the Association for Information Science and Technology. 53 (2), 134-144.

Watson, Leon. (2015) *Humans have shorter attention span than goldfish, thanks to smartphones*. The Telegraph. Available from: "http://www.telegraph.co.uk/science/2016/03/12/humans-have-shorter-attention-span-than-goldfish-thanks-to-smart/" [Accessed 25th August 2017].

Whitehead Jr, J L. (1968) *Factors of source credibility.* Quarterly journal of speech. 54 (1), p. 59-63.

Whitenton, K. (2016) *Website Logo Placement for Maximum Brand Recall.* NNgroup. Available from: https://www.nngroup.com/articles/logo-placement-brand-recall/ [Accessed 19th November 2017].

W3Schools. (2017) *Browser Display Statistics.* Available from: https://www.w3schools.com/browsers/browsers_display.asp [Accessed 25th August 2017].

Zajonc, R B. (1980) *Feeling and thinking: Preferences need no inferences*. American Psychologist 35 (2), p. 151-175.

8.0 Appendix

In the following section you will find all of the attachments for this study.

8.1 Attachment A: Experiment guide in Norwegian

Introduksjon

Hei og velkommen til dette eksperimentet vi skal gjennomføre i dag som en del av min masteroppgave i interaksjonsdesign ved NTNU i Gjøvik. Før vi starter ønsker jeg at du leser gjennom dette skjemaet og signerer om du er enig i det som står. Det forklarer litt av målet med eksperiment i dag, hvordan eksperimentet vil foregå og hva som vil skje med dataen som samles inn.

Det som skal skje nå er at vi skal kalibrere eye-trackeren så da må du følge med prikken på skjermen med øynene.

Nå vil du få se 28 nettsider en etter en i et og et halvt sekund. Etter hver nettside vil jeg lese opp tre påstander og du vil bli bedt om å si i hvilken grad du er enig i påstandene på en skala fra 1-7, hvor 1 er i meget liten grad og 7 er i meget stor grad. For å gå videre må du trykke på space og deretter se på krysset som dukker opp på skjermen i 2 sekunder før neste nettside kommer frem.

Etter eksperimentet vil vi gjennomføre en kort debrif hvor jeg skal forklare litt mer om eksperimentet og du har mulighet til å stille eventuelle spørsmål du vil ha.

Høres det greit ut?

