Investigation of the relationship between aesthetics and perceived usability in web pages

Raphael Kominis

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Mathematical & Computer Science

February 2014

Contents

List of Figures

List of Tables

Abstract

The main hypothesis of the thesis is that between two systems identical in functionality and usability, differences in aesthetics may positively influence users perceived usability.

To date, a narrow focus on the engineering aspects of aesthetics has adversely affected the scope and success of experiments, therefore previous work in the field needed to be revisited.

The thesis reviews literature and theory in usability and aesthetics, the latter from the point of the view of philosophy, theory, and application. It also explores the relationship between aesthetics, usability and user engagement; discusses a distinct new trend research that identifies a link between beauty and perceived usability of website interaction; and develops a pilot for an experimental methodology.

Based on conclusions from the review of the field of usability, two experiments where designed and carried out, an independent measures and repeated measures. The findings of these experiments confirmed the hypothesis that perceived usability was positively influenced by higher aesthetics.

Acknowledgements

There are far too many people to thank that helped me through the years of my PhD. They know who they are, but I will mention a few here. I would like to thank my supervisor for picking me up when her predecessor decided to quit academia. Her guidance has been invaluable through these years. If it wasn't for our meetings and her help I would had quit way before arriving to submission.

Maria Svoronou, who read the draft twice and helped me with correcting a lot of the text and making it readable.

Silia Vitoratou who with out which I would be at a loss with the analysis chapter. Her statistics insight was more than helpful.

My parents for supporting me morally and financially and having faith in me when I thought I would give up.

Nathalie Rendevski-Savaricas for giving me a job that helped me support my self while writing-up.

All my flatmates and close friends in Edinburgh and Greece that had to deal with me at my down periods when I thought that I was never going to make it. More so they made life these years a really pleasurable time. I will always think back to these years as the "the good years".

CHAPTER 1

Introduction

This Chapter is a short introduction to this document. It will present the domain of the research and its main hypothesis. In addition, a short summary of the overall document will be presented.

1.1 Motivation

For years usability has considered computers as tools to do a particular task. This work oriented and purpose-defined way of considering computers has led to usability becoming a field defining itself by a focus on functionality. ISO 9241-11 defines usability as: *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"*. Effectiveness is defined as the *"accuracy and completeness with which users achieve goals"*[3].

As computers are increasingly used to carry out many very diverse activities, ranging from work to leisure, like viewing movies, listening to music, shopping, browsing and so on, research started to take an interest in the User Experience (UE). In addition to

the strict limit of efficiency in a specific task, UE treats the use of an application or a website as a complete experience. Thus usability is viewed in a more holistic manner and looks at user engagement [33].

There are a growing number of researchers interested in the relationship between aesthetics and usability; Don Norman, Noam Tractinsky and Alistair Sutcliffe are some of the leading researchers on the domain. To this date their research on the subject, points to the direction of the existence of a relationship between the two; that aesthetics do play a role in usability [67, 91, 88]. Tractinsky goes as far as to title one of his papers "*What is beautiful is usable*"[91]. Don Norman's encompassing theory, for this, is based on emotions. A trio of emotional levels, *visceral, behavioural* and *reflective*, would be the mechanism that makes us like or dislike something.

However, there is some controversy in the field, as contradictory results have emerged. Hassenzahl suggests that there is no relationship between the two variables, aesthetics and usability; instead he proposes the notion of '*goodness*' in the scope of UE [33]. We can distinguish two camps: the "new usability movement" represented by Jordan and Green and the "*Funologists*" represented by Hassenzahl, Monk and Blythe amongst others. As cited by Hartmann [32], the first is aiming for a recosideration of usability in order to embody new user requirements while the "Funology" camp is proposing a science of fun. These trends will be reviewed in more detail in Chapters 3, 4 and 5.

In this new debate on usability, aesthetics and beauty are at the centre. Human Computer Interaction (HCI) as a field has become more and more interested in aesthetics and beauty, two areas that researchers traditionally looked at with suspicion. This suspiciousness goes as far as some researchers stating that even research in design icons products is unusable [78]. Nevertheless, as Hassenzahl states: "*Beauty mattered to them* (...Greek philosophers...) *as it still matters to every human. And as long as humans are essential elements in the study of Human Computer Interaction (HCI), to better understand beauty must be an important endeavour of the field.*"[33].

After reviewing previous work (see Chapter 5), we concluded that research to date

has its limitations and that there is a need to revisit experimental methodologies; we argue that a narrow focus on engineering aspects of aesthetics has impacted the scope and success of experiments to date.

The main hypothesis of this document is that for two systems identical in functionality and usability but different in aesthetics, user-perceived usability will be positively influenced by higher aesthetics. In other words, to paraphrase Tractinsky's title, "*If it is better looking is it perceived as easier to use?*". Our main proposal is to investigate what the link is between aesthetics and usability. This is presented in more detail in Chapter 2.

1.2 Outline of the report

This report contains 8 Chapters. We are going to briefly outline these Chapters here.

Chapter 2 Research questions This Section will further discuss the domain of the research; our main hypothesis is the main contribution of the thesis and our claim to novelty.

Chapter 3 Philosophy of Aesthetics, Literature Review A review of aesthetics from a philosophical point of view is presented. Different theories and viewpoints will also be presented. Aesthetics is a very old subject and has been debated since at least ancient Greece. There are almost as many aesthetics theories and viewpoints as there are artists and art philosophers. Aesthetics in one form or another, practical and/or philosophical, is a relevant subject to all cultures [20]. Chapter 3 will introduce a number of theories and then will present our own stand point. It is outside the scope of this document to try and present a fully comprehensive review of aesthetics. Nonetheless our purpose is to give a general overview of the field and provide a presentation of the main movements of aesthetics.

In order to organise the theories presented a taxonomy space will be presented for for an easier and better understanding of the field by the reader.

Chapter 4 Applied Aesthetics Applied aesthetics will be presented in this Section as a continuation of the previous Chapter but with a focus on the practical aspect of aesthetics. This Section presents some guidelines that designers use in their artefacts.

At the end of this Chapter we will present our position.The philosophical and practical debate of aesthetics makes it clear that there is not single answer to the subject of Aesthetics. There are many competing or comprehensive theories. As such we present our position and justify it.

Chapter 5 Past approaches to the problem Previous work is presented and reviewed. This Section examines research that is directly relevant to the field researched. Experimental design with a focus on the manipulation of the dependent variables is reviewed and issues are raised. These issues form the base for the next chapter, where our own experimental design is presented.

Chapter 6 Experimental Design This Section discusses our own experimental design. The Chapter is divided in two main Sections, our first experimental designs that reached various stages of pilot testing and the design of the main experiment. The first Section will detail the designs of the experiment and the results of the pilot testings. Conclusions and models are drawn from these pilots, which are then used to design the final experiment. The second Section details the factors taken into account.

Chapter 7 The experiments In this section is presented the design, methodology and the two experiments that where carried out for this document. To do that we will explain the tools used to gather the data as well as detail the results in their raw form and the manipulations applied to the results in order to draw our key

conclusions. The data is treated with various statistical manipulations to assess their statistical significance and and make them easily understandable.

Chapter 8 Discussion and Conclusions Our key findings are presented and critically discussed. Conclusions are drawn from the data gathered and presented to the reader. The results are discussed and some hypotheses are advanced. Further research is proposed.

Research questions and contributions

—Computing is not about computers anymore. It is about living.

Nicholas Negroponte [58]

2.1 Domain

The document explores the relationship between aesthetics, usability and user engagement. It evolves around a distinct new trend in usability research that identifies *beauty* as an attribute to interaction [67, 34, 86, 89, 88, 91]. Thus, this document will discuss and base itself on the joint domains of aesthetics, which includes the domain of applied aesthetics, and of usability.

2.2 Hypothesis

This thesis explores the hypothesis that aesthetics are an attribute of usability. When users visit a site, we believe that they are influenced by the visual aspects of the site and this reshapes their opinion of the usability of that site.

- Our main hypothesis is that for two identical systems in functionality and usability differences in aesthetics may positively influence the perceived usability of the site by a user.

- Several factors will have an impact on the positive influence of aesthetics on users perceived usability of a website.

  We are interested in the perceived usability of a website, not in its objective usability.

2.3   Contributions

We will develop an experimental procedure to establish whether there is a link between aesthetics and usability. This experimental procedure will be developed by drawing on work relevant to the subject of this report. We will start by critically evaluating approaches to the problem. This critical evaluation will be based on the literature review Section of this document. An experiment is developed and piloted. Pilot experiments were conducted before the final experimental design were drawn. The experiment itself is consequently conducted.

Due to the fact that this document is dealing with more than one domain, more so two domains that are very different - philosophy and engineering - language issues are arise. This document will attempt to explain some of these vocabulary issues.

2.4   Claim to novelty

The claim to novelty of this thesis consists of several positive results, which showed that usability was positively influenced by higher aesthetics; *easier to use upon first visit*, *GUS*, *helpfulness*, *tendency to find things around the website more easily*, *organisation of information on the system screens is clear*. We establish some raw positive links between aesthetics and usability, advancing the earlier work of other experiments.

Philosophy of aesthetics, literature review

Our first aim in this Chapter will be to review and define the concepts that we will be exploring in this thesis. This review serves as the basis for our experimental design. Chapters 3 and 4 will consider two aspects of Aesthetics: first the philosophical approach to aesthetics and then applied aesthetics, more closely related to web design considered in this work.

This leads us to our own position about aesthetics, which a is determinant for our experimental design. What we consider aesthetics to be is a key part of our experimental design. In an argument as complicated as that of aesthetics,which remains unsettled for two and half thousand years, our experiment needs a clear and concise position about what is beautiful.

Thus aesthetics will be the first notion that we will explore; we will give a brief overview of a number of philosophical approaches to aesthetics, art and/or beauty. We will not carry out an exhaustive review of philosophy of aesthetics, as the aim of this document is to present only the major philosophical movements and its representatives. In the history of philosophy of aesthetics, too long and complicated to be reviewed in

its entirety, there have been many opposing ideological positions; philosophical idealism considers beauty to be a property of the artefact while for materialists the purpose of the artefact indicates its aesthetic values and thus beauty has to do with the artist and the viewer as much as with the artefact. Universalist theories consider that the common biological resemblances of humans gives a common basis for what people like. In opposition, relativism stipulates that beauty is entirely in the eye of the beholder. Individualism and collectivism, two opposing sides of the last axis we are going to use, are the best way to explain fashion.

"*Taxonomy is always a contentious issue because the world does not come to us in neat little packages*"[27] and the same can be said of philosophy of aesthetics. We thus created a three-dimensional taxonomical space to help make the navigation and comprehension of this review easier. To use a space is a particularly good way to represent this taxonomy because no position fits neatly in just one end of the spectrum; theories are much more complex than just one position and an opposite. A space is thus a better way to represent the philosophies because it may include opposing ends of a spectrum but different positions do not have to be located at one extreme or the other:

Idealist - Materialist on one axis (section 3.1.1 p. 7 - Section 3.1.2, p. 11)

Universalist - Relativist on the second axis (section 3.1.3, p. 11)

Individualist - Collectivist on the third axis (section 3.1.4, p. 12)
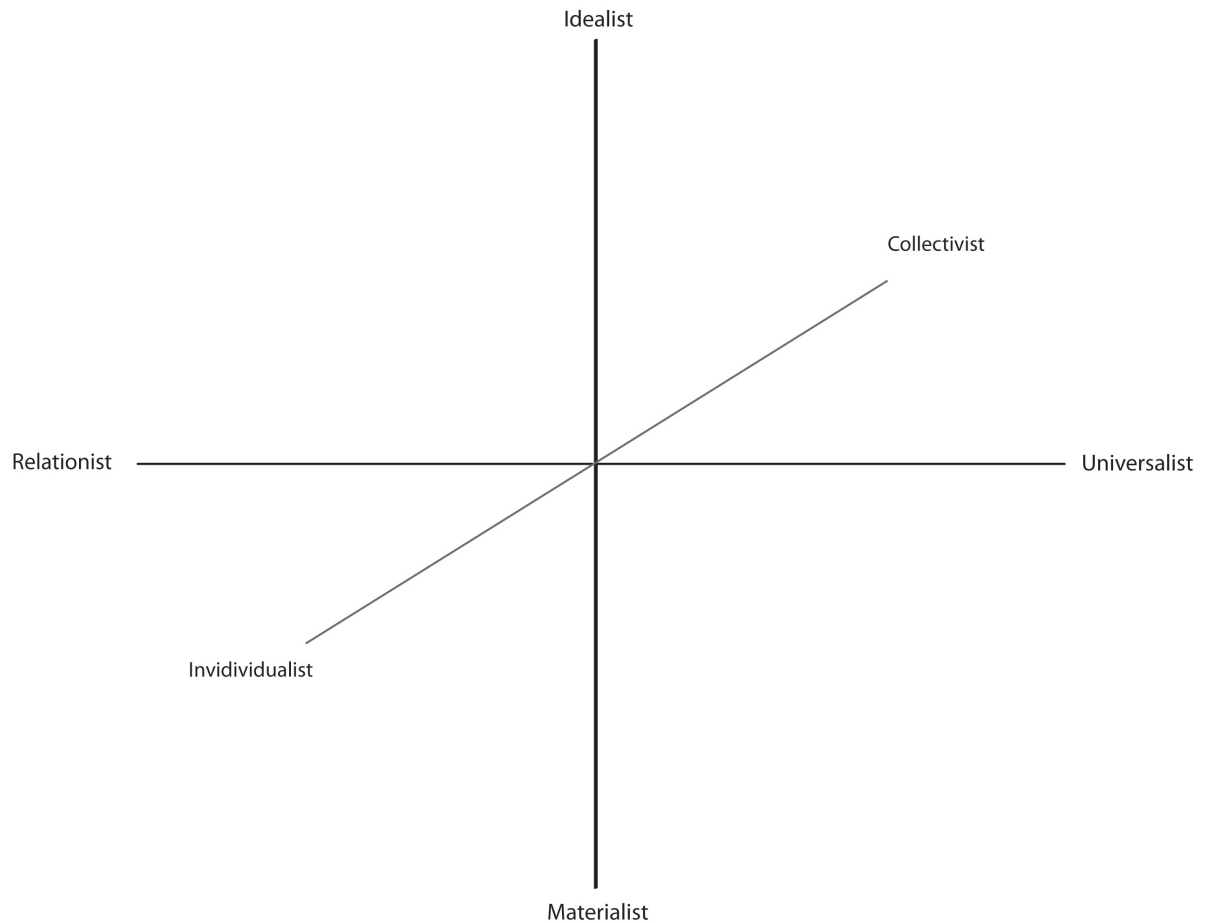
Figure 3.1: A representation of space to visualise the aesthetic theories

The taxonomy is there to help comprehend and give a sense of orientation in a complicated debate. It is a simplified space depicting only the philosophical positions presented in this document (see *figure 3.1*). Although these dimensions are not a definitiv description of the field, it is a simple way to categorise and to orient ourselves through this long and complicated debate.

After examining the philosophical space, we will present aspects of aesthetics that are closely tied to philosophy but are somewhat more practical. We will present some of the guidelines that designers use in their creative process and thus extract useful information for our experimental design.

We will address the field of semantics of design. A major part of design is to communicate or to facilitate the passing of information from the artefact [66], in our case a webpage, to the user. In this instance the word "design" has the meaning of

presentation. Therefore semantics is important and needs to be taken into account in our experimental design, as it explores the meanings given by different elements or the medium itself.

We will carry out a short review of usability standards. The main focus of this research concerns the relationship between aesthetics and usability, thus we review the current guidelines concerning usability.

We will present our own position on aesthetics. Having a specific position on aesthetics as well as beauty is an essential part of our experimental design.

Finally we will recapitulate and explain how our position, which varies from other theories, determines our experimental design.

3.1   Aesthetics

—Art is a human activity consisting in this, that one man consciously, by means of certain external signs, hands on to others feelings he has lived through and that other people are infected by these feelings and also experience them.

Leo Tolstoy, in Tolstoy and Maude, *What is art* [90].

The aesthetics debate has its roots in ancient times - Plato [72], Aristotle [70, 35], and so on. We believe that all humans make aesthetic decisions; to support this we quote Denis Dutton: "*All cultures display some form of expressive making of a kind European traditions would identify as artistic*" [20]. Furthermore, people have been preoccupied with aesthetics for a long time. The earliest artefacts found go back to the palaeolithic period 15,000 years ago (see Figure 3.2 paintings in Lascaux caves, France [11]). Yet to define beauty has proven difficult if not impossible; there is no universal agreement. To the present day there has been no theorist, psychologist, or aesthetician who has found *the* theory that encompasses all aspects of art, design and beauty. The aesthetics debate
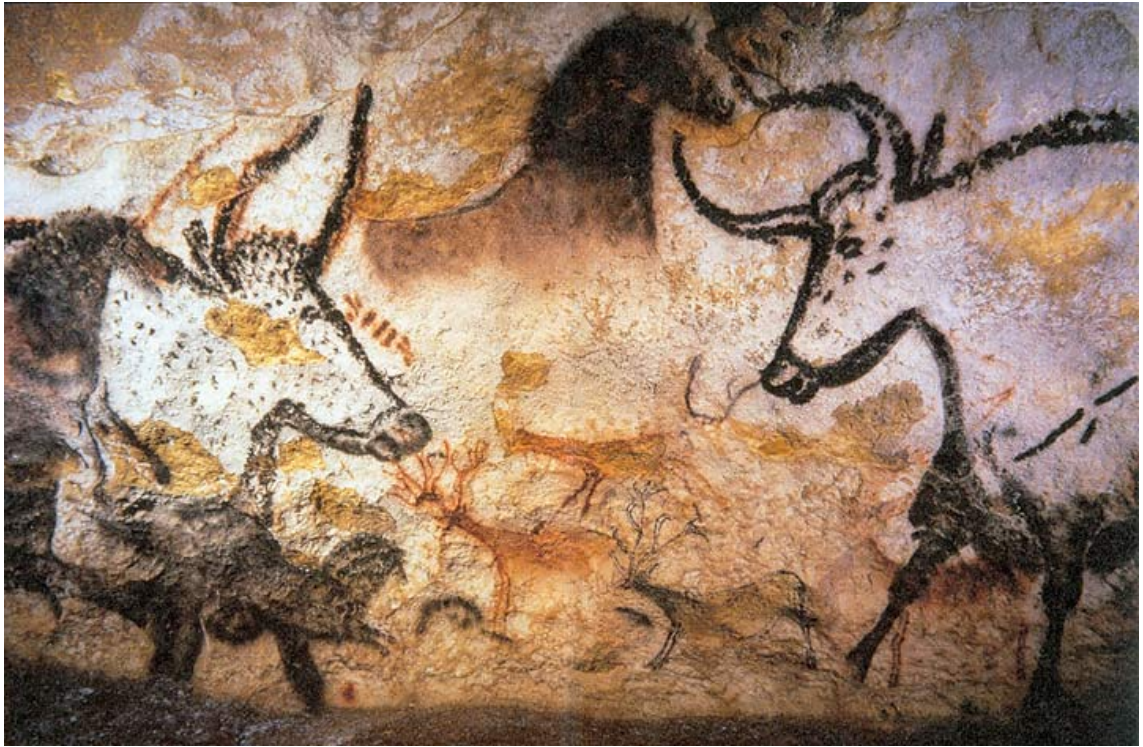
Figure 3.2: Paleolithic cave paintings at the Lascaux Caves in France, they are estimated to be 15000 years old (Bahn 1998)

has been approached through many different angles, philosophically and psychologically, mathematically, physiologically, by theories of evolution or experimentally through heuristics.

Immanuel Kant's philosophical approach is still one the most respected. For Kant beauty is a psychological process of which we become aware: "*the pleasing awareness of the harmony in the free play of our cognitive faculties*" [79].

We can occasionally, with some success, describe why we like something; but it is on very rare occasions, if ever, that everybody is of the same opinion. Finding a way, or a scientific procedure, to make something look good every time, to have the same success but to fit different conditions and produce different artefacts of the same aesthetic value, has not been discovered yet. For our experiment we would ideally need a way to represent aesthetics with a numerical score that one could analyse in order to measure the aesthetics of a design or artefact.

3.1.1 Idealists

We will start by introducing idealism as one of the main movements in aesthetics. We will present some of the aestheticians that have approached the subject from an idealist's point of view. Some of the major representatives of idealism would be Plato, Immanuel Kant, David Hume, George Berkeley; and so on. As an example of the complexity of the aesthetic question in philosophy we can say that even in the idealist position, there are differences; Plato's position for example, has a different approach than Berkeley or Hume's. His philosophical advances are fundamentally different in epistemology [52] but this part is beyond the scope of the present document.

The philosophical theory maintains that everything we know is through indivudual perception and thus is based on our minds and ideas. It takes the view that the 'external' world is undivided from the mind, consciousness, or perception. The idealist approach to aesthetics is that beauty is the ideal, the perfect, and we as humans will understand the beauty when we see it. In the *Republic*, Plato uses an allegorical story of people looking at the wall of a cave where the shadows of ideas are projected. These ideas are the reality and the projection is our perception. To climb out of the caves and look at the ideas themselves would be the job of the philosopher along with freeing the rest of the people in the cave [72]. There are many interpretations of Platos' writings and this is due to the fact that translation from ancient Greek is difficult. For the idealists, aesthetics of an artefact are independent of time, space, culture and the viewer. Beauty is transcendent and independent of the viewer and goes beyond individuals, their culture and personal taste. The person who looks at an artefacts for idealist philosophy is not important. The artefact's inherent properties make it beautiful independently of the opinion of the viewer, his culture and his experience. In that way we can say that the Mona Lisa by Leonardo da Vinci is beautiful whatever the hordes of tourists, who squeeze themselves in the Louvre to see it, think of it. What the viewer does is to acknowledge the beauty that is part of the artefact's properties.

Since we do not all agree on what is beautiful, the aesthetic experience becomes

something for the connoisseur, for the educated. We need to be educated to appreciate art, to appreciate beauty. Art then is not for everybody but rather for the person who has the knowledge and the education to understand and appreciate it. If a viewer does not appreciate the Mona Lisa it is a lack in the viewers part; the painting is beautiful no matter what this particular viewer believes.

Rules of art and design like the *golden mean* which is represented by the letter ϕ and *colour harmony*, could be considered an idealist point of view. We analyse ϕ and colour harmony in further details in Section 4.1 p. 16. We consider them as idealist because these rules are part of the artefact. If we assume that the relation of two forms in ϕ is beautiful, beauty is then an inherited property of the artefact.

Pythagoras, who had a great interest in ϕ, was an idealist, as were most of the artists and creators that we credit for using it along with other design rules. Pheidias, for example, the architect and sculptor of the Parthenon [28] (Figure 3.4 [39]) was one of the founders of classical idealism. These rules are considered to transcend culture and personal taste because they are a property of the artefact. Nevertheless this is not as clearcut as it appears, as one could defend the position that this is also a universalist theory, but this will become clearer further on when we introduce universalist theories.

3.1.1.1 Through mathematics

This Section will address a purely mathematical approach to the subject of aesthetics. There have been a number of mathematicians who have tried to measure, count, or otherwise put a number on aesthetics. Mathematics tries to conceptualise the real world through numbers, so by definition it has an idealist approach. The theory behind the practice is that beauty is there, it is a property of an artefact and subsequently we can measure it. If we can quantify beauty it means that beauty is transcendent of individuals, cultures and settings. If this is the case and we can quantify it on a scale in order to measure it, there would be a maximum figure. Since there is a maximum figure, which would be the ideal, it is an idealist theoretical position.
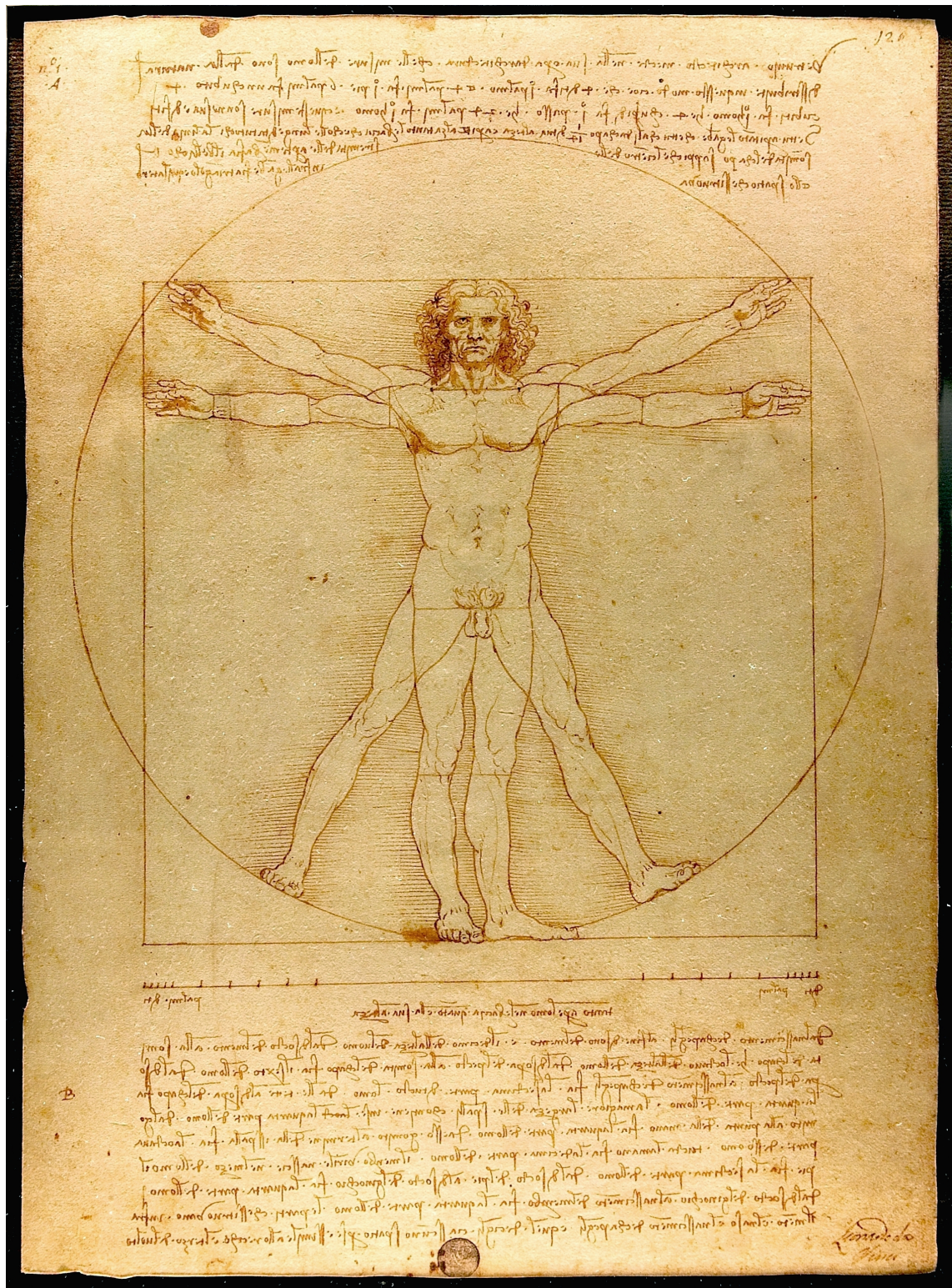
Figure 3.3: Man sketched having the perfect proportions according to Vitruvius, by Leonardo da Vinci.

Figure 3.4: Photograph of the Parthenon in Athens with the golden mean ratios superimposed (Jordan 2000)

Birkhoff [13] was a successful and well known mathematician who set out to find a mathematical measure of Aesthetics. He came up with this equation:

$$M = \frac{O}{C}$$

Where M is Birkhoff's aesthetic measure, O is order and *"depends on geometrical relations among identifiable segments of an evaluated object"* and C stands for complexity, *"the number of localities our sight will spontaneously rest on"* [83]. As Birkhoff states: *"Even in the most favorable cases, the precise rules adopted for the determination of O, C and hence of the aesthetic measure* M, *are necessarily empirical. In fact, the symbols O and C represent social values and share in the uncertainty common to such values"* [83].

Birkhoff seems not to take into account semiotics, emotions, culture and personal taste. He acknowledges that art has been approached from a *hedonistic*, *pedagogic* and *mystic* points of view. Hedonistic refers to the pleasurable aspects of art, pedagogic to the educational capacity of art and mystic to *"connotative or occult formal elements of order"*. But for Birkhoff the *"sound scientific"* point of view would be that none of them is or has a more important or dominant role in art. He seems to dismiss previous

nonanalytical theories, by declaring that they are approaching aesthetics from a more philosophical point of view. But at this point we see a controversy; for Birkhoff the variables C and O are empirical and social values. So if we cannot mathematically and precisely determine C and O the mathematical equation can more or less give us any result we might want by just changing C and O. This is a contradiction to the claim of a mathematical way of measuring aesthetics.

Birkhoff's book has been used as a point of departure for creating a more general theory. One of the most famous researchers who used his theory was Max Bense and his group of literary theorists in Germany in the 1950's. According to Scha and Bod, Bense with his group developed the theory of *information aesthetics* [79].

In 2007 a paper from University of Girona, Spain, tried to conceptualize Birkhoff's aesthetic measure using Shannons' entropy and Kolmogorov complexity equations: "*The initial uncertainty, obtained from the Shannon entropy of the repertoire (palette), is transformed into algorithmic information content, defined by the Kolmogorov complexity of the image*"[76]. It is partly on Birkhoff's and Bense's work, that Jaume Rigau et al. based their research. The paper applied the method on several works of three artists, Mondrian, Pollock and van Gogh. This treatment of the images, taking a picture of it and running an analysis that compares pixels it does not make sense according to Birkhoff, since O and C represent social values and not just simple changes of colours or stroke. In a painting, there is also texture, just to name one extra factor. Even a change in colour or shape does not automatically mean that it will be recognised as a "*locality of interest your sight will spontaneously rest on*" [13].

The papers that we have reviewed were expanding the calculation methods used based on Birkhoff's equation but none gave results or approaches as to testing the validity of the equation. [76, 83].

Scha and Bod suggest that developing and applying algorithms to images by analysing pixels does not stand to reason. This is "*because for most images encountered in practice, a construction out of adjacent discrete elements is not the perceptually relevant analysis*"

[79]. The eye of the user does not see pixels, it sees pictures. That is the point of having screens with high resolutions, to be able to make the pixels so small that they are not noticeable. As for the purpose of finding if something is universally aesthetic, from the human point of view, analysing pixels is probably not the right way. Reducing the aesthetic experience of looking at a van Gogh painting to a set of a few hundred thousand pixels like Rigau et al. are doing seems to lack something. For a digital image is not an actual description of a work of art, it lacks the texture to give but just one example.

But Birkhoff was not the only one who has tried to put a number on aesthetics in order to find a canonical approach to measure aesthetics numerically. Two more papers claim finding a numerical approach to aesthetics combining multiple factors into the calculation. While order and complexity is taken into account a further six factors are calculated for by Zain et al. [96] and twelve by Ngo et al. [59]. The two papers claim that their results from their applications are quite close to the results given by human subjects that reviewed the same web pages or artefacts.

The aesthetic measures developed are: "*balance, equilibrium, symmetry, sequence, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, rhythm and order and complexity*". 'Order and complexity' seem to echo Birkhoff's formula. According to the paper, *order* is the sum of all other measures counted. The inverse pole of order is *complexity*. So the result is a scale that has order on one side and complexity on the other.

$$\text{Order\_Complexity} = \frac{\sum_i^5 M_i}{5} \quad \in [0, 1]$$

Zain et al. developed an Aesthetic Measurement Application. The application measures web page interface aesthetics. It is mainly based on the findings of Ngo et al. although it only uses six elements instead of twelve: balance, equilibrium, symmetry, sequence, rhythm and order and complexity.

There are some parallels to be drawn between Birkhoff's equation and the equations of Ngo et al. and Zain et al. Both calculations and theorems evolve around the order

and complexity. The main difference is that Ngo et al. calculates O differently.

We are unsure as to the success of these two papers on finding a universal canonical numeration of aesthetics. Moreover a computational model of assessing aesthetics does not take into account, or even introduce the semantic dimension. And the aesthetic experience does not only comprise the perception of gestalts, shapes, but also the meanings that are given by the viewer to what they are perceiving [79]. We believe that cognition is a very important element in aesthetic experience.

3.1.2 Materialists

Materialism as a point of view holds that everything is matter, physical; and matter has its own laws, its own restrictions, so it also invites comparisons. Matter includes ourselves, our thoughts and our consciousness. It is very closely associated with physicalism and in some aspects the two terms are interchangeable [84]. Materialist aesthetics are divided into a number of different schools of thought, some of which are very politically engaged.

Aristotle is usually classified as one of the first materialists. He refers to aesthetics, not in the modern meaning but in the classical meaning, of being more related to perception. "*In Poetics 4 Aristotle says that human beings have a natural disposition to engage in and take pleasure in observing, imitation.*" according to Heath[35]. So for Aristotle all art is imitative. Furthermore as Parker has observed "*he believes very strongly that the intellect has a major influence on the emotions*" [70]. As such aesthetic pleasure has to be moral too in order to be pleasurable. Immoral pleasures are not good pleasures and as such are not pleasures.

One of the leading schools of materialist aesthetics is strongly tied to the Marxist camp. While Marx himself did not write much on the subject, we can see a continuity from Aristotle's aesthetics to Marxist aesthetics: "*Its originators, Marx and Engels, did little more than indicate to materialist dialectics the wide range of possibilities in this area*" according to Benjamin et al. [12]. But they never elaborated further. One of the Marxist views about art is that it has to have a political perspective according to

Burnham [19]. Further more for Marxists the debate did not end at the finished artefact and its sociopolitical implications. The production process and method was of interest as well; the way art was produced is also being analysed and taken into account, as explained in Wolffs' *The social production of art* [94].

Two important schools of thought in marxist aesthetics were socialist realism, developed in the Soviet Union and the Frankfurt school, that united a number of different theorists under a common base. In *Little Philosophic Dictionary*, Resenthal and Moiseevich's [77], a Soviet dictionary, we find for the word "*aesthetics*" explained in a clearly polemic language. The definitions are clearly materialist and Marxist and represent the position the Soviet Union was holding on aesthetics at the time of writing. The authors accuse the idealist position, of not being up to the task of explaining aesthetics "*scientifically*" and showing how it works. Diderot is presented as one of the major representatives in the "*fight*" in art between idealists and materialists.

The Soviet position on what constitutes good art has a set of rules. Not all of them are mentioned but those that are listed are as follows: "*The relationship of art with reality, the faithful copy of reality, the commitment to the idea, the ability of the artist to foresee and predict the new that comes to replace the old, the worn out, to help, with his art, (the new) to become standard, to be a fighter and not a philistine; the homogeneity of the form with its meaning etc.*". Clearly an ethical dimension is given to the aesthetics question and art is seen as a tool to educate.

The basic trent of the marxist approach, based on the philosophical framework of historical materialism[1] is that aesthetics must be firmly located in specific societies and cultures. This adds an inherently political dimension to it, given the centrality of class conflict in Marxist social analysis.

The claim given by the authors of the dictionary is that the materialist view developed by Soviet philosophers has rendered aesthetics to a scientific process: "*As for the evaluation of Soviet art, a decisive factor is to know to what extent an art piece helps*

---

[1] "*Historical materialism encompasses a wide range of approaches to the analysis of culture and society that variously have their origins in the work of Karl Marx (1818-1883)*"[12].

societies struggle, to the institution of communism, the education of the masses to the communist ideal, the development of Soviet patriotism and the feeling of national pride"

Furthermore Marxist aesthetics take a sociological approach to art. The understanding of art is closely linked to sociology. Edgar, in his *introduction to Adorno's aesthetics*, points out that there is a composite relationship between art, aesthetics and society. Art is a product of some particular social process but at the same time, in order for art to be able to criticise society, it must be independent of society. The paper concludes the following: "*Aesthetics is not, as such, reduced to sociology, but as an area of inquiry is justified by sociological insight*" [21].

3.1.3 Universalist, relativist

Universalist theories need not be based on the idealist position; they can also start from the fact that humans, as a species, are more or less alike. From some perspective, humans have relatively small genetic variations and all have more or less the same needs. For example, there is no human that can fly without the help of some flying device, as there is no human that can survive underwater without the help of scuba gear. Another example is that we more or less need the same nutrition —but and here is the interesting part— cuisines differ. The example of cuisine is perfect to show how different groups have devised different cultures from the same bodily needs. Thus it would be reasonable to hypothesise that we could find general trends of things we like that are not culturally mediated.

Such universalist theories would consist of, for instance, psychological theories about primitive emotions [22] — emotions that are considered as being universal. These would be emotions that all humans have in common as a result of a their common evolutionary history. Our more complicated and personal emotions would be constructed on those primitive emotions, although it seems that there is no consensus on what primitive emotions are. There is a number of proposals on the subject of primary, or basic emotions. There is even doubt about the scientific validity of establishing such a set of

emotions that underlie all emotions.

A brief aspect of how much difference there is in opinions of what constitutes "basic emotions" is shown in Table 3.1 (page 13) founded in "*What's Basic About Basic Emotions?*"[69]. Some of the emotions are reported by more than one researcher, but there does not seem to be any consensus on a list of basic emotions. It seems that some emotions appear more often than others in the table. As a result, there must be somewhat more consensus on some of them than others.

Paul Ekman [22] defines basic emotions as follows:

"*Characteristics which distinguish basic emotions from one another and from other affective phenomena*

1. *Distinctive universal signals*

2. *Distinctive physiology*

3. *Automatic appraisal, tuned to:*

4. *Distinctive universals in antecedent events*

5. *Distinctive appearance developmentally*

6. *Presence in other primates*

7. *Quick onset*

8. *Brief duration*

9. *Unbidden occurrence*

10. *Distinctive thoughts, memories, images*

11. *Distinctive subjective experience*

".

Reference	Fundamental emotion	Basis to action
Arnold (1960)	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness	Relations to action tendencies
Ekman, Friesen, & Ellsworth (1982)	Anger, disgust fear, joy sadness, surprise	Universal facial expressions
Frijda (personal communition, Sept 8 1986)	Desire, happiness, interest, surprise, wonder, sorrow	Forms of action readiness
Gray (1982)	Rage, terror, anxiety, joy	Hardwired
Izard (1971)	Anger, disgust, distress, fear, guilt, interest, joy, shame, surprise	Hardwired
James (1884)	Fear, grief, love, rage	Bodily involvment
McDougall (1926)	Anger, disgust, elation, fear, subjection, tender-emotion, wonder	Relation to instincts
Mowrer (1960)	Pain, pleasure	Unlearned emotional states
Oatley & Johnson-Laird (1987)	Anger, disgust, anxiety, happiness, sadness	Do not require propositional content
Panksepp (1982)	Expectansy, fear, rage, panic	Hardwired
Plutchik (1980)	Acceptance, anger, anticipation, disgust, distress, fear, sadness, surprise	Biological process
Tomkins (1984)	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	Density of neural firing
Watson (1930)	Fear, love, rage	Hardwired
Weiner & Graham (1984)	Happiness, sadness	Attribution independent

Table 3.1: A selection of lists of "basic" Emotions, from Ortony and Turner

For example disgust is included amongst them. These emotions have some very basic functions. Disgust is there to keep us from eating and touching potentially harmful things like rotten food and so on.

From this point of view we might theorise on the mechanism that people use to understand and evaluate art and beauty. If there are basic emotions, it would not be so farfetched to hypothesise that people use the same processes more or less to evaluate what they like or dislike in art and design. Nevertheless, such a theory is of little use to us when trying to find out what is beautiful. Basic emotions according to the theory are just that: *basic.* Other emotions, much more complicated, start piling up on top of them. And the way we judge art has nothing simplistic. Furthermore, hypothetically the processes might be the same but also very complicated. In addition the processes might be the same, but, depending on the individual, the outcome is not always the same .

The relativist position on the other hand advocates that beauty is entirely in the eye of the beholder. We, as viewers, attribute beauty to what we see. Beauty is not an attribute of the artefact but is given by the viewer. There is no universal beauty and aesthetic judgements are made on a personal basis. "*Kant insists that universality and necessity are in fact a product of features of the human mind (Kant calls these features common sense) and that there is no objective property of a thing that makes it beautiful*"[19]. Each individual depending on the personal preferences attributes beauty to different things. Personal taste here is the predominant factor. The criticism of such a position is this: If there is no right or wrong in aesthetic judgements then the theory is not subject to criticism as it can be neither right or wrong [97]. As such relativism is somewhat frowned upon by aestheticians that see it as an easy way out of the argument. Furthermore absolute relativism is frowned upon because it abolishes the concepts of masterpieces and classic art. It reduces aesthetics to a mere "I like it" or "I don't like it". It also abolishes, the social concept, the image, of the artist as a person who stands apart from society and has the ability to critique it. The art "expert" and the art "critic"

as social entities loose their meaning; they are then either people that can recognise trends or simply frauds.

3.1.4 Individualist, collectivist

In this Section we are going to discuss two other positions that stand more or less opposite of each other: individualist and collectivist. The notion of fashion will be of particular importance in this Section. Fashion will be mentioned, further down in later in Section 6.4.2 (*page51*). These two positions propose an explanation for fashion while other positions do not seem to be able to deal with fashion that well. For idealists for example if something is beautiful now it should also be beautiful later on. But fashion contradicts this. What was fashionable to wear some years ago is not anymore and so on. So how do we account for fashion? Can it be just a need for novelty?