Underveis – repeteres for hvert stimuli

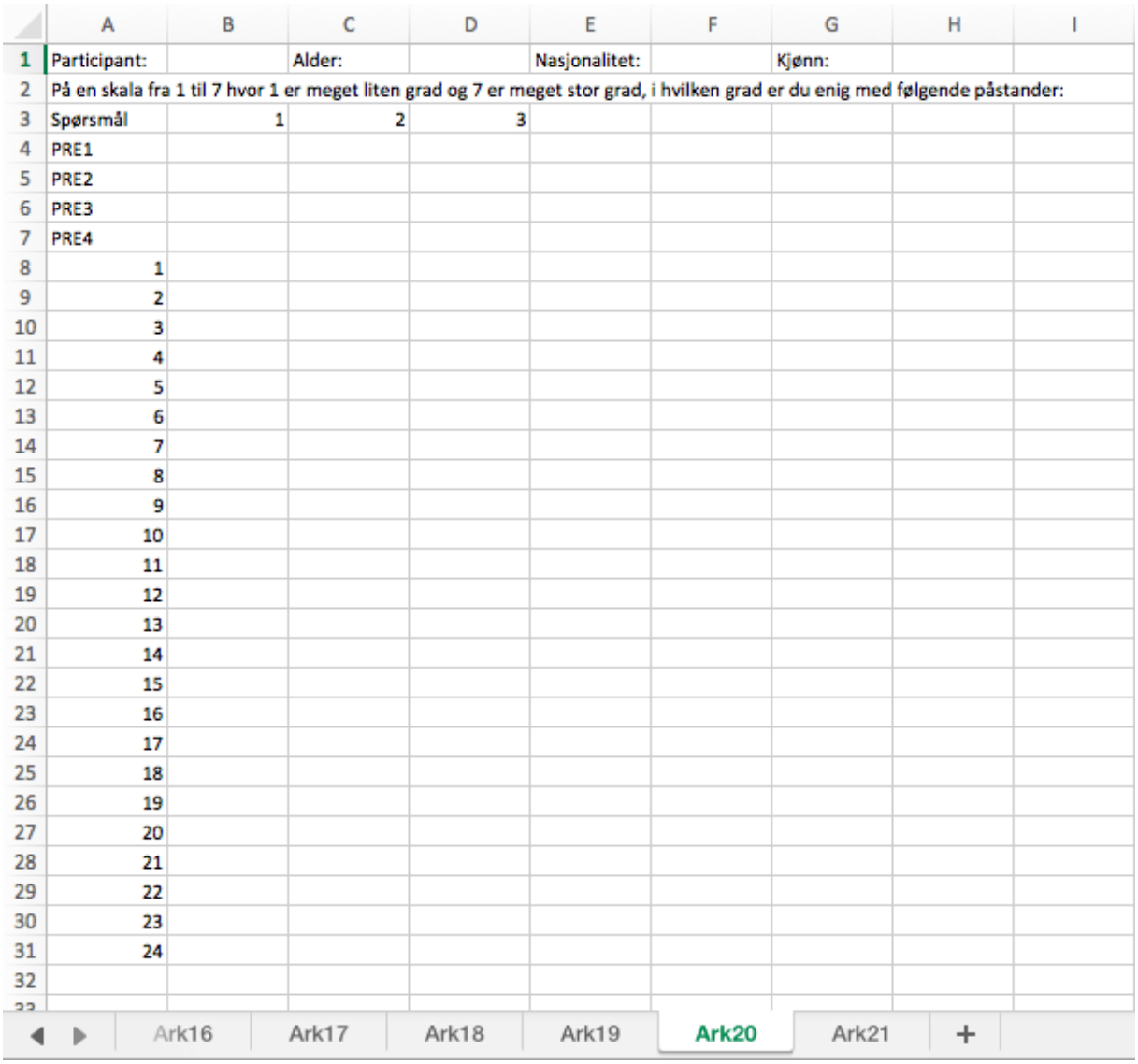
På en skala fra 1 til 7 hvor 1 representerer meget liten grad og 7 representerer meget stor grad, i hvilken grad er du enig med følgende påstander:

1. Nettsiden fremstår som profesjonell?
2. Nettsiden fremstår som visuelt appellerende?
3. Nettsiden fremstår som betryggende?

Debrief

Da var eksperimentet ferdig. Takk for din deltakelse! Da skal vi ta en liten debrief og så er vi ferdige. Først så vil jeg si at det er ikke ønskelig at du formidler bakgrunnen for eksperimentet videre til andre som ikke har gjennomført eksperimentet. Dette eksperimentet undersøker hvordan visuell kompleksitet på nettsider påvirker hvordan vi opplever troverdigheten til en nettside, eller bedriften. I tillegg til kompleksitet undersøkte vi om navigasjon eller logo også kunne ha en påvirkning på troverdigheten. Det undersøker jeg fordi vi bedømmer nettsider på en veldig kort tid og hvordan vi opplever troverdigheten er sett på som en viktig faktor til om brukere vil la seg overtale til å bli værende på siden. Igjen, takk for din deltakelse. Har du noen spørsmål?

8.2 Attachment B: Form used for noting participants' answers during experiment



	A	B	C	D	E	F	G	H	I
1	Participant:		Alder:		Nasjonalitet:		Kjønn:		
2	På en skala fra 1 til 7 hvor 1 er meget liten grad og 7 er meget stor grad, i hvilken grad er du enig med følgende påstander:								
3	Spørsmål	1	2	3					
4	PRE1								
5	PRE2								
6	PRE3								
7	PRE4								
8	1								
9	2								
10	3								
11	4								
12	5								
13	6								
14	7								
15	8								
16	9								
17	10								
18	11								
19	12								
20	13								
21	14								
22	15								
23	16								
24	17								
25	18								
26	19								
27	20								
28	21								
29	22								
30	23								
31	24								
32									

Ark16 Ark17 Ark18 Ark19 **Ark20** Ark21 +

8.3 Attachment C: Informed Consent for the experiment

Request for participation in research project

“Effects of first-impressions in web design”

Background and purpose
This is a request for you to participate in a research study that intends to map users first impressions of a set of web pages and investigate how those impressions are affected by visual aesthetics. The project is a part of Are Ingulfsen`s master thesis in the *Interaction design* master’s programme in NTNU, Gjøvik.

What does participation in the project imply?
Participation in the project means taking part in an session where you will view and rate 28 web pages. During the session there will also be used eye tracking in order to record how each website is viewed. Data collected during the session can be used in an master thesis publication. The session should not last for more than 15 minutes in total.

What will happen to the information about you?
Any personal data will be treated confidentially. Only the interviewer and supervisor will have access to personal data. Age group and gender will not be connected to your specific answer and will only be used as demographic data as background for the total group of respondents. Participants will not be recognizable in any form of publication.

The project is scheduled for completion by 10 January 2018.

Voluntary participation
It is voluntary to participate in the project, and you can at any time choose to withdraw your consent without stating any reason for doing so. If you decide to withdraw, no gathered data about you will be used in the project.

If you have any questions concerning the project, please contact:

- Are Ingulfsen: *tel. 902 43 826, email. arei@stud.ntnu.no*

Consent for participation in the study

I have received information about the project and are willing to participate

---

(Signed by participant, date.)

8.4 Attachment D: Interview guide in Norwegian

Intervjuguide

Hei og velkommen til dette intervjuet vi skal gjennomføre i dag som en del av min masteroppgave i interaksjonsdesign ved NTNU i Gjøvik. Før vi starter ønsker jeg at du leser gjennom dette skjemaet og signerer om du er enig i det som står. Det forklarer litt av målet med eksperiment i dag, hvordan eksperimentet vil foregå og hva som vil skje med dataen som samles inn.

Det som skal skje i dag er at jeg først vil stille deg noen spørsmål ang. din bakgrunn. Etter det vil du bli bedt om å rangere i hvilken grad du er enig med noen påstander om 9 nettsider som du vil få se i et og et halvt sekund. Til slutt vil du få se sidene på nytt i lengre tid og rangere de igjen, før vi vil diskutere litt rundt de spesifikke sidene.

Jeg vil repetere hva vi skal gjøre underveis så hvis du er klar så starter vi med noen generelle spørsmål.

Bakgrunn

1. Kjønn
2. Hvor gammel er du?
3. Hva gjør du til daglig?
4. Har du tidligere eller nå studert eller jobbet med design?
5. I så fall hvor lenge?
6. På en skala fra 1-7, hvor 1 er liten grad av kunnskap og 7 er stor grad av kunnskap, hvordan føler du din kunnskap rundt design er?
7. Hvor mye tid bruker du på internett i løpet av en gjennomsnittlig dag? 1-3 timer / 4-6 timer / 7 eller flere timer
8. Hva brukes tiden da i hovedsakelig på?

Første inntrykk

Nå vil jeg be deg se på 9 nettsider hvor du får se hver side i kun et og et halv sekund. Etter du har sett en nettside vil jeg be deg si i hvilken grad du er enig i tre påstander om nettsiden på

en skala fra 1-7, hvor 1 er i meget liten grad og syv er i meget stor grad. Dette repeterer vi for hver nettside. Påstandene er:

1. Nettsiden fremstår som profesjonell?
2. Nettsiden fremstår som visuelt appellerende?
3. Nettsiden fremstår som betryggende?

Diskusjon og andre inntrykk

I denne delen vil vi diskutere 6 av sidene du har sett på tidligere. Før vi går inn i en diskusjon rundt sidene vil du først få se sidene på nytt i 10 sekund hver og du vil bli bedt om å rangere sidene på nytt slik du gjorde i forrige øvelse.

1. Nettsiden fremstår som profesjonell?
2. Nettsiden fremstår som visuelt appellerende?
3. Nettsiden fremstår som betryggende?

Nå vil du få se sidene på nytt på noen ark her, og så skal jeg stille deg noen spørsmål rundt hver av de?