Individualists believe that personal taste is the predominant aspect of aesthetic judgement. Our personal taste, that could coincide with other people's taste, is something that belongs to the individual.

A Collectivistecdl point of view would be that it is culture that mediates our decisions. Depending on where we come from, where we grew up, our social class and surrounding environment we make culturally mediated aesthetic judgements. It would be a way to explain fashion as an aesthetic issue.

A good indication of the aesthetics of fashion being a controversial subject is the existence of the debate on: "*is fashion art?*". Art critics, fashion designers and sociologists do not seem to agree on the subject. A paper by Sung Bok Kim [43] has some details on the debate, including Remy G. Saisselin[2], Michael Boodro[3], just to name a few of the debate's protagonists. They agree that both fields, fashion and arts, have a lot in common and that the link between them is becoming stronger; but they do not agree about fashion being an integral part of art.

Fashion is a cultural product of the society that surrounds us. Idealist positions

---

[2]Scholar of eighteenth-century French art
[3]Art ctitic

cannot account for fashion in any other way than by considering that people are making a mistaken "aesthetic judgement", or that they have "bad taste". For Kant, fashion has nothing to do with judgements of taste. He considers it as pure imitation and as such *"the opposite of 'good' taste"* [30]. But nevertheless both Kant and Simmel, whom we are going to introduce in more detail later on, thought that trying not to follow fashion was completely futile.

From a sociological point of view fashion seems very interesting. Georg Simmel, one of the "fathers" of sociology, published three essays on fashion; for him there is a "duality" in every person, between the individual and the group he or she belongs to. Fashion is analysed as a social class phenomenon. Simmel sees it as a *"practical conflict representing socialism on the one hand or individualism on the other, [...] we have always to deal with the same fundamental form of duality which is manifested biologically in the contrast between heredity and variation"*. Heredity here can be understood as the environment the individual has grown up in. Variation is the individuality that the person brings to that environment. *"Fashion is the imitation of a given example and satisfies the demand for social adaptation"*, he further states. We imitate our parents and peers and in order to be accepted by our group and peers we need to follow the rules that are set. Wearing a suit at the office would be a good example. *"[...] Fashions differ for different classes - the fashions of the upper stratum of society are never identical with those of the lower; in fact, they are abandoned by the former as soon as the latter prepares to appropriate them"*, he says, explaining the role of fashion as a social delimiter. For Simmel, the lower social classes are always trying to imitate the higher ones. The constant change is not then only by will to change but also by necessity, to keep the demarkation between the classes clear. *"Thus fashion represents nothing more than one of the many forms of life by the aid of which we seek to combine in uniform spheres of activity the tendency towards social equalisation with the desire for individual differentiation and change"* [81]. Fashion thus fulfils not only a social role, a sort of mark up or of delimiter of the social classes, it also fulfils the need for novelty

and change of the individual in the social class.

For Simmel, individualism and collectivism not two conflicting theories but are an integral part of the duality of the individual. His view is that these forces exist inside each one of us, we belong or desire to belong, to a group but we also desire to be an individual inside that group.

3.1.5 Norman's emotional design

Don Norman has a very pragmatic position about how we judge aesthetics. His hypothesis is mainly based on design and emotions. This position would be in the space between individualist and collectivist, as it is taking into account cultural and personal taste, but also between universalist and relativist as the proposed explanation takes into account issues that we have all in common as well as personal taste.

In order to better understand design, Norman [67] identifies three different levels: visceral, behavioural and reflective.

Behavioural design is the part where HCI (Human Computer Interaction) comes into play; where we can formulate rules for designers to follow. It is the aspect of design on which the HCI community has centred its research.

Visceral design has to do with visual aspects. It deals with beauty and aesthetics. There are no firm rules but many timeless guidelines. While no scientific methods exist, the guidelines, golden ratios, colour harmony and so on, are widely used around us. Some of them are almost mystical. Figure 3.4 p. 10 is an example of use of golden ratios in history. It shows the Parthenon with golden rectangles superimposed on top of it. Golden ratios are extremely old and are found in many different cultures, for example in the Great Mosque at Kairouan build the 670 AD[15].

Reflective design has to do with users mentality, culture and individuality. Designers here have to take decisions based on their intuition [34]. Norman goes as far as

to say that the reason why product developers ignore HCI experts is because the latter are only interested in behavioural design, ignoring all other aspects that are equally important in the success of a product.

For Norman, emotions play a determining role in how we judge things around us. It could explain individuality in taste but also the similarity between different people as well as cultures. In art history there have been many rules and guidelines that have survived in art since ancient times, including synthesis theories, colour harmonies, rhythm and so on.

Here too we see a reconciliation between group and individual as in Simmels' theory. The *behavioural* aspect refers to the group; as argued above, all humans are similar in some aspects. *Visceral* is the idealist side of the theory; it is about the factual aspects of the artefact. Lastly, *reflective* is a combination of the two; it has an individualistic aspect but also a collectivist aspect. In this instance, mentality, or personal taste, is the individualistic aspect and culture the collectivist aspect.

3.2 Recapitulation

In this Chapter philosophical standpoints on aesthetic that are important to this research, were introduced. To do that we have devised a philosophical space, whose sole purpose is to make this exploration easier to understand. The space comprises three axes (Figure 3.1 on page 6):

Idealist - Materialist dimension

Universalist - Relativist dimension

Individualist - Collectivist dimension

The idealist position gives more importance to the *ideal* than to the perception of reality around us. We talked about the mathematical approach to aesthetics by Birkhoff

and how the formula he advanced to calculate aesthetics has been used in more recent computer science research.

Materialist theories evolve around the main idea that everything in the world consists of matter. Marxist theories were then reviewed and the historical division between Soviet aesthetics and the aesthetics advanced by Adorno and others in Europe, is mentioned .

The universalist - relativist dimensions was then presented. Universalist theory considers that similarities in art appreciation come from the fact that humans are similar in many aspects. Based upon the fact that human beings are similar on a genetic level, universalist theories consist of psychobiological theories like *basic emotions* that would be universally common.

For the Individualist - Collectivist dimension we presented the theories of Georg Simmel that connect both stances. For Simmel there is a duality in each person: we want to be unique but at the same time we need to belong to a group.

We lastly presented a pragmatic position by Don Norman, who advances that we appreciate an object in a mix of three emotional levels, *behavioural*, *visceral* and *reflective*.

CHAPTER 4

Applied Aesthetics

> —*Design is that area of human experience, skill and knowledge which is concerned with mans ability to mould his environment to suit his material and spiritual needs*

Archer B, The Need for Design Education [10]

This Section of the document will present aspects of design relevant to this work. It will present some more practical approaches to the design process. These are what designers commonly use in the creative process. This Section links the previous Section that dealt with more philosophical and theoretical approaches to the practicalities of the the design process. In this Section some of the examples used are not from the web but from different artistic fields. This is because web design is still a very young field. There are no landmarks of design in web design. This will be further discussed in Section 4.2.2 (page 21)

4.1 Some rules of design from designers

In this Section we are going to present some of the rules or guidelines designers use to create, compose, frame, present, design, their artefacts. Thesea are used as tools when needed.

One of the first things that art schools teach is to learn the rules and then decide if they need to be broken in order to serve the purpose of the artefact. Most the guidelines or principles, are common across the arts. It does not matter if it is a photograph, or a painting or a webpage being framed. As Lidwell et al. say in *Universal Principle of Design*; "*The best designers sometimes disregard the principles of design. When they do so, however, there is usually some compensating merit attained at the cost of the violation. Unless you are certain of doing as well, it is best to abide by the principles*" [50]. The message is the same in O'Brien and Sibley's in *The Photographic Eye*: "*Break as few rules as possible. For any photograph, try to stick to as many of the established rules as you can. [...] If you can't get the results you want, then — and only then — break another rule. [...] If, instead, you break all the rules at once, you're likely to end up with chaos that no one else will understand, or want to understand*" [68].

4.1.1 Symmetry, asymmetry, proportions and rhythm

In the introduction of their book *Photographic Composition* Grill and Scanlon emphasise that photographers need to know classic design in order to compose their photographs properly. On the question of *balance* they state: "*Complete symmetrical balance within a composition is usually uninteresting. Only when the actual point of the photograph is to emphasize the condition of symmetry should the composition be precisely symmetrical*"[29]. On the contrary Lidwell et al. [50] state that symmetry has three fundamental types: *reflection, rotation* and *translation* and that symmetry has been long associated with beauty: "*Symmetry is the most basic and enduring aspect of beauty. Use symmetry in design to convey balance, harmony and stability. Use simple symmetrical forms when*

recognition and recall are important and more complex combinations of the different types of symmetries when aesthetics and interestingness are important" [50].

Symmetry in art and design is one of those rules that is a mere guideline. There are advocates of symmetry, like the ancient Greeks (Figure 3.4 of the parthenon on page 10), or the Egyptian pyramids, (Figure 4.1a, page 18); we see examples in modern architecture like the glass pyramid in the Louvre museum (Figure 4.1b). The next step is not a counter example but something in between, where the architect decided to break the symmetry but not completely, (Figure 4.1c) of the Glasgow School of Art by Charles Rennie Mackintosh for example. A complete counter example to the where the architect decided to completely break the symmetry is the building of the Berliner Philharmonic, (Figure 4.1d). These examples are from architecture but the argument we make is clear: symmetric or not the artefact can have a claim to beauty.

Rhythm would be the most difficult term to explain in visual arts. Rhythm come from the Greek word ρυθμός (rhythmos). Rhythm exists in most art forms, particularly in music. In music it is very easy to understand it but when it comes to something visual, rhythm is more difficult to explain. It is the regularity with which forms, elements or colours are organised in the picture without always implying a form of symmetry. It creates a flow and a sensation of movement. Rhythm in practice is the organisation of elements, a pattern in the picture that makes the eye of the viewer jump from one point to another inside the frame.

(a) Photo of a pyramid in Egypt.

Photo: Khalid Almasoud

(b) Glass pyramid in front of the Louvre in Paris.

Photo: Glenn Crouch

(c) Glasgow School of Art.

(d) Berliner Philharmonic asymmetrical building.

Photo: Andreas Levers

(e) Painting by Jackson Pollock,
One: Number 31, 1950.

Photo: Wally Gobetz

Figure 4.1: Examples in architecture and painting, varying from symmetrical to having elements that break the symmetry, to complete asymmetry. The painting is an example of complete asymmetry but a good example of rhythm.

4.1.2   Proportions, the golden mean, the Fibonacci sequence

Proportion is the relationship between two or more related elements; in our case parts of an artefact. We already mentioned the *golden mean* in Section 3.1.1, page 7.

"*A rational number is one that can be put in the form of* $\frac{h}{k}$*, where* h *and* k *are integers with* $k \neq 0$*. Real numbers like* $\sqrt{2}$ *which are not rational are said to be irrational*" [65]. The *golden mean* is a ratio that can be found in nature and in many artefacts. In maths its symbol is ϕ. The ratio is expressed by this equation:

$\phi = \frac{1+\sqrt{5}}{2} = 0.6180339...$

and is one of more well known irrational constants like π and e.

Maybe one of the most famous examples of the *golden mean* is by Leonardo da Vinci *The Vitruvian Man* (Figure 3.3 page 9). Da Vinci here sketched a man with the perfect analogies according to Vitruvius. Others based their theories on this ideal analogy. Pythagoras or at least the Pythagorean school believed that things shaped using the golden mean would be more attractive.

The golden mean ϕ is one of those numbers that some people think are hardwired in the human brain as being a standard of beauty. Matila Ghyka [26] advocates that the golden mean and other elements like the pentagram are essential in art and life. His work has been highly criticised though as unsubstantial and non-scientific [38]. There is surely somewhat of a mystical side in what he writes[1]. There is a great polemic as to whether the golden mean has any aesthetic value at all see Green [28] for example. The use of the mean is even debated as to if it was used on ancient monuments like the Parthenon but this is beyond the scope of this document.

The great polemic is presented by Green in his paper "All That Glitters: A Review of Psychological Research on the Aesthetics of the Golden Section" [28]. According to him, first scientific study in psychology was done by Fechner [23] in 1860. But the results have been contested or supported numerous times ever since then by people that

---

[1]Ghyka attributes to ancient Greek architects mathematics that they could hardly have accomplished; i.e. $\sqrt[4]{\phi}$ which needs logarithmic calculations studied by Neper at the beginning of the 16$^{\text{th}}$ century [38].

were for or against the golden mean idea. Green in his review concludes that the golden mean holds some aesthetic value but it is a very fragile one.

Two more theories concern the *Rule of Thirds* and the *Fibonacci sequence.* The Rule of Thirds is a way to frame the subject taken in pictures used widely by photographers and cinematographers just to cite two categories of artists. The rule of thirds is a derivative of the golden mean; it is easier to mentally divide a frame in three than to find $\phi = 1.6180339...$ while taking pictures on the field.

The Fibonacci sequence is a series of numbers that are in the following succession:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...

$F_n = F_{n-1} + F_{n-2}, ...$

The relationship between two sequential numbers in the Fibonacci series is tending to be closer to ϕ: for example the two last numbers in the sequence above gives $\frac{144}{89} = 1.617977...$; which is very close to ϕ. Thus when designing something it is easier to use the Fibonacci series to choose two proportions than to try and Figure out the proportions using ϕ. Furthermore the Fibonacci sequence consists of integers, which is very convenient for digital media artists since pixels are not dividable.

4.1.3   Colour harmony

Colour theory for the arts explains, how do colours mix and what visual impact do certain combinations of colour have on viewers. The most widely known representation is Isaac Newtons colour wheel, which is a basic representation colour theory (Figure 4.2 page 20). In colour theory the colours are divided in Primary colours, red, yellow, blue, (Figure 4.2b). By combining one primary colour with its next one we get secondary colours, (Figure 4.2c). Then by combining, a primary colour with a secondary colour, you get a tertiary colour, (Figure 4.2d).

Combining colours from different categories has different effects. Colours that sit opposite of each other in the colour wheel will give a high contrast, and are called complementary. Colours that are next to each other are called analogous colours, and

usually one of them dominates the other. The colour combinations are called colour schemes. There is a plethora of software programs online and offline that can assist designers in choosing colours according to the scheme they want to use.

(a) The colour wheel

(b) Primary colours

(c) Secondary colours

(d) Tertiary colours
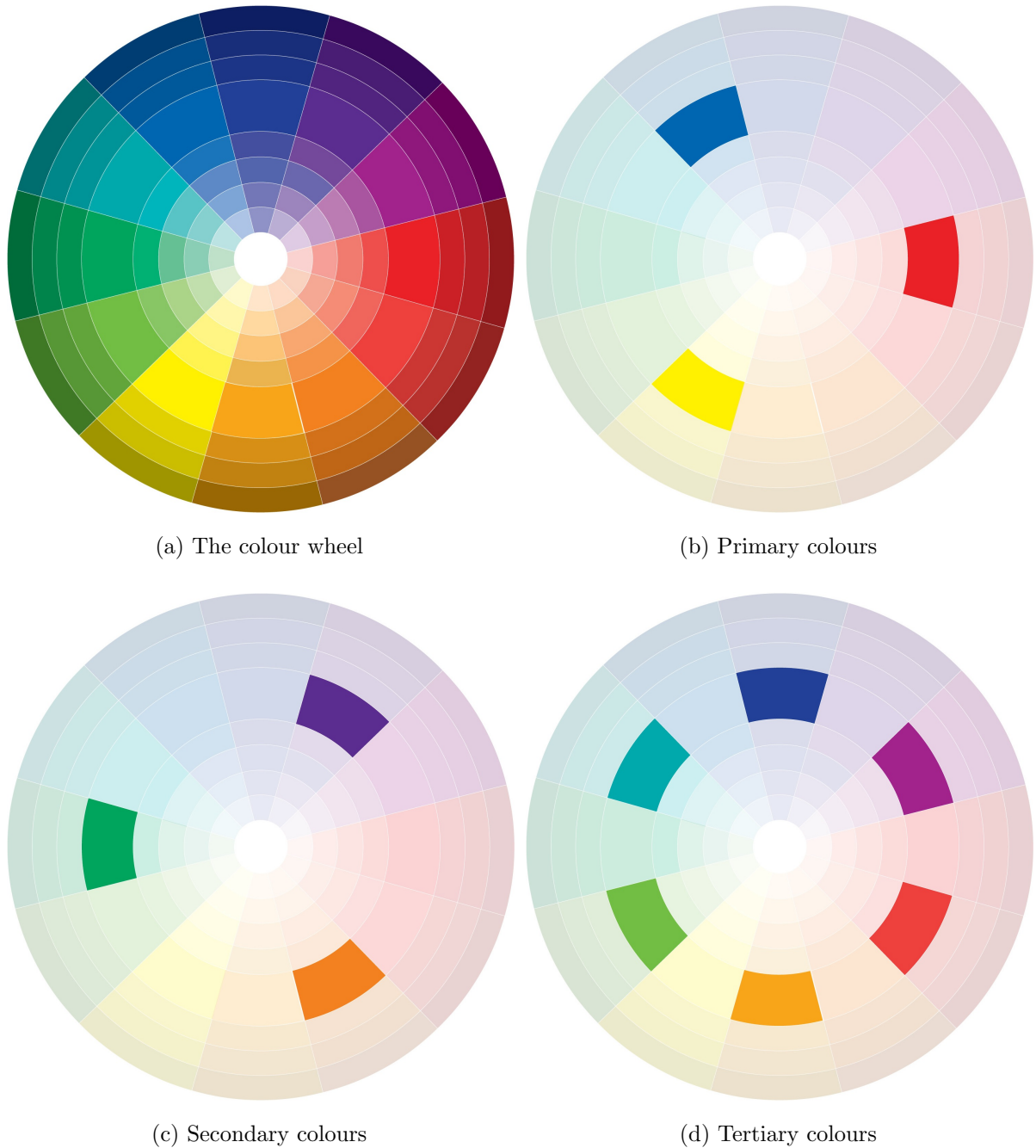
Figure 4.2: A visualisation of colour theory.
illustrated by Chad Engle

4.2   Art and design, aesthetics and applied aesthetics

In aesthetics there are two different fields, art and design. A way to separate them is aesthetics and applied aesthetics. We need to say a few things about the differences between these two branches.

For Immanuel Kant beauty is something that should not have any function. For him there are two kinds of beauty and they are distinct: "*There are two kinds of beauty: free beauty (pulchritudo vaga), or beauty which is merely dependent (pulchritudo adhaerens). The first presupposes no concept of what the object should be; the second does presuppose such a concept and, with it, an answering perfection of the object*" [40]. So according to Kant, we appreciate beauty for what it is, its semantic attachment to concept, when the object we are contemplating has no purpose. If it has, we automatically link it with its purpose and as such we stop judging it for what it looks like or for its semantic meaning but for how fit it is to accomplish its task. A painting for example does not have any practical use, so it is contemplated only for its aesthetic qualities. A car on the other hand has practical use, so when judging it we cannot help but linking it to its use as a car, to the way it handles on the road, to its spaces inside and so on.

For Kant beauty can not be proven, either: "*Proofs are of no avail whatever for determining the judgement of taste and in this connection matters stand just as they would were that judgement simply subjective.*"[40].

4.2.1   Applied aesthetics

Applied aesthetics is particularly relevant to the domain of this research. A simplified explanation of the difference between aesthetics and applied aesthetics is that Aesthetics is a notion mostly reserved for fine arts, when applied aesthetics is a notion that relates to things we use. The National Gallery of Art in Washington defines applied arts as "*art made for a practical purpose (e.g., weaving, metalwork, ceramics, woodworking, graphic*

design, etc.) [2].

Applied aesthetics regards art and life as two units of a greater sum. Applied Aesthetics in contrast with traditional aesthetics are not concerned about the study of beauty; instead they deal with "*aesthetic phenomena*" such as light, shadow, space, motion and sound, as well as how the observer responds to them[98]. In addition the medium, such as newspaper, television, computer screen, and so on, is not neutral.

4.2.2 Web design

Web design is one of the disciplines that belongs to the Applied Aesthetics field. It has a purpose, it is limited to the fact that is on the web but at the same time it involves a few different media: typography, photography, illustration, video, music, sound design, animation and so on. It employs a lot of people in the Internet industry and many universities offer web design courses. But still as a discipline, it is considered new. One design blogger, Armin Vit [93], raises a very interesting question: Are there any canonical web-designs? From the point of view of aesthetics there are no canonical designs in web design like there are in print. In print there are famous examples that we can use as landmarks, like Andy Warhol's prints (Figure 4.3).