1. Hva synes du om denne siden?
2. Ved tilfelle av at bruker har endret betraktelig(2 punkter) sin vurdering av en side så spør hvorfor de rangerte den lavere eller høyere denne gangen?
3. Var det noe spesifikke elementer du bet deg merke i?
4. Hva appellerer til deg ved siden?
5. Hva appellerer ikke til deg ved siden?

Avslutning

Det var siste siden. Takk for din deltakelse! Da skal jeg bare fortelle litt om eksperimentet, og så vil du ha mulighet til å stille noen spørsmål om du ønsker. Det er ikke ønskelig at du formidler bakgrunnen for intervjuet videre til andre som ikke har gjennomført intervjuet. Dette intervjuet er som sagt en del av min masteroppgave som undersøker hvordan kompleksitet i design påvirker hvordan vi opplever troverdigheten til en nettside, eller bedriften bak den. I tillegg til kompleksitet undersøkte vi om navigasjon eller logo også kunne ha en påvirkning på troverdigheten. Det undersøker jeg fordi vi bedømmer nettsider på en

veldig kort tid og hvordan vi opplever troverdigheten er sett på som en viktig faktor til om brukere vil la seg overtale til å bli værende på siden. I eksperimentet vi gjorde i starten var målet og se hvordan ditt inntrykk var i løpet av en kort eksponering, og så fulgte vi opp med en lengre eksponering for å se om inntrykket ditt endret seg om du fikk mer tid til å undersøke siden. Spørsmålene til slutt var for å lære litt mer om hvordan folk vurderer sidene, hva de tenker på og legger merke til. Har du noen spørsmål?

8.5 Attachment E: Informed consent for interviews

Request for participation in research project

“Effects of first-impressions in web design”

Background and purpose
This is a request for you to participate in a research study that intends to map users first impressions of a set of web pages and investigate how those impressions are affected by visual aesthetics. The project is a part of Are Ingulfsen`s master thesis in the *Interaction design* master’s programme in NTNU, Gjøvik.

What does participation in the project imply?
Participation in the project means taking part in an interview where you will be asked to view and rate 9 web pages. Additionally you will be asked to answer questions around the 9 web pages and your background. Data collected during the session can be used in an master thesis publication. The session should not last for more than 45 minutes in total.

What will happen to the information about you?
Any personal data will be treated confidentially. Only the interviewer and supervisor will have access to personal data. No names will be saved and participants will not be recognizable in any form of publication.

The project is scheduled for completion by 10 January 2018.

Voluntary participation
It is voluntary to participate in the project, and you can at any time choose to withdraw your consent without stating any reason for doing so. If you decide to withdraw, no gathered data about you will be used in the project.

If you have any questions concerning the project, please contact:

- Are Ingulfsen: *tel. 902 43 826, email. arei@stud.ntnu.no*

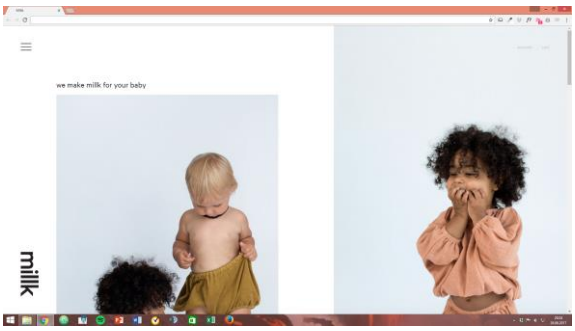
Consent for participation in the study

I have received information about the project and are willing to participate

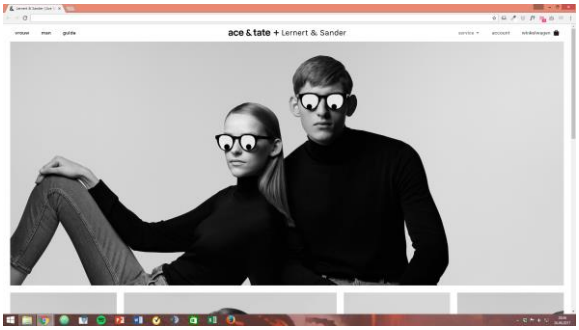
---

(Signed by participant, date.)