Print media have different designs, *long sheets*, *tabloids*, *magazines* and so on. Each school of thought is easily identifiable on a news stand. Long sheets tend to have less photograph and more information; tabloids on the other hand tend to use bigger fonts, especially in the titles, pictures dominate the pages and the language is kept simple and sometimes even vulgar. Magazines are printed on different types of papers and so on.

But in web design there are no landmarks that are timeless, that we can look and aspire to. "No classics" would be the most accurate description. Though there seems to be no or very little research on that topic, we will advance some possible answers to the question, because it is relevant to the domain of this work. Important considerations may include the fact that the medium itself is new. The underlying technology of web design is still changing quite rapidly. Print has been a widespread medium for hundreds

(a) Marilyn Monroe, by Andy Warhol.
Photo by Rael Garcia Arnes.

(b) Art Work done by Andy Warhol for a music record by the Velvet Underground & Nico in 1967.
Photo by Jeremy Chan

(c) 32 variations of Cambells Soup cans by Andy Warhol, 1962.

Figure 4.3: Three famous examples of pop art, from posters, to ordinary cans of soup. All three prints are famous. Warhol was a very influential artist and personality in western art.

of years. Furthermore, in the case of websites, new designs replace old ones much more rapidly than buildings or books and much more completely than artistic artefacts such as paintings. Figure 4.4 (page 23) shows the differences in the BBC's news site of the BBC between 1999 and 2010.

(a) Screen grab, using wayback machine (http://www.archive.org/), of the BBC news site as it was on October 13, 1999.

(b) Screen grab of the news site of the BBC on March 8, 2010.

Figure 4.4: Screen grabs of the BBC website in 1999 and in 2010.

Some like Joshua Porter in will argue *Do canonical web designs exist?* [73] that Google is a prime candidate for this. But is Google a canonical web design? Can Google with its extremely simple search page, single text area and two buttons sitting under the logo of the company, be described as a design decision? Or merely as a better understanding of what users want out of a search engine? Is the success of Google as a site and a search engine based on the design of their main web page only or is it also based on the superiority of Google as a search engine? These are rhetorical questions. The most probable answer, is that all of those factors have to do with Googles' success. The fact is that any web designer would be hard pressed to pick a web site that exists or existed as a canonical design, as a memorable classic of web design.

When a user enters a website, it is the presentation of the website that they see first. The presentation is the one of the elements that will give the user the information needed to understand what the website is about and what it offers. Wroblewski says: *"Because images tend to have large file sizes, few images on a web page might be considered good in terms of usability. However, this guideline can be wrongly interpreted to mean that less visual elements provide better web experience. Such thinking could very well produce web pages that download quickly, but only confuse users through poor layout and are as interesting as staring at the 'Smith page of a phone book'"* [95]. Wroblewski further identifies three factors that a web designer needs to take into account when designing a web page: presentation, organisation and interaction (see Figure 4.5 page 24). Those three factors are overlapping and are closely related. Organisation is seen by the user through the presentation and interaction; it is regulated through the organisation and understood through presentation.

According to Hoffmann and Krauss [47] there is very little literature out there that has to do with visual aesthetics for the web compared to website usability. This is even when visual performance is thought to have an impact on the users decision making concerning quality and trust of products or services [42, 47, 98].
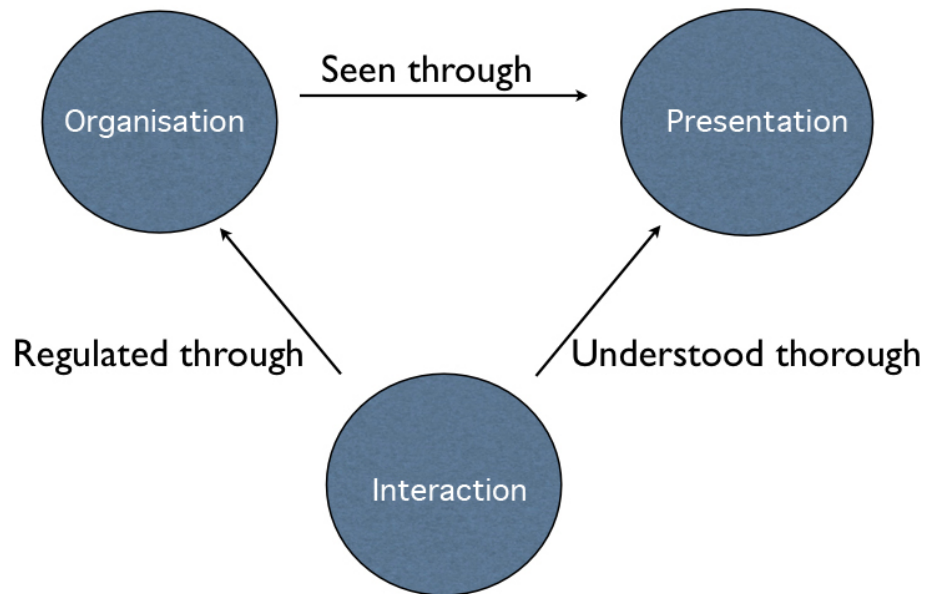
Figure 4.5: Organisation, presentation, interaction, after Wroblewski.

4.2.3 Semantics, design, message and appropriateness.

The aspects of semantics that involve design and are of interest to this work for two reasons:

- Semantics are an integral part of the work of a web designer when producing a page.

- Design has to do with communication, explicit and implicit.

When testing different web designs, as well as designing an experiment as we are doing, we need the design to be in harmony with the subject.

The medium itself carries a message and influences the content of the message. Some go as far as to say that *"the medium is the message"*[54]. Even as an overstatement, what it really means is that the medium and the visual aspects of it, carry a big part of the message [25]. McLuhan does not really believe that the content is of no importance, but by overstating the medium he emphasises its importance over the content. The important information here is that the two meanings have to be *appropriate* for one

another. Message and medium include messages and if those two clash, it could lead to very adverse results. But we will come back to this later in the document.

Similarly when giving advice on how to design for the web, Wroblewski says that we need to learn how to "speak web". Which means a lot of things in this instance. Not only does it address the tone of the text used online but also the way of presenting the information as well as the structure of the website. For him the message is not only the text of the website, it is the website as a whole, "*images, text, colours, sounds, motion and so on*" [95].

It is the semantics of our designs that we need to be careful of, in order to make sure that the result is *appropriate* to the subject. By appropriate we mean that the message is uniform and that the user does not read into it something that we did not intend to communicate.

According to Zettl [98], we understand things depending on their changing associations and their context; in our perceptual process, we judge aesthetic phenomena and processes in a contextual frame of reference. The three major points are: stabilising the environment, selective seeing and the power of context. For Zettl, the medium itself creates the structure and the context, which echoes McLuhans argument about the medium.

So in order to design something for the right context we need to understand our audience. Sometimes we think that our audience is everybody. While that could be true in theory, it rarely is. For instance, a website can be accessed by whoever has an Internet connection from any part of the world, but the audience would be finite, depending on the subject of the website . The person who browses to buy something off the Internet has the qualities of an online shopper and when he or she is trying to find how to fix their coffee machine, they have the qualities of a household appliances engineer [95]. Therefore, we can find ways to define our intended audience and its qualities. Another way is to start defining what audience we intent to target, what services we are offering and who is going to use them.

Figure 4.6: Kathy Dawkins, Interference (Plaque), authorised language cited by Crow.

As we previously stated, the medium carries a great part of the message. When the messages of the medium and the content clash then the results can be unexpected. Kathy Dawkins [39] is an artist who carried the written content of graffiti, to an official language (Figure 4.6 page 26). The result is shocking not so much due to the message, as very few of us are shocked by the message of the graffiti on the street. But when these messages take an official form, in this case a plaque of the Liverpool Council, the medium carries a message of seriousness and officialness while the content is a message of profanity.

Zettl states that "*elements of visual aesthetics should work in concert to support the message, else it is ineffective communication*" [98]. This further reinforces our argument about appropriateness.

4.3   Usability, ISO standards

Jordan states that a review of the human factors literature will show that the field does not traditionally give humans any terms of description other than age, gender, education or profession [39] . Satisfaction is typically perceived as the mere avoidance of physical or cognitive discomfort and, as proof of that, Jordan cites the definition of satisfaction as defined by the International Standards Organisation: *"the level of comfort that the user feels when using a product and how acceptable the product is as a vehicle for achieving their goals"*. He continues, the traditional human factors approach, is to treat the user as a *"mere physical and cognitive component of a system comprising the user, the product and the environment of use"* [39]. Beauty, aesthetics and pleasure are meanings that are completely absent from the traditional field of human factors.

4.3.1   ISO 9241

The International Organisation for Standardisation, issued a new ISO 9241-151 guidance document in May 2008. This documents sets as clearly as possible the standards that websites should follow. These should concern everyone but it is most common to find corporate and more importantly government and institutions websites are the ones that make sure they are following these rules. About aesthetics the document states this:

"Web user interfaces are presented on a personal computer system, mobile system or some other type of network-connected device. While the recommendations given in this part of ISO 9241 apply to a wide range of available front-end technologies, the design of mobile web interfaces or smart devices could require additional guidance not within its scope; neither does it provide detailed guidance on technical implementation nor on issues of aesthetic or artistic design" [5].

It is not our intention to accuse the ISO committee of disregarding aesthetics. It is extremely difficult to set standards that are clear for subjects of the likes of Internet websites at the usability level. Setting specific standards for aesthetics would

be practically impossible. As already mentioned, there is no aesthetic theory that can be used as a clear guide for designs to be liked by everybody. That is why setting standards for aesthetics would be extremely difficult, if ever achievable.

4.3.2   Usability as a dissatisfier element

In this Section we will introduce a hypothesis made by Jordan: we expect a product to be easy to use, if it is not we get aggravated, if it is we do not notice that it is [39]. Instead of being pleasantly surprised when confronted by an easy-to-use product, users get dissatisfied by difficult-to-use products. According to Norman [34] as interviewed by Hassenzahl and Blythe, when users are performing tasks and everything goes well, they might even feel relief; relief is not a positive feeling and surely not the feeling that we want our users to feel. Relief means that while using our product users were feeling stressed about the things that could go wrong and this is not a pleasurable experience. Jordan believes that what he calls a pleasure-based approach to human factors will not replace usability-based approaches. However, it may well be that easy to use products, are not pleasurable to use, or are ugly. Despite this, the chances of success of a difficult-to-use product to succeed are slim, that does not mean that there have not been successful products that where difficult to use. Anyone who remembers the first TV remote controls, or the VHF recorders. But these where products that their alternative was to physically get up from the couch and go to the TV to change the channel for the remote, or miss your favourite show while being stuck at work. Which suggests that the usability is not to be disregarded and should be taken seriously along with aesthetics.

Jordan also states states that people have social interaction with products; they are not just tools but are more like living objects that we have relationships with. Objects have the power to create emotions in us, like happiness, satisfaction, anger, fury, pride and shame [39].

According to Jacob Nielsen, Internet users are driven mainly by their goals. In

other words, they seek some kind of information or service. Functionality, ease of use and uniformity of design are very important things for the user. For Nielsen, users do not appreciate any obstacles standing between them and the information they seek. Therefore, everything has to be as simple as possible. Nielsen has a very functional analysis of the web. Nonetheless, nowhere in what he says is there any mention of aesthetics or beauty [62]. Instead, He tends to make a prohibition rule whenever something does not work, without taking into account aesthetics or beauty [34].

4.3.3   Usability in the context of our experiment

We are going to address usability in the context of our experiment in more detail in Chapter 6. Nevertheless it is important to state here that for our experimental design we are aiming to create a usable website. Our hypothesis is not that aesthetics is a more important factor than usability. For our experiment we need to create a site that will be usable and not aggravating for the users in any particular way that was predetermined.

4.4   An ontological misunderstanding between the art and computer science communities

Arguably there are ontological differences between the artistic community and the computer science community. Our belief is that part of the problem lies in our understanding or indeed our differences in understanding of the terms used in common by the two communities of engineers and designers.

When Nielsen writes about design for different resolutions and screens, his use of the word "aesthetics" is more related to functionality and ergonomics: for example, "*are captions directly under the images*"[63].

An engineers understanding of the word design for example is completely different than an illustrators understanding of the same word. One would understand design as a technical drawing or a concept of a program while the other would understand

the presentation of it. Kristina Karvonen in her paper The Beauty of Simplicity[42], debates the understanding of the word simplicity. That means that simplicity is a word understood in different ways by the two communities in question.

A purely engineering school of thought to which Nielsen belongs, regards simplicity as a design stripped of all flashy extras.The artistic school, sees simple design as a clean design that intends to be beautiful. Karvonen is of the opinion that a multidisciplinary approach to the subject of web aesthetics is needed in order to better understand it.

4.5 Our position

In this work we take a position closer to relativism but with some modifications. We argue that in aesthetics what matters is the pleasurable play of our cognition that makes us see beauty. Because cognition is personal, we believe that beauty is found according to personal taste, which is developed by an individual's experiences in life, as well as by the culture in which the individual is brought up or later submerged in. Our bodies also probably play a great role in the sum of these things, as to whether we are short-sighted, far-sighted, colour-blind and so on. To these we must add the fact that humans have few differences as a species. We are genetically very close. So it would be reasonable to assume that there are some things that we consider beautiful in common, that are hardwired into the human brain or we commonly find them beautiful because of our common physiology. This could explain the transcendence of some art across cultures, and is not far from the primitive emotions theory. Also and this is just an example, Stone and Collins in 1965 pointed out that our field of vision is close to the golden rectangle in shape. They advanced that this could be an explanation of why there is a human preference to the golden rectangle[2]. We also argue that Norman's theory about emotional design, with its three emotional levels, seems very promising. It does not clash with our relativist position.

We reject the criticism that a relativist theory is imposing because it tries to stop

---

[2]as cited by Green

people from thinking of as right or wrong [97]. The criticism is a sophistry that holds that since in relativism all judgements are right, then the relativist theory has to be right and as such whatever a relativist declares is right, because there is no wrong. Only whoever accepts that there can be right and wrong aesthetic judgements can accept that they might be wrong. Nevertheless our relativist position is not about *life, the universe and everything*, as Douglas Adams already answered that, (it is 42^3. Our relativist position is about aesthetics and only aesthetics and us such is as open to criticism as any other position.

We conclude that in aesthetics, there will be trends; a significant percentage of people will find beautiful or just pleasing, some particular design and will find displeasing another one. It is this common ground that we hope to find in our experiment in order to explore the link between aesthetics and usability.

4.6 Aesthetics in the context of our experiment

For the design of our experiment we need to take into account a number of factors. In this part of the document we will address the issues that arise from the literature review and concern the experimental design.

- We need to consider guidelines of design, colour harmonies, proportions and others used by designers that seem to be pleasing to most people. Our intention is not to create some ground-breaking revolutionary web design; this is neither feasible nor within the scope of this document. The intention is to produce a design that would follow some basic rules in order to be pleasing to as many users as possible.

- We need to consider appropriateness. We need to make our design appropriate to the content of the website in order to keep the users from being startled or annoyed by a discontinuity of the message between design and the content of the

---

[3]This is a joke referring to Douglas Adams, *The hitch hikers Guide to the Galaxy* where a super computer called *deep thought* is asked the ultimate question; to calculate the Answer to the Ultimate Question of Life, the Universe and Everything. The answer given is 42, but no one knows 42 what [8].)

site.

- We need a design that would be able to address a culturally varied audience.

- We need to create a site that is pleasing rather than avant-garde, in order to avoid any strong reactions by the users.

These considerations are in line with our aesthetics postion. The first consideration about design guidelines is in accordance with humans sharing common features and thus trends in what they like. The second consideration about appropriateness is self explanatory and we cannot see a clash with our position. It can be argued that it is consistent with most aesthetics positions, even when an inappropriate design is used like in Figure 4.6, where in order to get the desired effect the artist used contradictory messages between content and presentation. Lastly, anticipating cultural differences is again in line with our position, of acknowledging cultural differences in aesthetic judgements. This issue is one of the most difficult to resolve and the solution that we proposed was to keep the design as less flamboyant and *"opinionated"* as possible.

4.7   Recapitulation

In the Section *design* we mentioned some rules that designers use in their creative process to understand more practical aspects of the creative process. *symmetry, asymmetry proportions and rhythm* were presented as aspects of design. *proportions* were presented more analytically by introducing some proportions widely used by designers, like the *golden mean* and the *rule of thirds*. Following this, *colour harmony* was introduced.

The field of *applied aesthetics* was presented and more specifically the aspects of *web design* that are of interest to this research were introduced. Aspects of *semantics* were introduced thereafter.

Usability and, more precisely, the *ISO 9241 standard* was introduced, as well as how lack of usability can become a *"dissatisfier"* element. The section concludes by presenting usability in the context of our experiment.

This Chapter was concluded with the presentation of our *position* on the various theories presented and our position in the context of our experimental design.

CHAPTER 5

Past approaches to the problem

In this Section we critically review research already carried out in the field of usability and aesthetics. We argue that an over-narrow focus on the engineering aspects of aesthetics has impacted the scope and success of these experiments. Our understanding of aesthetics is an important factor in determining what experimental methodology we are going to develop. However it is important to review experimental methodologies of past research in order to draw information and conclusions about the experimental design.

In this Section we present five papers, central to this research: one by Noam Tractinsky and four papers by Alistair Sutcliffe. While this Section does not claim to give an exhaustive review of all the literature of the domain, the most significant and relevant works have been chosen. Tranctinsky is one of the first researchers that experimented on the relationship between aesthetics and perceived usability. Suitcliffe used websites for his experiments which are particularly relevant to our research. This is a list of papers that we are going to review:

1. Tractinsky's ATM experiment

2. Assessing the Reliability of Heuristic Evaluation for Website Attractiveness and Usability

3. Interaction, Usability and Aesthetics: What Influences User's Preferences?

4. Investigating Attractiveness in Web User Interfaces

5. Assessing Interaction Styles in Web User Interfaces

These critiques have been published in Kominis [46]. This Chapter presents a fuller version.

5.1   Tractinsky's ATM experiment

Tractinsky recreated an earlier experiment by Kurosu and Kashimura [49][1]; he thought that cultural factors were at play and that it was unlikely to produce the same results in Israel as in Japan. The experiment consisted of testing the usability of several ATM (Automated Teller Machine) screens, [91]. The experimental design consisted of identical screens in which the positioning of the elements forming them were changed. By changing the layout, the aesthetic balances, proportions and symmetries were changed and thus the aesthetics of each screen was different. The Japanese result of the experiment was that people found it easier to use the better-looking ATM screens. The rerun of the experiment gave even stronger results in Israel than in Japan [67]. Tractinsky anticipated the reverse results, arguing that Israeli people would not be interested in aesthetics as much as Japanese people, hypothesising that the phenomenon was cultural. The initial hypothesis here would be that cultural changes influence users judgements and as such this position is not dissimilar to collectivist approaches.

We have reviewed Tranctinsky's experiment and its experimental design. The idea behind the experiment is sound at a first glance. The idea is that by changing the placement of the elements on the screen the aesthetics are affected. The proportions

---

[1]as cited by Norman [67]
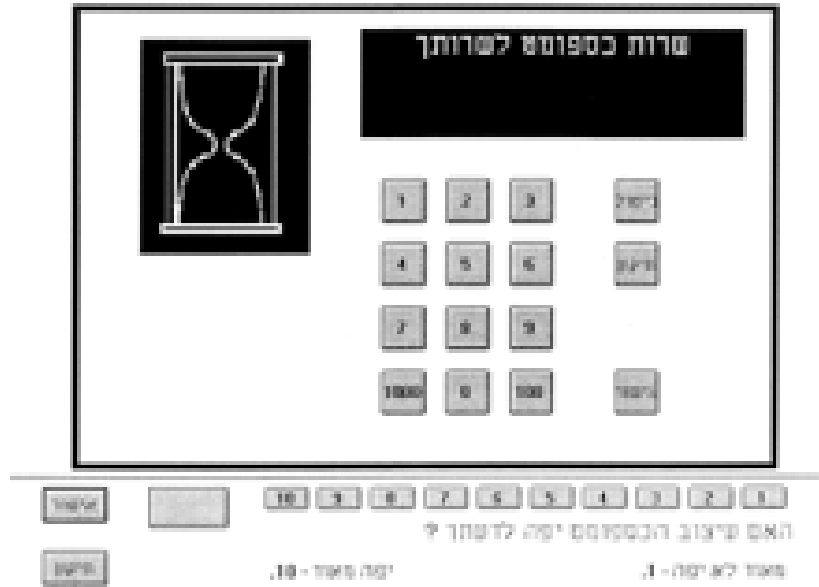
(a) ATM screen sample, high aesthetics score
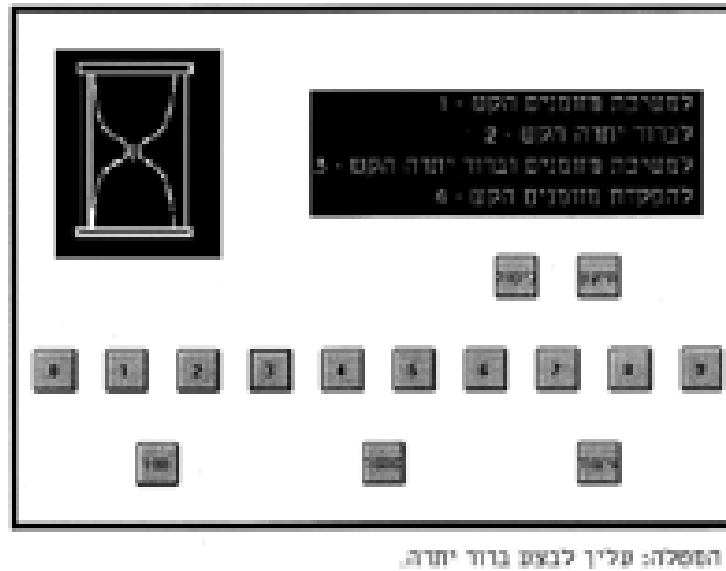
(b) ATM screen sample, low aesthetics score

Figure 5.1: Screens from Tractinsky's experiment

and ratios of the elements on the screen change, and thus the aesthetics are affected, by moving the elements, the synthesis, and the framing. As a result, the screens become either more or less attractive to the user. Examples from the screens used in the experiment in Figure 5.1 page 31. The experiment was based on ranking the outcomes by asking users to assess a sample set of ATM screens.