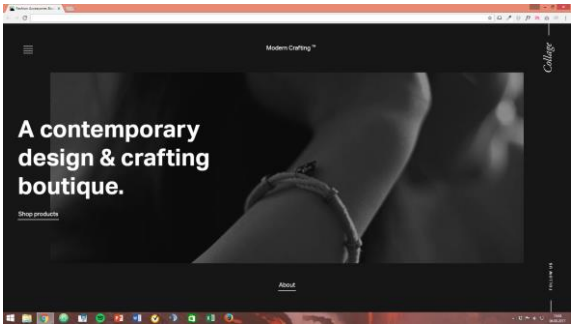
8.6 Attachment F: Stimuli used in the study



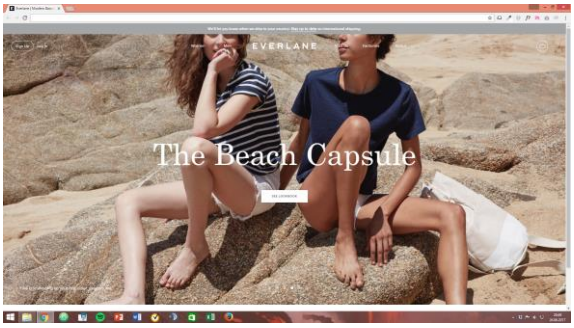
Stimuli 1 – Low complexity stimulus in business category



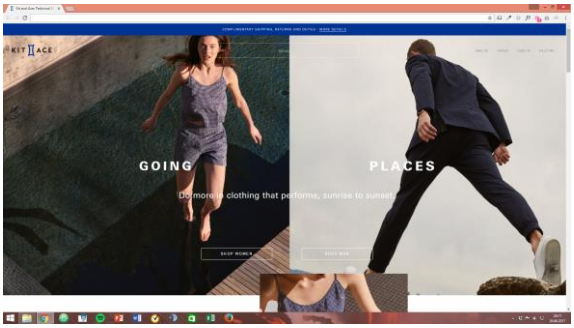
Stimuli 2 – Low complexity stimulus in business category



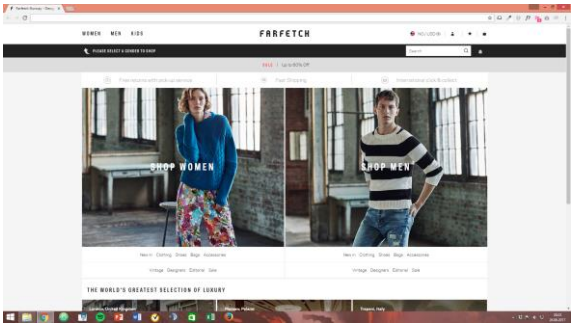
Stimuli 3 – Low complexity stimulus in business category (also used as stimulus in interview)



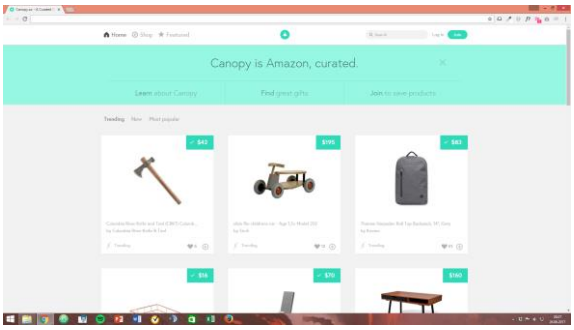
Stimuli 2 – Low complexity stimulus in business category