A possible flaw we see in this experiment is that changing the position of the same graphical elements on the screen is more a change in ergonomics than in aesthetics. The users, when ranking the screens, might be ranking what they see as easier to use rather than nicer to look at. The experiment was ranked both on an aesthetics scale and a usability scale; however, it seems possible that when users are asked to rate aesthetics, the answer carries elements of ergonomics that would affect usability.

In this paper Tractinsky is trying to find out if beauty actually changes the perception we have about usability. Separating usability from aesthetics is very difficult, as discussed in the previous Chapter, since organisation, presentation, and interaction are closely related, (Figure 4.5). Users might not always see the difference between looks and ergonomics, nevertheless it is essential to avoid ambiguity as to what the exact question is the users are answering.

Thus, the experimental design needs to address and minimise questionnaire ambiguity. From this experiment we deduce that we need to try and distinguish ergonomics from aesthetics.

5.2   Assessing the Reliability of Heuristic Evaluation for Website Attractiveness and Usability

In his paper, Assessing reliability of the Heuristic evaluation for Website Attractiveness [87] Sutcliffe uses the websites of three airline companies to assess heuristics describing attractiveness and a heuristics describing content. The following tables explain two

Title of heuristic	Explanation of Heuristic
Judicious use of colour	Colour use should be balanced and low saturation pastel colours should be used for backgrounds. Designs should not use more than 2-3 fully saturated intense colours.
Symmetry and style	Visual layout should be symmetrical, e.g. bilateral, radial organisation that can be folded over to show the symmetrical match. Use of curved shapes conveys an attractive visual style when contrasted with rectangles.
Structured and consistent layout	Use of grids to structure image components and portray a consistent order; grids need to be composed of rectangles which do not exceed a 5:3 height to width ratio.
Depth of field	Use of layers in an image stimulates interest and can be attractive by promoting a peaceful effect. Use of background image with low saturated colour provides depth for foreground components.
Choice of media to attract attention	Video, speech and audio all have an arousing effect and increase attention. Music can attract by setting the appropriate mood for a website.
Use of personality in media to attract and persuade	This principle applies primarily to e-commerce websites when use of human image and speech can help to attract users and persuade them to buy goods by being polite and praising their choices .
Design of unusual or challenging images that stimulate the user's imagination	And increase attraction; unusual images often disobey normal laws of form and perspective.

Table 5.1: Attractiveness heuristics as quoted by Suitcliffe

heuristics sets (Tables 5.1, 5.2)

In the abstract it states: "*The heuristics are tested by evaluating three airline websites to demonstrate how different attractiveness and traditional usability trade-offs contribute to the overall effectiveness*". We know that the heuristics used here are derived from studies done by Kristof and Kristof [48] and Mullet and Sano [56]. In the airline website examples, if you have ever flown with a cheap airline company you would know that a user might develop strong feelings about this mode of transportation. Such feelings amongst users might interfere with the results of the experiment.

This procedure, furthermore, relies on an analytical deconstruction while aesthetics seems to result from synthesis: the whole is greater than the sum of its parts. So

Title of heuristic	Explanation of heuristic
Consistent visual style	This heuristic is on the borderline between the two sets. Visual style is generic in the sense that a website needs to be consistent in terms of layout and image, but the style also needs to reflect the corporate values. Hence a website targeted at the youth market should use arousing material, whereas a site targeted at older users may use more restful, natural images. For tranquillity, choosing natural world content is advisable; conversely the image of a modern, dynamic organisation is reinforced by technological subject matter (e.g. racing cars, jet aircraft, spacecraft).
Visibility of identity and brand	The effectiveness of this heuristic depends on the strength of the brand image and corporate identity. The design principle just recommends making the identity visible in a consistent manner.
Matching arousal to users' mood and motivation	This heuristic focuses on the match between the user model and website content. Variations to be expected are between age and gender. Ultimately this is a complex topic dealt with in many books on marketing research.
Stimulating users' interest by secondary motivation	Attractiveness can be increased by adding functionality that is not geared to the sites primary purpose, but may attract for another motivation. Some examples are placing games and simulations on e-commerce sites for users' amusement.
Selecting content to suit users requirements	This should result from a sound requirements analysis, but poor content display may confound a thorough requirements analysis. Content related to users' requirements should be clearly stated, in unambiguous language, with clear cues on how to find it."

Table 5.2: Content heuristics as quoted by Suitcliffe in his heuristics evaluation paper
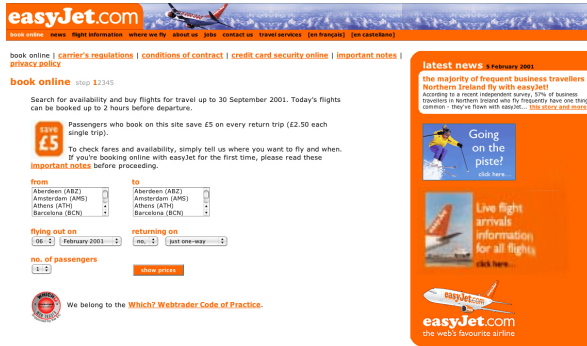
(a) EasyJet web front page, which has a somewhat irregular layout but to its credit is good at presenting its basic message: cheap flights and the economic incentive for booking on the web.

(b) Virgin website, illustrating the contrast in graphic design with Figure 5.2a. The jet windows and keyhole metaphors contain animations to attract attention.

(c) British Airways front page, with a low-key corporate identity (compared with Figure 5.2b), but the structure is well laid out and the content meets with users requirements for flight browsing/booking.

Figure 5.2: Screen grab of the websites that Sutcliffe used for his experiment, with the original captions from his heuristic evaluation paper.

deconstructing the aesthetics is unlikely to give us a good understanding of its value [41]. One more aspect to consider is that in lab conditions users do not have the same reactions that they would have at home or at work (real conditions), when performing their own tasks. The users' reactions in the privacy of their home would be very different than their reaction in the lab where they are observed.

In addition we have some reservations about the heuristics used in this experiment. For example we believe that using the symmetry and style heuristic as a firm rule to judge a design is flawed; in the design world, there have been advocates for symmetry, non-symmetry, proportions and/or rhythm. Some of these issues have been addressed earlier in this document, in Chapter 4, and, more specifically in Section 4.1. Just to reiterate our point, we do not hold that symmetrical web pages are necessarily non-aesthetic, this would be unfounded; but we hold that non-symmetrical web pages have an equal claim to aesthetic qualities. In some forms of art, like photography, courses actually recommend against the use of symmetry except under specific circumstances, because the end result is often a boring picture.

In the case of depth of field the use of a background image that interferes with text in the foreground can render the outcome not only unreadable but also looking poorly designed.

For *choice of media to attract attention*, we agree that the choice of media is very important; usually music is discouraged when used in web pages. Music could indeed set the mood of a website, but it could at the same time be a nuisance. A user that is surfing from their work place might depart in a haste if it starts playing audio. Another possible case scenario —is that users do more than one thing on their computer—, they listen to music while working, surfing the web and so on. So when a site starts playing music unexpectedly it can be an *irritation* as it will interfere with the other multimedia activities the users might be performing at the same time. In any case imposing audio on users when they did not explicitly ask for it, could have some very negative effects.

Use of personality in media to attract and persuade could be flawed, as the research

conducted by Murano concludes "*on the whole showed that the results were not one directional, either in each respective domain or across the set of nine experiments*"[57]. Murano's research shows that there is no conclusive evidence that anthropomorphic elements help users in their tasks consistently.

We argue that the deconstructive approach chosen in this research is flawed. We think that Zettl is right in this case when stating that: "*The elements of visual aesthetics are interdependent and should be studied accordingly (the total is greater than the sum of its parts).*" as cited by Krauss and Hoffmann [47].

5.3 Interaction, Usability and Aesthetics: What Influences Users' Preferences?

In a second experiment, Sutcliffe et al, evaluated two websites that have the same content but different interactive metaphors as stated in the abstract on what influences users [86]. According to the abstract: "*In this paper we describe an evaluation of two websites with the same content but different interface styles (traditional menu-based and interactive methaphors). A formative usability evaluation was carried out with heuristic assessment of aesthetics and questionnaire assessment of aesthetics, content, information quality, usability and post-test memory.*"

In Figure 5.3 the lower screen grab is an example of what they call *expressive aesthetics* and it was found to have serious usability defects in contrast with the first screen grab.

However, it can be argued that differences in stylistics and the difference in usability, render the results weak. The stylistic or presentation difference, makes it difficult to compare the aesthetics of two images that are not of the same kind. One of the designs will be aesthetically preferred; there are problems involved in measuring the aesthetic difference between different styles. What is more beautiful, a pre-Raphaelite painting or a cubist Picasso? A Rembrandt or a Kandinsky? Where the style is so different,

(a) Menu-based (upper) and metaphor-based (lower) interfaces for the Renaissance Connection website.

(b) The Innovations 1400-2020 Section in the menu-based (upper) and metaphor-based (lower) interface

Figure 5.3: Screen grabs with their original captions from the web pages used by Sutcliffe and De Angeli for the experiment

comparison may be very difficult. The usability problems that the expressive site has in contrast to the menu-based site could push people to say that they do not aesthetically like the site because they are irritated by poor interaction and thus confuse results even more.

A key outcome of this analysis is that in order to see how users react to aesthetic differences it is desirable to test sites that are stylistically and ergonomically the same, but are aesthetically different: one conforming to agreed forms of beauty within its chosen style, versus one that breaks these forms without losing its usability structure.

5.4 Investigating Attractiveness in Web User Interfaces

In a third experiment, Hartmann and Sutcliffe [89] use three websites that have the same content. The websites of departments of Stanford University are tested: Stanford Design division, Stanford HCI Group; and Hasso Plattner Institute of Design DSchool (Figure 5.4 page 38). The abstract states: "*A theoretical framework for assessing the attractiveness of websites based on Adaptive Decision Making theory is introduced. The framework was developed into a questionnaire and used to evaluate three websites which shared the same brand and topic but differed in aesthetic design [...] The implications of framing and halo effects on users' judgement of aesthetics are discussed.*"

The *framing effect* and *halo effect* mentioned here refer to psychological effects influencing cognition and decision making and come from the field of psychology:

Framing effects is the different reaction that people have to the same choice presented either positively or negatively. Users that have been presented with the same problem positively rather than negatively will change their decision. "*The effects of frames on preferences are compared to the effects of perspectives on perceptual appearance*" according to Tversky and Kahnemann[92].

Halo effect is generally defined as "*the influence of a global evaluation on evaluations*

of individual attributes of a person, but this definition is imprecise with respect to the strength and character of the influence. At one extreme, the halo effect might be due simply to an extrapolation from a general impression to unknown attributes. Global evaluations might colour presumptions about specific traits or influence interpretation of the meaning or affective value of ambiguous trait information" according to Nisbett and Wilson [64].

Figure 5.4: Screen grabs from the sites used in the attractiveness experiment .

One of the questions in the survey is described as follows: *"Subjects were asked to briefly revisit all three websites and then compare them in two scenarios. In both scenarios they were asked to imagine being a student who will be going to study at Stanford University and to rank the three departments in order of preference. In one scenario the subjects were asked to imagine being a Bachelor student who is going to Stanford University for a one-month summer internship, whereas in the other the subject was asked to imagine being a Masters student who is going to Stanford for a PhD research studentship. The order of the scenarios was randomised between subjects"*. Asking users to choose a school depending on what they thought of the schools' websites might be strongly influenced by the users' previous subject prejudices, especially when they have to choose between a design school and an HCI school.

Sutcliffe stated that when asked to express the basis for their decision in relation to the qualities of the website, 87% of the subjects showed the framing effect by indicating that the look and feel of the website was the most determining factor for the summer internship, whereas for the PhD research-studentship, the content of the website was

decisive. However it seems likely that there are influences from previous preferences that affect the subjects' responses. As in Sutcliffe's previous experiment, the sites are very stylistically different and thus there is a similar problem. Rating the aesthetics of an artefact is quite difficult without taking into account substantial stylistic differences.

For this second experiment, it was assumed that the different schools would score differently on an aesthetics scale. However, each school presents its courses differently. The designers may well have put some thought into the appropriateness of the presentation method for the particular school. So the context changes depending on the school and thus the experimental *constants* vary.

The paper notes that are considerable usability differences amongst the schools' websites; the design school ranks lower on the usability scale. Nevertheless, all of the three sites present different usability issues which make the results more difficult to interpret. "*The departmental websites differ significantly in the amount of content provided, as well as in their visual presentation. [...] The complete website of the DSchool consists of 26 pages; in contrast, the websites of the HCI Group and Design Division each contain over 100 pages*"[89]. To this we would like to reiterate our critique of the previous paper and argue that due to the schools' different curricula the designs are stylistically different in order to reflect the different purposes.

5.5   Assessing Interaction Styles in Web User Interfaces

The final experiment reviewed here [85] also involves an evaluation of two websites. One was mainly Flash-driven, using the control room of a space ship as a metaphor, the second a more classic design, laid out in columns with pictures and text (Figure 5.5a and 5.5b). As the abstract states: "*A formative usability evaluation was carried out with heuristic assessment of aesthetics followed by post-test memory. The subjects had more problems with the metaphor-based site, but rated it more favourably on the*

aesthetics heuristics. There was no difference in free memory recall between the sites. The implications for website design and evaluation are discussed." [85]. At the end of this experiment the researcher asks the users which site they would use to teach young children about the solar system.

In this experiment Sutcliffe and De Angeli make a clear distinction between the layout, presentation and graphics/font aspects of aesthetics and the interactive part of the sites. But design and aesthetics are not just the font colour or colour scheme, and so on they are an ensemble. The paper presents the informal assessment of the interfaces: "*To summarise the informal assessment, the two interfaces had minor differences in the layout, presentation and graphics/font aspects of aesthetics; both used the same fonts and colour scheme, although the metaphor site used more interesting graphical and shading on titles and layout frames for the information*" [85]. We have some trouble with this statement since we do not see as being minor differences in the layout and presentation of the two websites (Figure 5.5, page 39).

At the end of the experiment Sutcliffe et al. asked their subjects which of the two sites they preferred and would use again in the future; all except one voted for the Flash version of the site. However, when asked which of the two sites they would use to teach secondary school children, the subjects' choices were split evenly between the two sites.

When choosing for their own use the subjects said that the Flash-based site was more interesting and thus they would prefer it. When thinking about teaching use, the subjects' reasons for their choices were:

- simplicity and ease of use by the children, when chosing the menu-based site

- engagement and interaction, for those that chose the metaphor-based site

This underlines the importance of context when judging a website. The paper demonstrates that when visiting a webpage users like to be engaged and like interactiveness. In our opinion, the experiment does not give results that link aesthetics and usability, thus also appearing problematic. The styles of the two sites are different and so is their
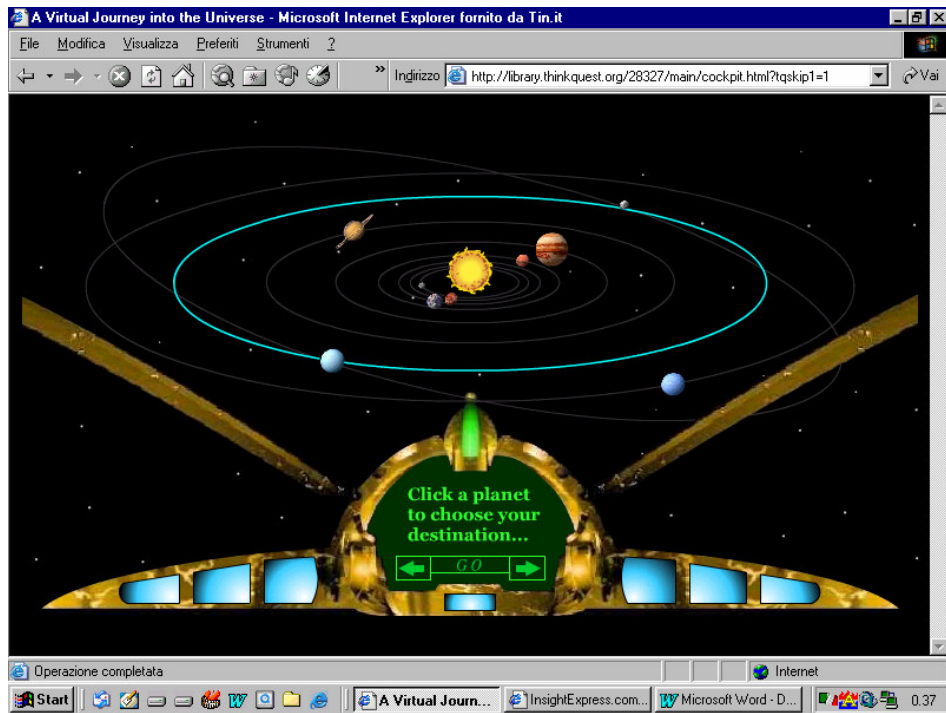
(a) Metaphor style interface showing the planets and cockpit metaphors.
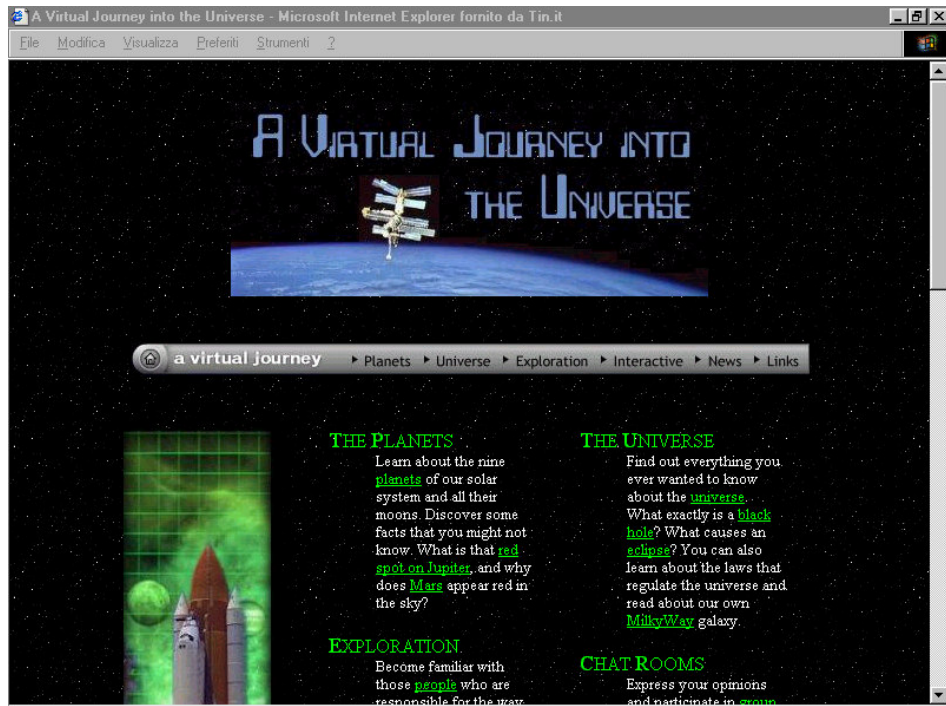
(b) Menu based style showing the same information with the metaphor style interface.