Stimuli 5 – Medium complexity stimulus in business category

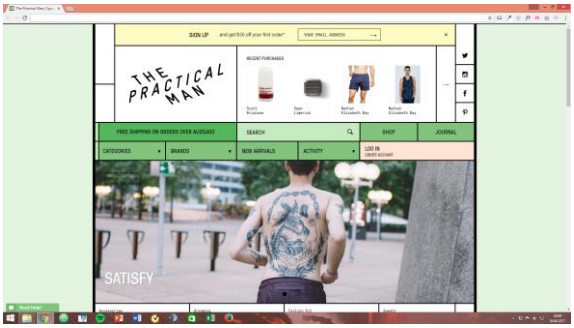


Stimuli 6 – Medium complexity stimulus in business category



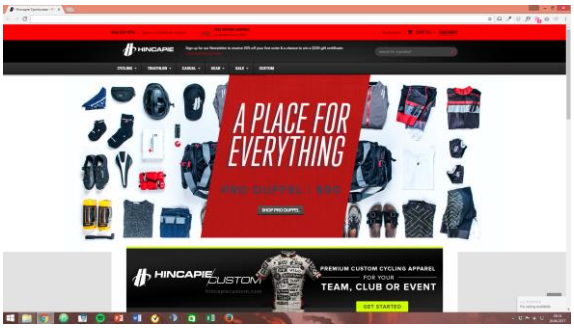
Stimuli 7 – Medium complexity stimulus in business category (also used as stimulus in interviews)

Stimuli 8 – Medium complexity stimulus in business category



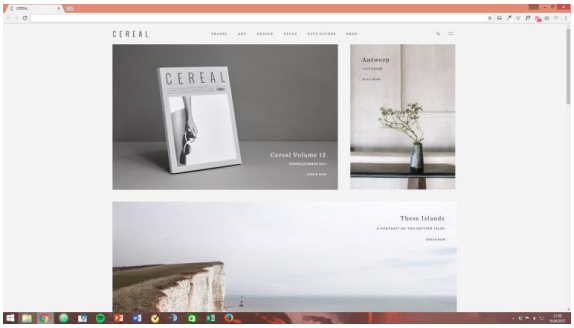
Stimuli 9 – High complexity stimulus in business category

Stimuli 10 – High complexity stimulus in business category

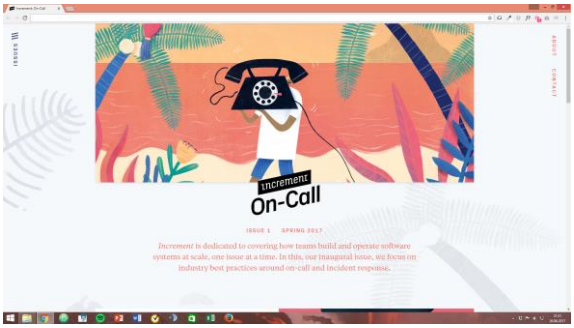


Stimuli 11 – High complexity stimulus in business category

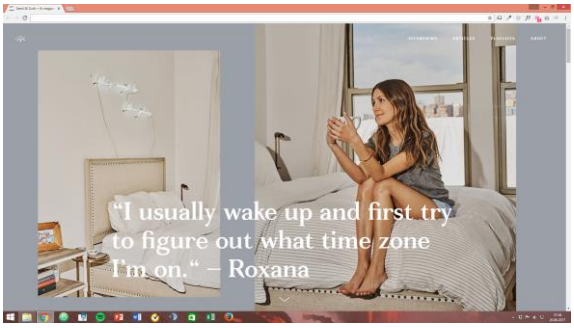
Stimuli 12 – High complexity stimulus in business category (also used as stimulus in interviews)



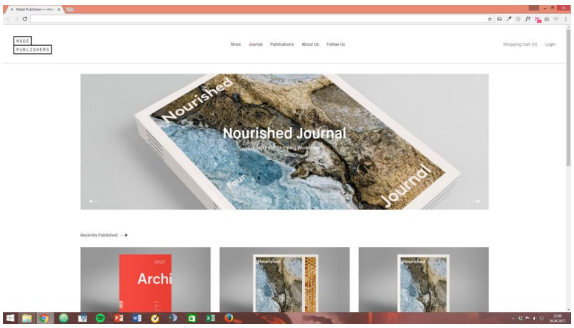
Stimuli 13 – Low complexity stimulus in information category



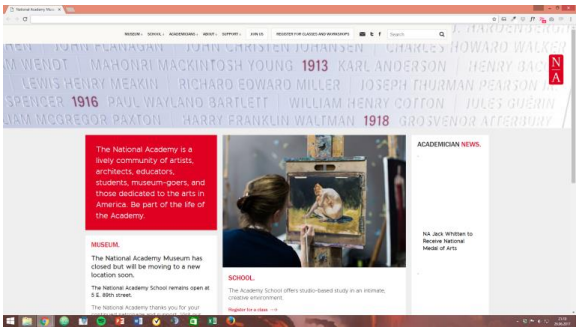
Stimuli 14 – Low complexity stimulus in information category



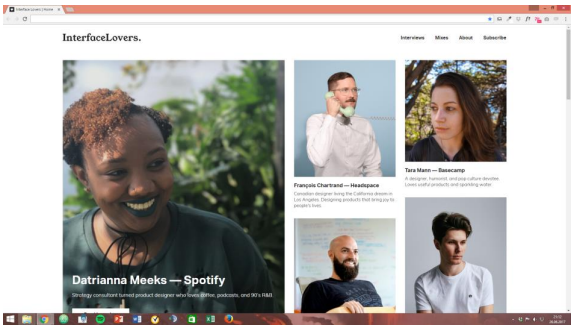
Stimuli 15 – Low complexity stimulus in information category



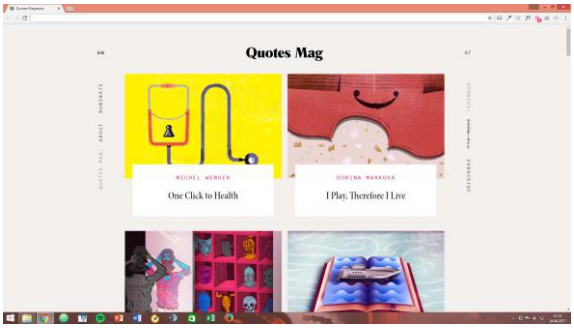
Stimuli 16 – Low complexity stimulus in information category (also used as stimulus in interview)



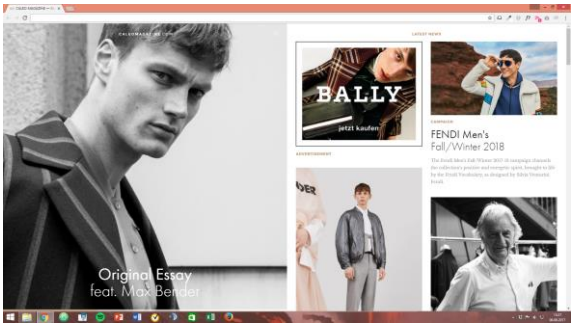
Stimuli 17 – Medium complexity stimulus in information category



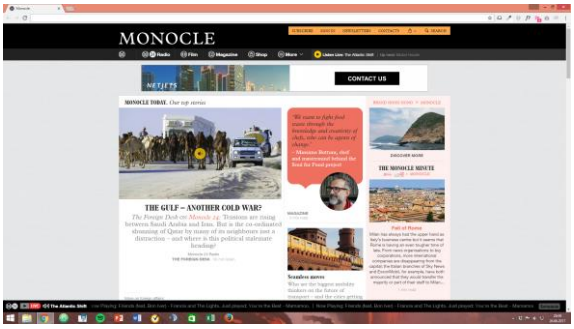
Stimuli 18 – Medium complexity stimulus in information category (also used in interview)



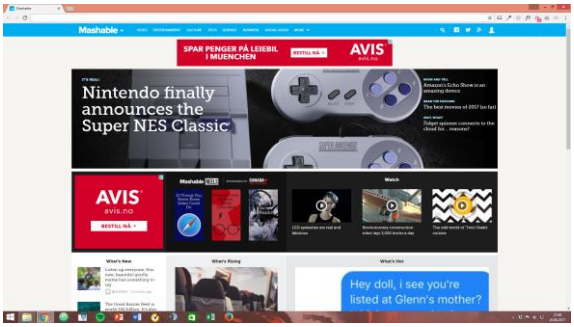
Stimuli 19 – Medium complexity stimulus in information category