Figure 5.5: Screen grabs with their original captions from the web pages used by Sutcliffe and De Angeli.

functionality. In Section 4.2.3, starting on page 24, we addressed the issue of context and appropriateness. These two factors seem to count for a great part of the users' visual experience. A possible case scenario here is that a teacher would prefer using a more engaging site over a more usable one in order to keep the attention of pupils. Thus a possible confounding variable here could be that, if users think what they are using is appropriate, then they can be inclined in some cases to favour a site that is more interactive and engaging, and can be prepared to overlook usability problems. What we might be seeing here is different teaching schools rather than judgement directed towards aesthetics and usability.

We have already presented our view on aesthetics and the fact that we believe that it is a synthetic mechanism, thus not something that can be deconstructed into sections. Some theorise that design, being too complex and multileveled resists quantitative testing approaches [41]. Thus equating two designs on the single criterion on having the same fonts and colour schemes is problematic.

5.6   Conclusions

In this Chapter five papers have been introduced and discussed:

1. Tractinsky's ATM experiment [91]

2. Assessing the Reliability of Heuristic Evaluation for Website Attractiveness and Usability [87]

3. Interaction, Usability and Aesthetics: What Influences Users Preferences? [86]

4. Investigating Attractiveness in Web User Interfaces [89]

5. Assessing Interaction Styles in Web User Interfaces [85]

The list of papers on the subject we have reviewed here is by no means exhaustive. Only key papers of the domain that are particularly relevant to this research have been

reviewed. The particular relevance of these papers lies not only in that they investigate very similar areas of research but also in that they use webpages in their experimental design.

The most important criticism is that the experimental designs considered approach the subject of aesthetics with a deconstructive logic rather than a synthetic one. This appears to be a flawed approach to Design and Aesthetics while we argue that the result is greater than the sum of its parts.

From this Chapter we would like to reiterate the key points that derive from each paper reviewed:

Section 5.1 From Tractinsky's paper, the problem of ergonomics is of particular interest.

Section 5.2 Heuristics, being a deconstructive method, we believe is flawed as an approach.

Section 5.3 As stated above, the key conclusion of this analysis is that in order to see how users react to aesthetic differences, sites that are stylistically, usability-wise and ergonomically the same but aesthetically different need to be chosen and tested.

Section 5.4 It is difficult to compare sites that have a different purposes however they resemble each other. As stated above, in our critique of the previous paper, this is the more so as the different school curricula affect the designs, which are also stylistically different.

Section 5.5 Similarities between some elements of the two sites do not make them the same aesthetically. Context of design is a factor that needs consideration in our experimental design.

Despite our criticism we acknowledge the enormous importance of this body of work. The profound methodological implication of this work is used in this research to direct and develop our own experimental procedures.

Development of the experimental design and pilot testings

This chapter presents the first three iterations of our experimental approach. The resulting final experiment is then presented in Chapter 7.1.2.

The methodology used to develop the final experiment was an iteration between pilot testing of the experimental design and evaluation of the results.

From each redesign, piloting data was collected with the experimental design in mind. This indicated which aspects of the design needed to be corrected, changed and redesigned in order to meet the overall aim of this work.

Of course from those iterations the *literature review* (Chapters 3 and 4) got richer as more and more aspects were covered from issues that were raised from the pilot testings.

6.1   Development of the experiment 1

This experiment was designed to contain two steps; first, users were presented a set of pictures, of which the best and the worst were supposed to be used as backgrounds in a simulation of an ATM. The users then filled a usability questionnaire. In the case

of significantly different results, the null hypothesis that aesthetics do not affect the perception of aesthetics, could be rejected.

The rationale behind the initial experimental design was to repeat Tractinskys' experiment (Section 5.1 on page 30); however, instead of changing the layout of a set of elements, we decided to change the background. In order to choose a background it was decided to create four different ones and get users to vote on one picture as the most aesthetically pleasing and one as the less aesthetically pleasing. These pictures were used in the next step of this experiment, the usability testing. During the course of gathering data for the first step, *off the record* questions from the users regarding how to grade the pictures since they were not all pictures made us aware of this inconsistency in our method. There were two textures, one landscape photo and shaped background. For the second step the textures were chosen.

As mentioned in the previous chapter, Tranctinsky theorised that there were cultural effects at play and Japanese users would be more affected by aesthetics than Israeli ones. His results suggested that this is not the case. Our experimental design tests for differences between engineering students and art students in an effort to cover a different sort of cultural differences. However due to gender imbalances between the two Schools, the engineering school having more male students and the art school more female students, it was decided to use only female students in order to avoid gender imbalances in the results between the two schools.

6.1.1   Methodology

The next step was an in *between subjects / independent measurements* design rather than a *within subjects / repeated measurements*, Figure 6.1. Between subjects design has a potential for the most statistically powerful results [24].

For the second step of the experiment new users were chosen to participate. The System Usability Scale (SUS) is a questionnaire using a Likert scale. It is a *"a reliable, low-cost usability scale that can be used for global assessments of systems usability"*[17].

The questionnaire has been developed for an easy and reliable method, a *"quick and dirty"* way as as it is called by its creators, to evaluate the usability of a software.
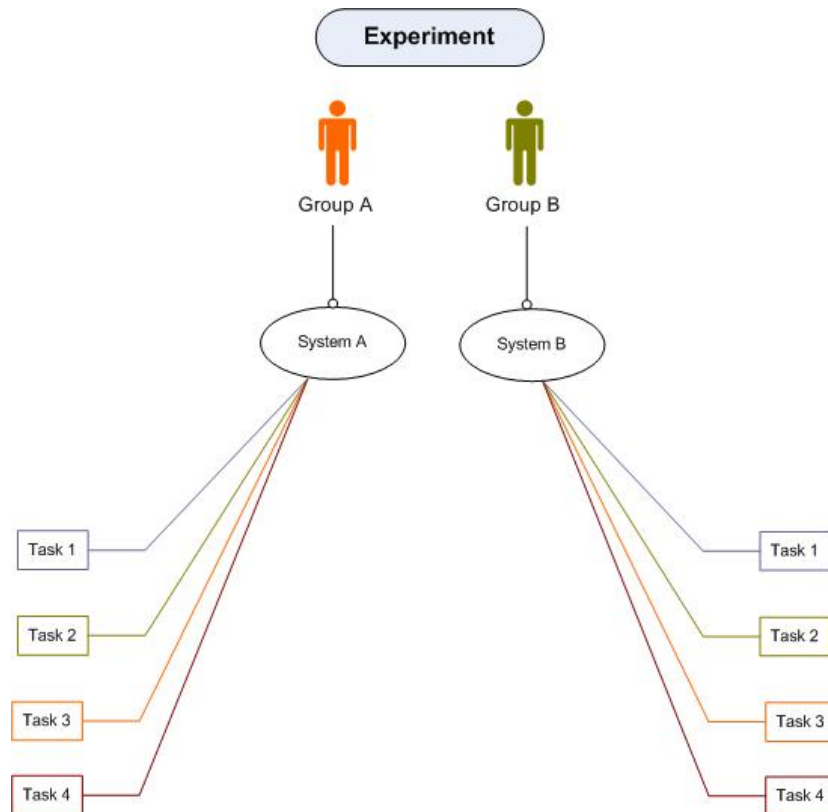
Figure 6.1: Experimental design for the first trial

6.1.2   Implementation

The experiment was carried out on an Apple PowerBook G4 (1.25MHz 1Gb or RAM). The systems tested by the users were designed in Macromedia Flash 8 and viewed in a stand-alone Flash player.

Users were presented with the software and given a printed set of instructions to follow. At the end of the instructions (Table 6.1) they were asked to fill in a questionnaire.
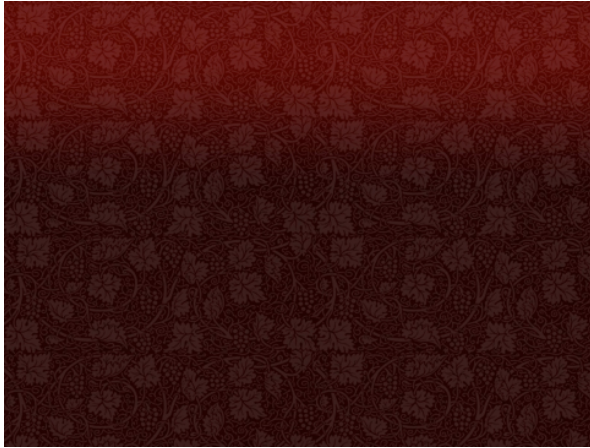
(a) Picture 1, *pattern*
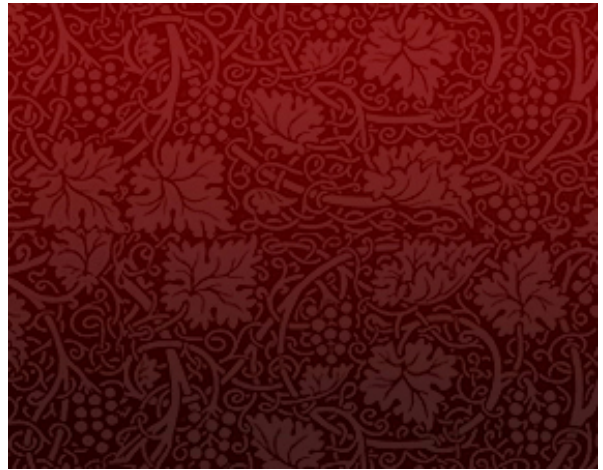
(b) Picture 1, detail of the pattern
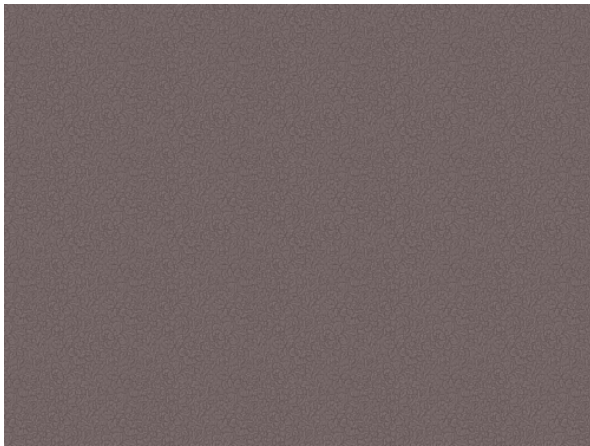
(c) Picture 2, *pattern*

(d) Picture 2, detail of the pattern

(e) Picture 3, *landscape*

(f) Picture 4, *shapes*

Figure 6.2: The backgrounds that were evaluated in the first part of the experiment.

- Please check your balance.

- Now withdraw 250 quid

- Check your balance again

- Try and withdraw another 250

- Ok now that you know that you do not have enough money deposit 200

- Your mobile is out of credit so top it up 10

- Check your balance again

- That's it you are done. Please fill in the form provided and dont forget your biscuit*

*Users were promised a biscuit, as a humorous form of payment for participating in the experiment

Table 6.1: Users were given these instructions to follow.

The questionnaire was the regular SUS questionnaire with one extra question about the perceived usability of the system.

A Flash version of the questionnaire was initially considered but rejected. The users might have felt that the form was an extension of the tested system. Furthermore some users that had no computer background could have felt uncomfortable with such a questionnaire. Thus, we proceeded with asking the users to fill in simple paper version of the questionnaire. The backgrounds and systems used in the second stage of the experiment are shown in Figure 6.3 (page 45).

6.1.3 Problems with the experimental design

In this Section the problems of the experimental design are discussed prior to presenting the results, as results are influenced by the ramifications of this discussion.
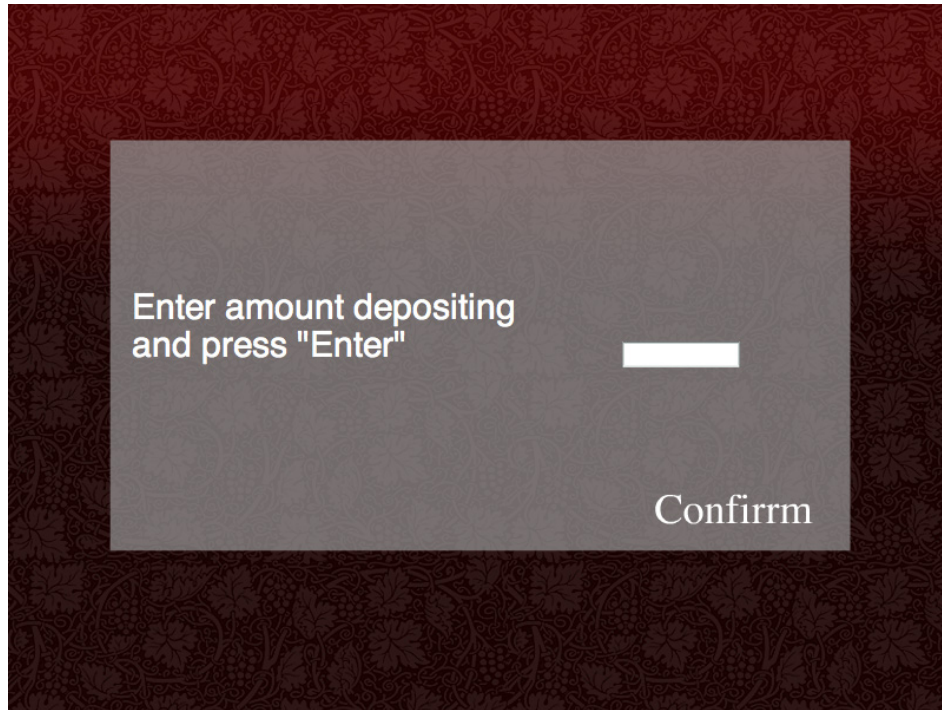
(a) First template (System A) used in the first experiment
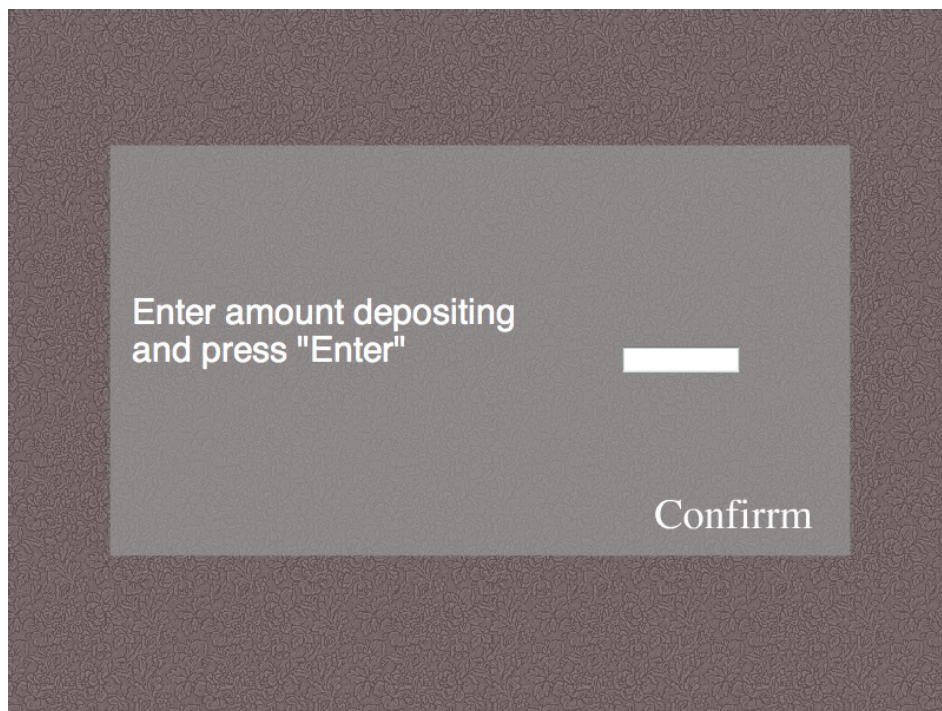
(b) Second template (System B) used in the first experiment

Figure 6.3: Screen grabs of the two templates used in the experiment.

6.1.3.1 Population tested

The number of subjects for this experiment was extremely low: 12 users in total from both universities. Limiting the experiment to female users was the reason for this. Thus there are no statistically significant results.

The most important outcome from observing the users while performing their tasks was that the steps performed the tasks too easily and that they did not seem to present them with any challenge. No absolute times were taken, but this was noted as a remark by the observer at the end of the experiment. The very high scores of the SUS support this.

6.1.3.2 Aesthetics classification, ranking *vs* absolute scoring

Users were asked to vote for best and worse image out of the four images presented to them. A problem with the classification of the images was perceived. When ranking images, users were expressing which image they preferred against the other images; this is not to say they liked them in any absolute sense. Thus their choice had a higher aesthetic value was higher than the ones it was compared too, in their opinion, but its "actual" aesthetic value remained unknown. The problem arising from such procedure is that we do not know if the users just ranked a set of images they thought were all of low aesthetic value, and thus they just chose one over another. Thus a plausible scenario is that they scored a low aesthetics template and a lower aesthetics template.

Our initial assumption that we needed a number of random images to rank as a first step of the experiment was thus flawed.

A common question when they were asked to rank the images was to ask where the picture is going to be used. The difference in style and *nature* as well as lack of purpose of the pictures was confusing the users. This question related to what was discussed in Section 4.2.3, context and appropriateness.

While reviewing the results of the ranking, as mentioned earlier the opinions were divided; three out of four images got votes as best and as worst image. It was thus

decided that two images of the same genre were best suited for the next step. The positive outcome of the experiment was the realisation that pictures of the same genre were needed: patterns for example, or landscapes; pictures of different genres undermines the validity of the comparison. This refers us back to the issues of style discussed in 5.3 on page 36.

6.1.3.3 Questionnaire

The SUS is a simple quick and, more importantly, a reliable questionnaire [17] about the usability of a system. But, it has its limitations: it delivers a single number. It does not offer a multilevel analysis. *"SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own"*, according to Brooke [17].

As mentioned earlier in this report, to define what is beautiful and what is not is one of the hardest aspects of this work. In this experiment aesthetics were not measured, pictures were ranked and then seen by our second-step users on their own, without any possibility of comparison.

6.1.3.4 Complexity of task and length of experiment

The complexity of the task set and the length of usage of the tested application is a variable in testing for a link between aesthetics and usability. We do not know what the results would be if we were to test a specific programme, for example a word processing programme, or an email client or even a corporate intranet website —a system with a certain level of complexity— questioning people over a long period of time. Such an experiment is rather difficult to conduct with limited resources. As mentioned above users seemed little challenged by the tasks set in the experiment. It took them very little time to work through them. It did not appear to them a convincing simulation of an ATM machine. Thus we derive two points from all this:

- The testing platform needs to feel real, not a dummy website.

- The experimental protocol should make users browse the site for a period of time that would offer the opportunity to get a real use of the test website.

6.2   Experimental design points to take into account

This experiment, notwithstanding that it yielded no results due to experimental design weaknesses, did present insight to issues that need to be taken into account in this research.

One of the most illuminating answers from a Heriot Watt student was written as a comment underneath the SUS questionnaire. It said that the system was too beautiful to be an ATM; it thus posed questions about the *appropriateness* of the design.

A recapitulative list is given:

- Aesthetics ranking is not a solution, as users are giving a opinion of comparison and not a general opinion.

- Appropriateness of design: The design of the site, its presentation, needs to be appropriate to the given purpose of the site.

- The testing platform needs to feel real.

- The testing site should be complex enough to engage the users. A too-simple site presents no engagement or challenge to the users and thus leaves very little margin to score it.

6.3   Development of experiment 2

For the second experiment, our first idea was to degrade the aesthetics of an already existing website instead of creating a new one. This would solve the problem of ranking and would require half the number of users in order to get statistically significant data.

Half because there would be no need for a first group of users to rate the aesthetics of the sites. A possible approach was to try and get permission from a website that is already regarded as aesthetically pleasing and in an experiment try and degrade the aesthetics in such a way that would leave all other factors intact. The problem with this approach is getting permission for a such a project from the site owners. The assumption here was to find something that would be generally considered well designed and use it for the "*high aesthetics*" element of the experiment.