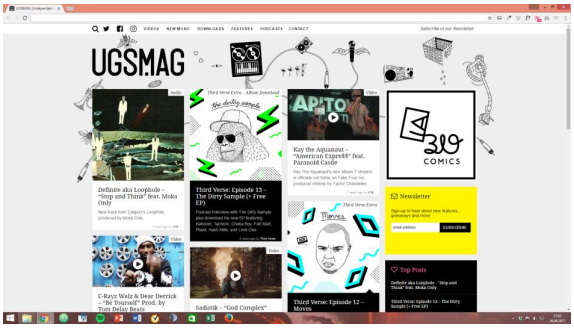
Stimuli 20 – Medium complexity stimulus in information category



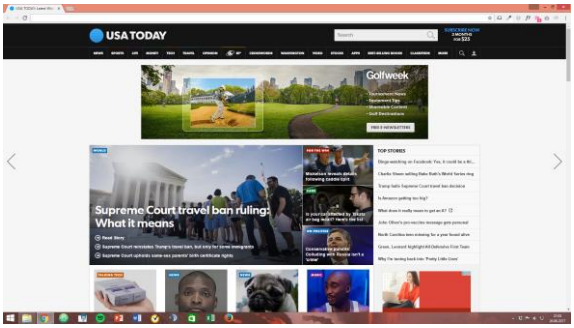
Stimuli 21 – High complexity stimulus in information category



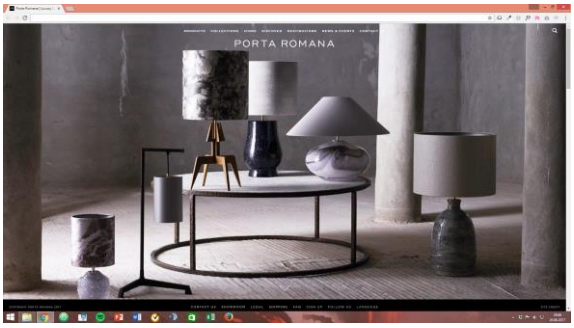
Stimuli 22 – High complexity stimulus in information category



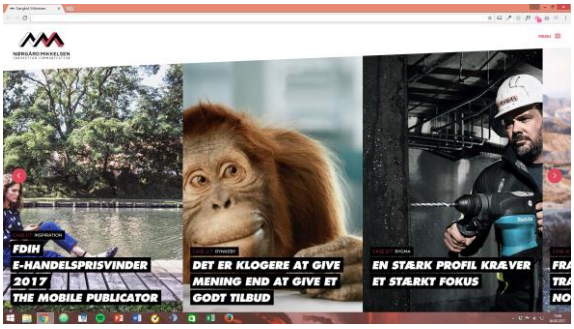
Stimuli 23 – High complexity stimulus in information category (also used as stimulus in interview)



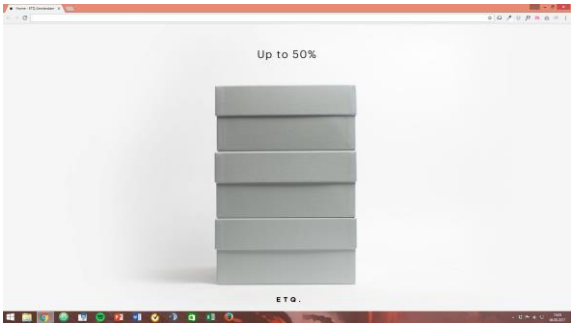
Stimuli 24 – High complexity stimulus in information category



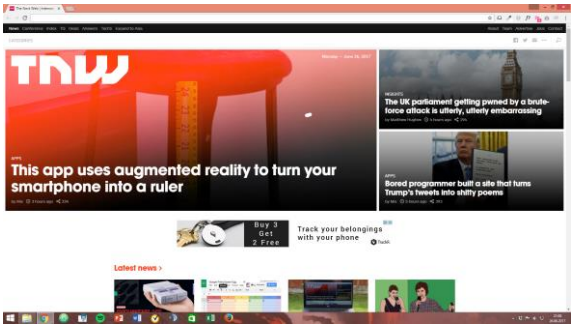
Pre stimuli 1 – Stimulus used in start for allowing participants to get used to the process



Pre stimuli 2 – Stimulus used in start for allowing participants to get used to the process



Pre stimuli 3 – Stimulus used in start for allowing participants to get used to the process



Pre stimuli 4 – Stimulus used in start for allowing participants to get used to the process