6.3.1   The design

An alternative to this solution was to use Apple iWeb software to build a web page based on one of Apples templates and then tweak it to disfigure it. The rationale was that Apple is known to be very design-conscious. These templates that were made by Apple and by extension by the design team of Apple, should produce a good-looking website. Consequently a series of alterations on the original theme were created, (Figure 6.4 page 48).
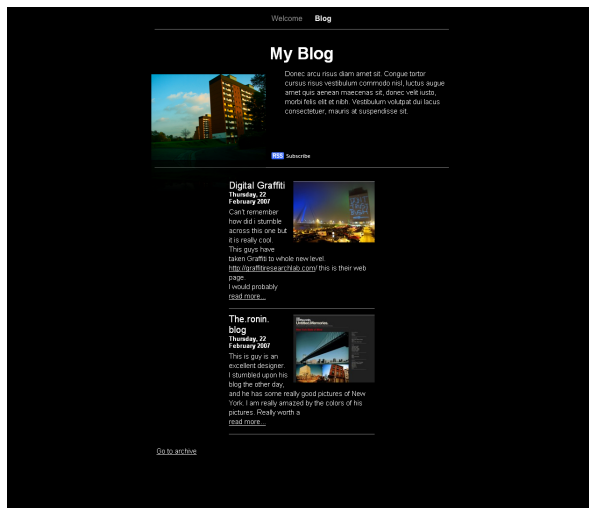
(a) Web page created with Apple iWeb. No variation from the original design
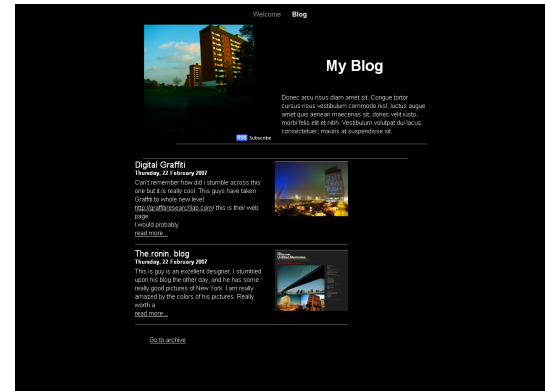
(b) Web page created with Apple iWeb. First variation on the original design
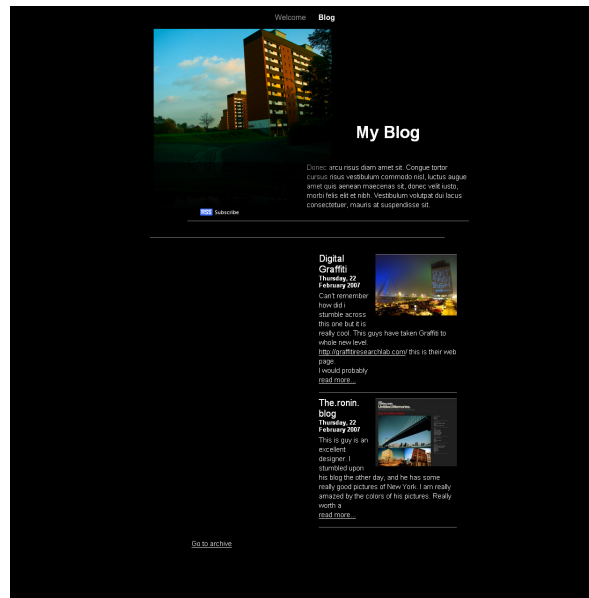
(c) Web page created with Apple iWeb. Second variation on the original design

Figure 6.4: Screen Grabs from the second pilot experiment. Variations on Apple's iWeb designs.

6.3.2 Piloting

A pilot experiment was conducted with a very small base of users in order to find possible pitfalls with this approach. The initial target for piloting the experiment was to find 10 users. However, the experiment was abandoned after the 5th user because of the

results obtained. The experiment consisted of users viewing picture grabs of different versions of the site created and ranking them according to which they thought more aesthetically pleasing. It became clear very quickly that the users did not really see any significant aesthetic differences between the versions they were ranking. Comments and questions mainly concerned usability issues, such as "Why are the fonts so small?" "Why aren't things aligned to conserve space?" Even when asked to judge the aesthetics of a page, the users were evaluating it depending on function. Thus a good experiment should keep the function the same and change only the aesthetics, in order to see if users will find it easier to navigate.

6.3.3 Results

The pilot indicated that differences cannot be subtle for this type of experiment. *Low* and *high* aesthetics need to be distinct. The aesthetic differences need to be obvious.

6.4 Development of the experiment 3

A webpage with an informative purpose explicitly designed to be aesthetically pleasing would be an appropriate choice. The web-page should have good design as a central feature. The OFFF conference in Barcelona in 2007, was such a page [4]. This event is a more of a festival with workshops and so on, rather than a conference as such. It is for multimedia designers, animators and generally people in the digital art and design world.

6.4.1 Experimental design

The approach taken was to strip the page of all the functionally unnecessary graphics. We created two different versions of the original; one a bit more bare than the other. Figure 6.5 (on page 50) shows for screen grabs of the website used for the experiment.
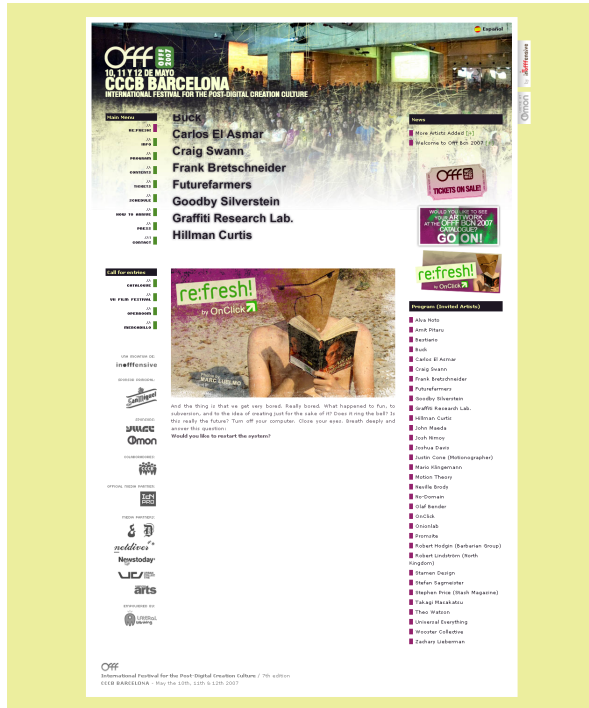
(a) Original webpage for the OFFF festival in Barcelona.

(b) The first level of taking out the embellishments of the web page.

(c) The stripped down version of the OFFF festival web page.

Figure 6.5: Screen grabs of versions of the website used.

The first version lacks the surrounding yellow colour, and the top picture is cropped and is not set up as a background graphic. A Flash component in the middle of the page has been removed. The purpose of the Flash component is to present the names of the artists in a fluid way that permits the user to scroll up and down, just by moving the mouse inside the boundaries of the Flash component. This has been replaced by a simple list in an *<iframe>*. The outcome worked in a similar way; however, there was no automation of the perpendicular movement of the text caused by the positioning of the pointer, it was replaced by a simple scroll bar. In the third version we completely deleted the top graphic of the page and left the logo. The reasoning behind this was that the logo was needed for branding purposes but the top picture, a photograph of the previous year's event where the colours were accentuated, was there only for illustrative reasons. The second picture in the web page has been left untouched, as it was part of the information provided on the page.

Initially users ranked the aesthetics of screen grabs of the three versions of the website. The hypothesis was that the users would rank as aesthetically superior the original page first, the one that was deprived of some graphics second and the one that had no graphics last.

A pilot ranking was carried out to see if this hypothesis was plausible.

6.4.2   Pilot testing outcome

After reviewing the results of the pilot ranking of the pictures the experiment was rejected. The results were not uniform, and no clear trend was seen. For some participants the original version was the most aesthetically pleasing and for others it was the least aesthetically pleasing. Such big contrasts would probably distort results and so this approach was also abandoned.

Here we can see the impact of fashion. The colour scheme features, as pointed out by one user, were unusual for a website, thus it created what some perceived as an *avant garde* effect and others as low aesthetics.

The experiments

7.1 The first experiment

7.1.1 Implementations of the experimental design for the first experiment

This Chapter will describe how the first experiment was designed to take into account the lessons learned from the previous experiments reported above.

7.1.1.1 Design of the first experiment

A number of lessons have been drawn above from the aesthetics theory, past experiments and pilot testings. This suggests a number of factors essential to a good experimental design:

- Ergonomics

- Aesthetic appearance of the website

- Rations

- Colour

- Fashion

- Style

- Content

- Experimental methodology

- Number of users

7.1.1.2 Ergonomics

As previously mentioned in Section 5.1, *ergonomics* is defined, in our case, as the placement of the buttons and the overall layout of the web page.

As stated in Chapter 5, to minimise the effect of different ergonomics impacting usability, we decided to keep page elements at the same place in different versions. A website was created using *Joomla!*[1] CMS (Content Management System) that was duplicated and "dressed" with two different templates. A CMS defines content and organisation independently of the context and the presentation. This offers an opportunity to either change the template on the same site, thus conserving the exact same context and organisation, or duplicate the website and dress it with different templates.

Joomla! was chosen as a popular open source system. In November 2010, *Google.co.uk* reports 23.3 million hits for *"powered by Joomla!"*. *Joomla!* thus has a vast community of users. A significant number of users for a system means that there is good documentation and a community that helps with possible problems.

In the second phase of the experiment, with an open population, Google Analytics (GA) was used to gather data; this is a statistics service currently used by 56.7% of

---

[1]Joomla! is a open source CMS server; for more information, see http://www.joomla.org/

the top 10,000 websites [18]. GA provides detailed statistics about a website's traffic. Installing this extra data gathering tool, allows more data to be gathered about the testing population, the amount of time they spend on the site, how deep into the site they went and whether they returned to see the site later.

The two templates are mainly a template and a copy of it that has been stripped of all embellishments. All pictures, boxes, dressing of the HTML are stripped and the only parts left are the instructions that concern placement on the webpage and organisation. Figure 7.1 shows the two templates designed and used in the experiment.
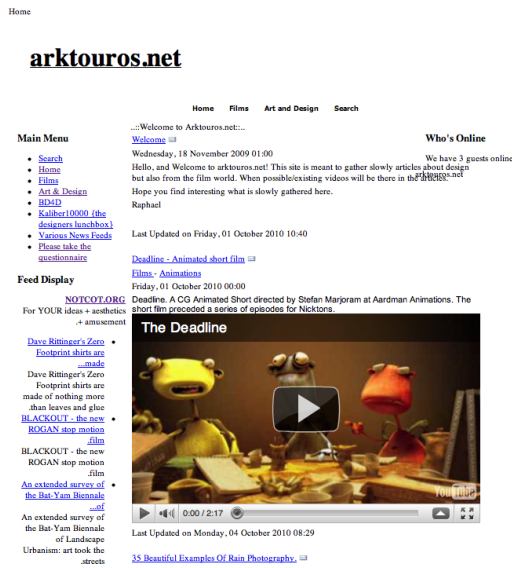
(a) Low aesthetics template.
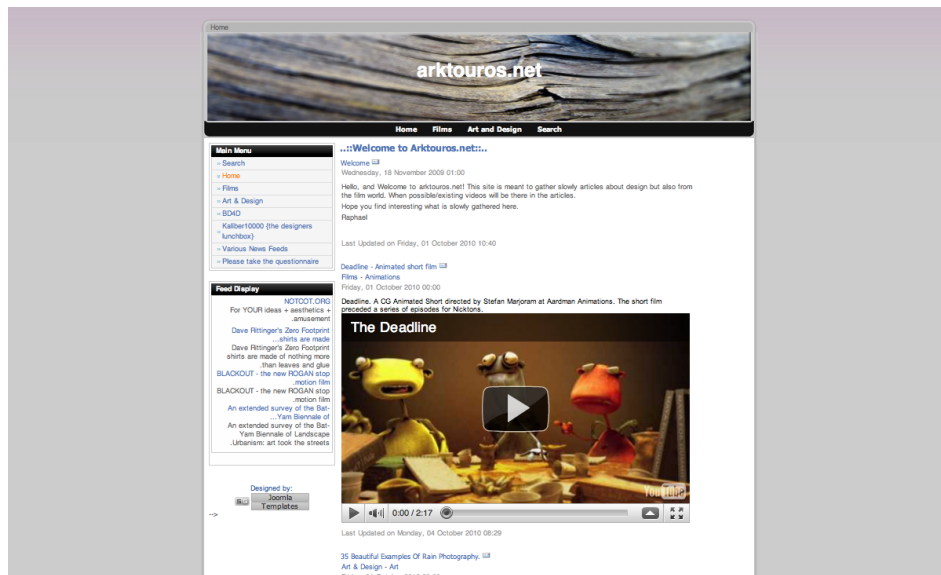
(b) High aesthetics template.

Figure 7.1: Screen grabs of the two templates used in the experiment.

7.1.1.3 Aesthetic appearance of the website

The high aesthetics template was designed with simplicity in mind. The intention was not to create a radical new interface, or a big breakthrough in web design. We were interested in creating a template that would not surprise the user and would be similar to the templates used by a majority of blogs.

Joomal! has a selection of free templates that can be used and tweaked. In the process of creating a design that would be suitable for the experiment we came upon different templates that were changed, tweaked to make them suitable for our purposes.
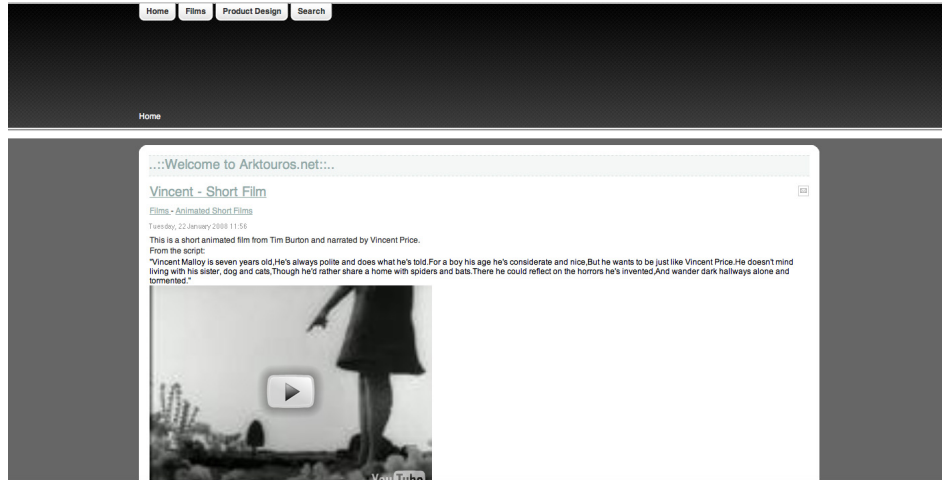
Figure 7.2: First iteration while designing the template

One of the first design iterations that was considered is the one, seen in figure 7.2. Our design decisions are based upon a few different factors.

- Use of ratios

- Use of colours

- Fashion

We will address these factors one by one consequently.

7.1.1.4   Ratios

Ratios, More specifically the *golden ratio* or *golden mean* were mentioned earlier in this document in the Literature Review Section (in Section 3.1.1 and Section 4.1.2). The Fibonacci scale was also mentioned as a very good approximation to the golden mean (0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, and so on).

Our template uses widths of columns from the Fibonacci scale. The widths for the left, right and main column are 233, 144 and 610 pixels respectively.

On the aesthetics low template the ratios were tweaked so as not to follow any aesthetic rule.

7.1.1.5   Colour

The *high aesthetics* template was designed to have a complementary colours scheme. This was chosen instead of an analogous, for example[2], due to the fact that analogous ones have almost no contrast and the colours are made very "tame". The colours that were chosen are not too bright in order not to create major reactions to the users that would be testing the site. The only exception is for the links that are chosen. (Colour harmony was mentioned in Section 4.1.3).

Colours that are in the *banner* top image (see figure, 7.1b on page54) showing a close up of piece of wood are complementary to the background gradient colour from grey to purple.

The *low aesthetics* template on the other hand was stripped down of all colours and left with the default settings for HTML links, which is to be underlined in blue.

7.1.1.6   Fashion

A conservative design was chosen on purpose. In this experiment we were not trying to create a radical new design that would be exceptionally bold. We are trying to test two templates for the same site, their main difference being that one is of higher aesthetics value and the other of lower. A new design that would be daring and innovative, breaking with the main stream of blog design, a more "*avant-garde*" design, could have some adverse effects. This issue already occurred during the pilot testing of the second experiment as we have seen in Section 6.3.

In Section 3.1.4 we talked about fashion and, as Kant [30] cited by Gronow and Georg Simmel [81] have said it is pointless to try to break away from fashion. Fashion will always be influential. What we tried to do is minimise the effects of fashion as much as possible by making a template that was trying to be a "*passe partout*".

---

[2]See Figure 4.2 page 20 for the different colour shcemes.

7.1.1.7 Style

Style was dealt with in a similar way. Presentation of information on the screen is in the end a series of design decisions. We were trying to keep the *high aesthetics* template close to what blogs look like in general.

On the *low aesthetics* template we were trying to go around the issue of style encountered by previous researchers by stripping the ornaments of the *high aestetics* template. We understand that it is practically impossible to do major changes to the website and leave its style intact. Nevertheless by stripping the design of its ornaments the style change is kept to a minimum. By keeping the same layout the style of the webpage is kept similar for both versions, high and low aesthetics.

7.1.1.8 Content

The webpage needed to be substantial enough to seem like a real project, more than two or three pages. The amount of information in the webpage should be on an amount that would take a user at least 10 minutes to see it all. It needed to be substantial in order for users to be able to navigate around it for long enough to form an opinion about usability.

In order to create a big enough site data was required to populate it. Due to the challenge of finding users in sufficient numbers to participate in the experiment, the population of users should not depend on a particular field. Thus a subject that users from many different groups could relate to was required. Of course some people will relate to some subjects more than to others but this effect can be minimised by finding a subject general enough in order to be of interest to a large number of users.

The most readily available datasets are weather data collected by various weather stations. This data was in the form of *xml* data sheets populated by a series of weather measurements, temperature, barometric pressure and such. However this kind of dataset is not very relevant or accessible for users that are not interested in that particular field.

The IMDB (the Internet Movie DataBase) would have been a good choice. However,

while it was available for download, we were not given permission to use it. IMDB has strict rules on the Digital Rights Management (DRM) of their material that came at a cost that we were not able to cover.

Eventually content about art and design was chosen in order to make it interesting to a wider public. The works and designs of different artists and companies were gathered in the blog. Such a subject might not be interesting to everybody and some groups of people, like artists, designers and architects and so on might be more interested in it than computer scientists would be. However this met the criteria of widespread interest reasonably well.

The site was populated with 22 articles in 2 categories, Films and Art and Design. Films has two subcategories, Short Films and Animations and Art & Design had three subcategories, Misc Designs, Houseware Design and Art.

7.1.1.9 Experimental methodology

We decided to choose an *independent measures* design for our experiment [24]. We chose it over a repeated measures methodology in order to have more significant statistics. Repeated measures have the disadvantage that users are influenced by the first *measure* they have experienced. Thus, their answers on the next measurement will be different, influenced by what they have already experienced. There are statistical manipulations that can be applied to the results to counteract the effects of repeated measures, nevertheless the results are much more significant if the design is based on independent measures.

A ready-made questionnaire was chosen for reasons that will be discussed in the next chapter. A questionnaire that was developed to measure Internet website usability specifically was desirable, not one developed with a wide variety of applications in mind, such as the SUS, which is more adapted to office applications. An individually designed questionnaire would not have met the strict standards and efficiency of other questionnaires that have been widely tested and calibrated.

7.1.1.10 Number of users

Finding users can be challenging. Three groups of users were used. The first two groups were a controlled population from Napier University and Heriot Watt University.

The second batch of users was is an open population. For this part of the experiment users were contacted in various ways:

- Facebook

- Email

- Word of mouth

- Other universities

For the open population groups, both the *low aesthetics* group and *high aesthetics* group, we changed the first questionnaire. It was replaced with a simple usability questionnaire [75] but one that also has questions about the age and sex of the users. This is due to the fact that WAMMI (Website Analysis and MeasureMent Inventory) [45] in the form that we are allowed to use does not include those two extra questions. The first questionnaire is there to double-check WAMMI. Questionnaires are discussed in more detail in the next Chapter.

7.1.2 Data analysis

In this section we will describe the way the experiment was conducted, the questionnaires used to gather data, the statistical treatments done to the data gathered and the results of the experiment. The objective of this chapter is to report procedures and results in detail. There will be a synopsis at the end of each section to state simply the findings and main points of the section.

A discussion about the results in relation to the hypothesis and to previously done experiments will be presented in Chapter 8.

7.1.2.1   User collection

The experiment was conducted using two different populations. The first was a "controlled population", students of two different universities. The second was conducted with an "open population"; in this case the users were invited by various means and methods to do the experiment: emails, social networks, word of mouth.

The experiment was conducted using an online questionnaire, WAMMI. Four groups were used to run the experiment. The subjects were given a procedure to follow, which of the procedure was to make the users visit the site and get acquainted with most of its features and content. What we wanted was for the users to get a real feel of how the site works and what it contains.

7.1.2.2   Procedure

Users were given a false brief before starting the procedure; if users requested for more information before the procedure, they were told that they are rating a blog before it goes live. The truthful version, along with explanations of what it was they did rate, was explained at the end if so requested by the users. Using a false brief to the users raises ethical questions. We believe that in this instance misleading the users was harmless. This is because believing that they were rating a blog that is about to go live does not affect them in any way, physically or psychologically. The site is not and will not be a commercial site in the future and thus there is no question of profit, either. If requested, the users were informed in full of what they really did rate. With these facts we assessed that the ruse was perfectly harmless to the users.

We misinformed them because the experiment is based on an independents sample procedure. They were led to believe that they were rating a blog that was about to go live because we wanted to assess users' reactions to a website that was as close to a real one as possible.

The following is the procedure that was handed to the users at the beginning of the experiment:

Procedure

Welcome to www.raf.arktouros.net.

Please follow these instructions. If you can not do something or feel that you can not find something proceed to the next step.

- Please find the welcome note of the site.

- Please navigate to Films>Animations>Crow by Psyops and take a look at the video.

- Can you find out what is "my room in a box"?

- In the category Art of Art and Design is there a video that you can see?

- Find the animation made by Tim Burton.

- Find the drop clock screen saver

- Please take a wander around each category.

Now please go back to the home page and click the Please take the questionnaire link and please be as detailed as you can in your remarks about the web site.

For the second questionnaire you will need a password. The password is: ******

Thank you!

The procedure was easy to follow and the users were left to their own devices as to whether they wanted to see the videos that they were instructed to see, or not. This was done to gather some qualitative data. What we were interested in was to see if there was a difference in how much interest the users had in the site. Part of what we wanted to see was if they would explore more things than the procedure explicitly told them to see. Gathering *speed of task* data in this instance would not have been very informative. Being an online site, with video streaming embedded from YouTube and other services, downloading times can vary from one case to another. A more significant measure would have been to ask the users if they wandered around the site more than what the

procedure instructed them to. After the initial controlled population experiment was completed, we used Google Analytics to gather extra data to how much each template was visited, for how much time and depth did users got into the site.

At the end of the manipulations the users had to fill two online questionnaires. The first one was there to reinforce the belief that they are just scoring some blog that is about to go live and we were testing it to fix possible problems. Although after the controlled population experiment was completed we introduced a questionnaire designed by Utah University [75], that gathered some additional information. We introduced some extra questions mainly concerning the demographics of the user population. The second questionnaire is the WAMMI questionnaire. The users at this point had to use the code provided by the *procedure* given to them. The code is different for each group of users and is there to separate the low aesthetics from the high aesthetics user results.

As stated in the previous chapter, the experiment was conducted on two distinct populations, Group A and Group B. These two groups were randomly distributed into two sub groups that tested the "*aesthetics high*" and "*aesthetics low*" templates.

7.1.2.3 WAMMI questionnaire

In this section we are going to discuss in more detail the WAMMI questionnaire and how it works. We used a questionnaire developed by Dr. Jurek Kirakowski and Nigel Claridge, the Human Factors Research Group in Cork, Ireland. WAMMI mission statement explains that: "*WAMMI carefully chooses questions to capture users' personal views on a website's ease-of-use. Our questionnaire has been thoroughly tested and used on numerous websites to ensure the results point toward improving user-experience. The questions are standardised and may not be changed.*"[45]

We used WAMMI questionnaire because it has been developed specifically for websites, unlike other questionnaires that are more generic or more application-oriented. WAMMI questions are balanced and a procedure that will be explained subsequently is used to render the results more accurate, and easier to interpret. WAMMI reliability

ranges from 0.90 to 0.93, (reliability results are shown in Table A.2 page 84).

WAMMI uses twenty questions that are rated on a Likert scale. The questions are organised around five dimensions. The dimensions with the explanations as described in WAMMI reports are shown in Table 7.1:

Attractiveness	*An Attractive site is visually pleasant and also offers much of direct interest to the respondents, whether it be functionality or information.*
Controllability	*If a site scores well on Controllability the respondents most probably feel they can navigate around it with ease and do the things they want to do. Poor usually means a poorly organised site that disrupts the way they normally expect to do things.*
Efficiency	*When respondents give a high Efficiency rating they feel they can quickly locate and do what is of interest to them in a effective and economical manner. They feel that the web site responds (possibly, the pages load) at a reasonable speed.*
Helpfulness	*A site which is high on Helpfulness corresponds with the respondents' expectations about its content and structure. A site low on Helpfulness can be misleading about its layout and content.*
Learnability	*When Learnability is high, respondents feel they are able to start using the site with the minimum of introductions. Everything is easy to understand from the start. When Learnability is low, respondents feel that the site may be using concepts or terminology which are unfamiliar. More explanations are needed.*
Global Usability Score (GUS)*	*Global Usability centres round the concepts that a site must make it easy for respondents to access what they need or want from the site, that there is a good, understandable level of organisation and that the site 'speaks the respondents language' and meets their expectations.*

*Global is the mean of the five dimensions. The raw gathered data processed in order to make it more easily interpretable.

Table 7.1: Explanations copied from a WAMMI report. These are given as information as to how to understand in order to interpret the results of each category.

7.1.2.4 WAMMI data procedure

The following is the procedure used to transform the raw data collected by the question-naire into data that is presented in the five dimensions previously stated.

1. The first step is that individual responses are weighted. Weighting in statistics means that the responses are multiplied by a factor, in this instance the responses are multiplied with a coefficient that is correlated to the dimension to which the question belongs; thus high correlation means high coefficient.

2. The results are then indexed, and each dimension is the sum of the weighted results for that dimension. So each result is produced like this:

$$S_n = (Q_1 * C_1) + (Q_2 * C_2)...(Q_{20} * C_{20})$$

S= Score of dimension, C= coefficient of question

3. The scores are turned into percentiles. The use of percentiles is a method to express a score in comparison to a large pool of other scores in order to standardise the results and make them easier to understand. Thus the scores are compared to a database of results.

   This database has the results of usability tests done on many different websites with a lot of users. Percentiles are the percentage of results in the database that are *less than, or equal to* our own result. This percentage is rounded to the nearest whole number. [3]

A report is then generated with the transformed results and the list of comments the users have written about the site.

Zimmerman and Zumbo, argue that significance testing and data analysis can be applied to percentiles, even if it is widely discouraged by many textbooks: *"[...]Contrary to the received view described in many texts on tests and measurements, it is appropriate*

---

[3]Personal communication by Dr. Jurek Kirakowski for more information, see Kirakowski et al. [44].

*to conduct statistical analysis on the percentiles.[...]Depending on the population distri-
bution, using percentiles may or may not bring about a substantial gain in power, but at
the very least, one can be confident that their use will not lead unavoidably to incorrect
statistical decisions"*[99].

7.1.3 Results

This section will detail the results of the experiment along with the statistical manipula-
tions done on the results. In each section there will be a short synopsis to explain the
results given and what these results show.

7.1.3.1 Controlled population results

The controlled population used for the experiment consists of postgraduate computer
science students from Heriot Watt University (from now on referred to as *Group A*) and
porstgraduate usability course students from Napier University (from now on referred
to as *Group B*)

7.1.3.2 Group A results

Due to a lack or users, *Group A* cannot bear results of any statistical value. Low
aesthetics template for *Group B* has 5 users and high aesthetics template has 2 users.

   Thus we treated this part of the experiment as a pilot to see if there were any
glitches or other factors that we had not accounted for. Nothing worth mentioning was
discovered.

7.1.3.3 Group B results

Group B has 10 users for the low aesthetics template and 13 users for the high aesthetics
template.

   For more information is available in Tables A.4 and A.6(page 86) for the means,
table A.7 (page 86) for non-parametric significance tests.
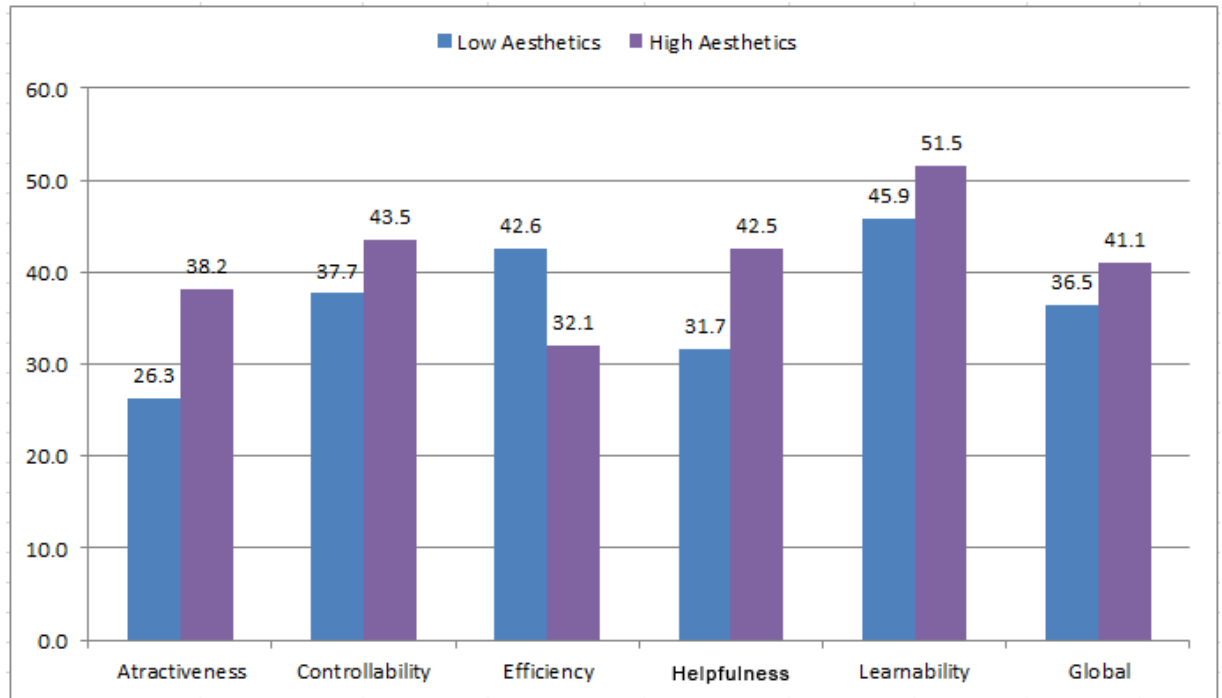
Figure 7.3: WAMMI questionnaire means from Group B.

The results indicate that the sample is too small for such small differences in the results. We can see differences of 2points but with a sample of 23 people, effectively two samples of 10 people the results are not significant enough to be sure that we are not just seeing noise in the data and differences that resulting from pure chance.

7.1.3.4 Open-question answers for Group B

Group B users' answers to the question, "*Is there anything you think is missing from this web site?*" are quite similar amongst themselves. Most of them do not seem eager to point out that anything is missing. Two users from the low aesthetics template mentioned the aesthetics of the site. For the high aesthetics template a student among other things mentioned that the template needed some corrections to make it look nicer. The full answers are shown in tables A.31 and A.31 (page 96).

7.1.4 Results of the open experiment's first questionnaire

The first questionnaire showed very interesting results. It was answered by 110 users, 55 for the low aesthetics group and 55 for the high aesthetics group. The group consisted of 49 males and 61 females. The age groups where divided initially into 4 groups, 15 to 24 years old, 24 to 38, 39 to 58 and 58 onwards. The last group was answered by too few people to get any statistically usable data. Thus we merged the two last age groups to form a larger group from 39-year-olds and above.

The list of questions used in the first questionnaire are as follows:

Question 1	It is easy to navigate through this website
Question 2	It is easy to find what I want on this website
Question 3	This website loads too slowly
Question 4	The graphics on this website are pleasing
Question 5	It is easy to use this site upon my first visit
Question 6	Clicking on links takes me to what I expect
Question 7	The organisation of information on the system screens is clear

Table 7.2: Questions from the first questionnaire of the open experiment.

According to the Mann-Whitney test, the two sets of values differ significantly for Question 4 and 5. Question 4, *low aesthetics*($\mathsf{mean} = 2.31$, $\mathsf{SD} = 1.263$) is significantly lower than *high aesthetics*($\mathsf{mean} = 4.02$, $\mathsf{SD} = 1.097$). (Mann-Whitney $\mathsf{U} = 478$, $\mathsf{Z} = -6.347$, $\mathsf{p} < 0.001$). Question 5 mean of *low aesthetics* ($\mathsf{mean} = 3.40$, $\mathsf{SD} = 1.285$) is again significantly lower than the mean of *high aesthetics* ($\mathsf{mean} = 4.11$, $\mathsf{SD} = 1.031$) (Mann - Whitney $\mathsf{U} = 1020$, $\mathsf{Z} = -3.064$, $\mathsf{p} = 0.002$). More details are shown in Figure 7.4 and Tables A.8, A.9, (pages 87, 87).
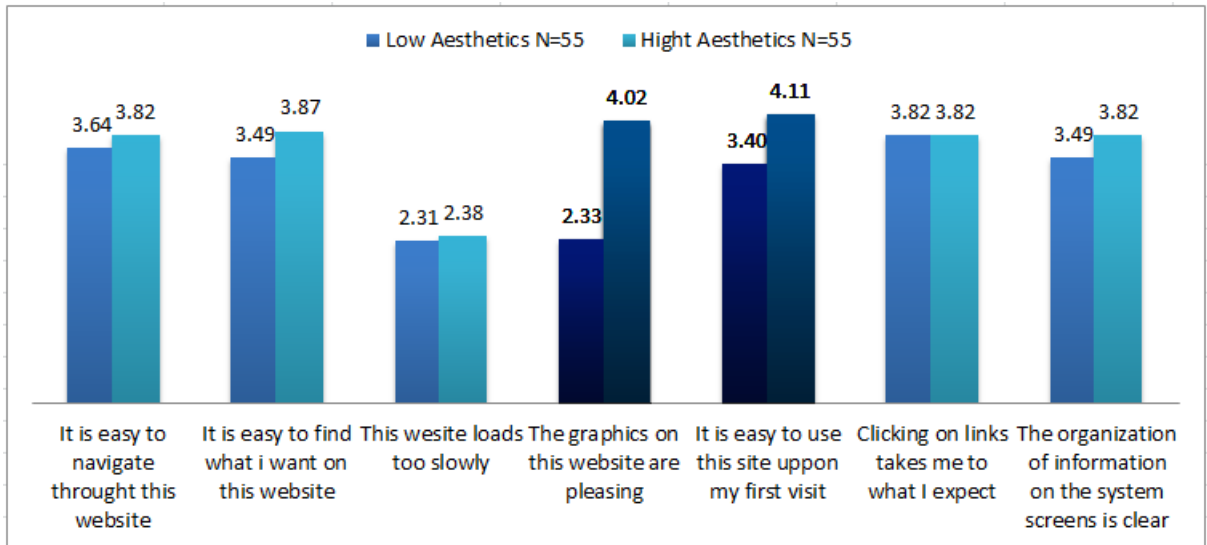
Figure 7.4: Means of the 'low aesthetics' and 'high aesthetics' templates, the statistically significant differences are in bold.

There are no statistically significant differences between genders. All *asymptotic significance* values are higher than 0.05 (Table A.10 page 87).
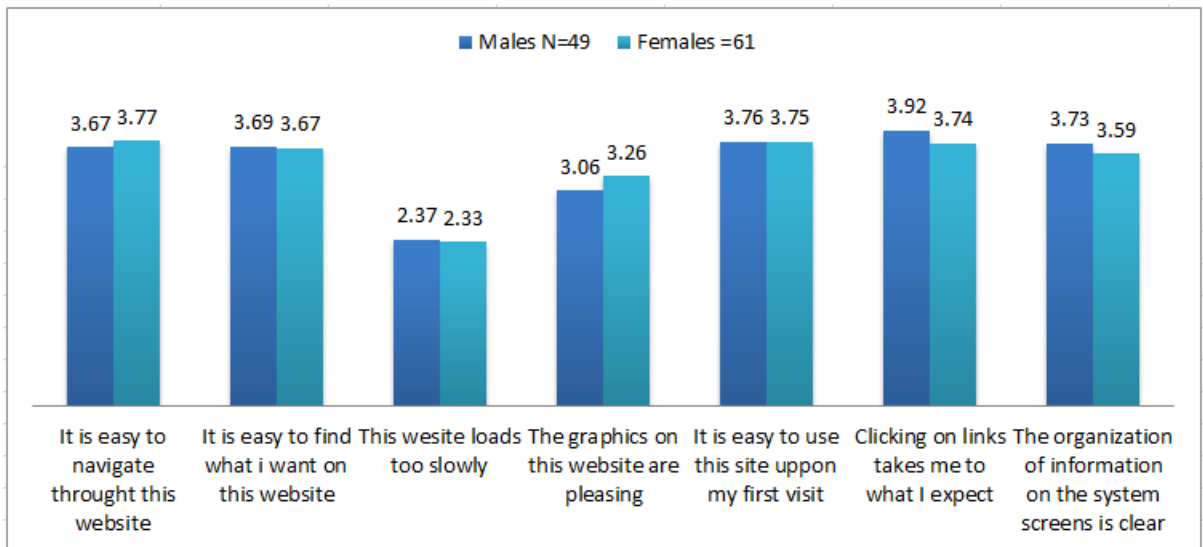
Figure 7.5: Means of Genders

Age groups, have statistically significant differences. According to the Kruskal Wallis test, mean values differ significantly for question 5 between age categories. Means for *15 to 24, 24 to 38* and *39 onwards* are respectively ($\text{mean} = 3.58$, $\text{SD} = 1.261$), ($\text{mean} = 3.64$, $\text{SD} = 1.240$), ($\text{mean} = 4.39$, $\text{SD} = 0.85$); (Kruskal Wal. $x^2 = 1.368$,

df $= 2$, p $= 0.043$). See chart 7.5 on page 63 and for more details see Table A.12 and A.13 on page 88.
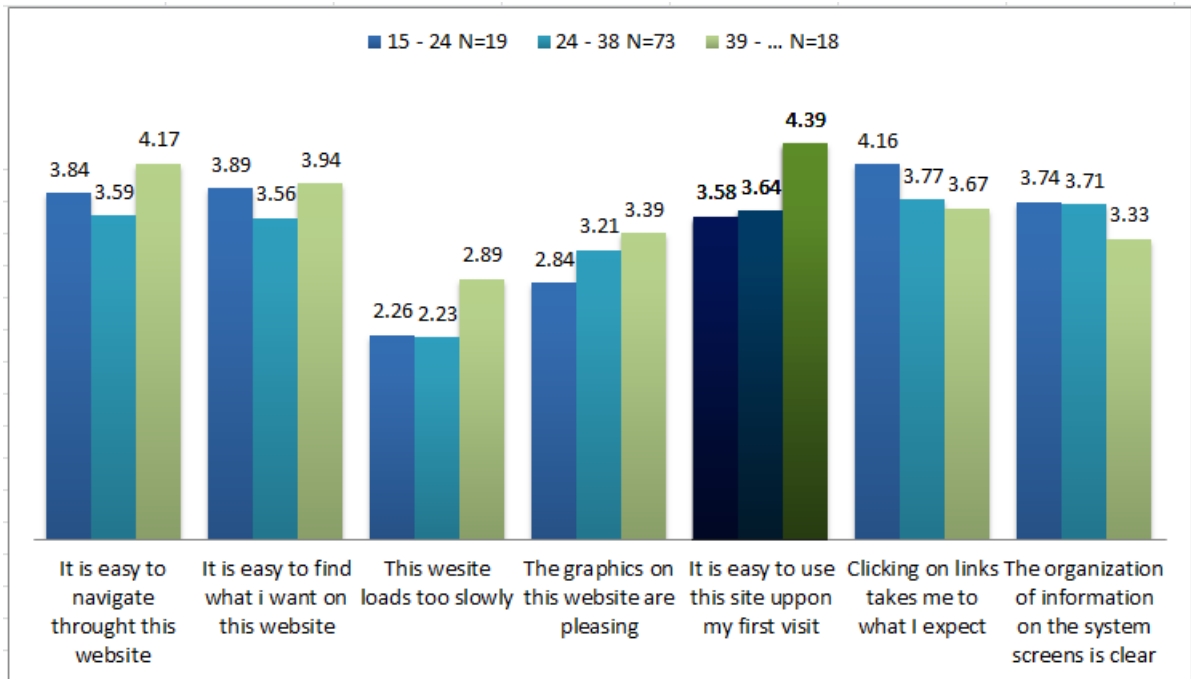
Figure 7.6: Means of age groups, with statistically significant differences in bold.

The degree of web design experience users have (the question is in the form of *how many web pages have you designed*) also showed some statistically significant variations in answers given by the users to question 4. Significant differences emerge in the means of groups: *none* (mean $= 3.44$, SD $= 1.274$), *occasional pages* (mean $= 2.82$, SD $= 1.527$) and *lots of pages* (mean $= 2.38$, SD $= 1.188$); according to (Kruskal Wallis $x^2 = 6.683$, df $= 2$, p $= 0.035$). Figure 7.7 (page 64) and Table A.15 (page 89) for more details.
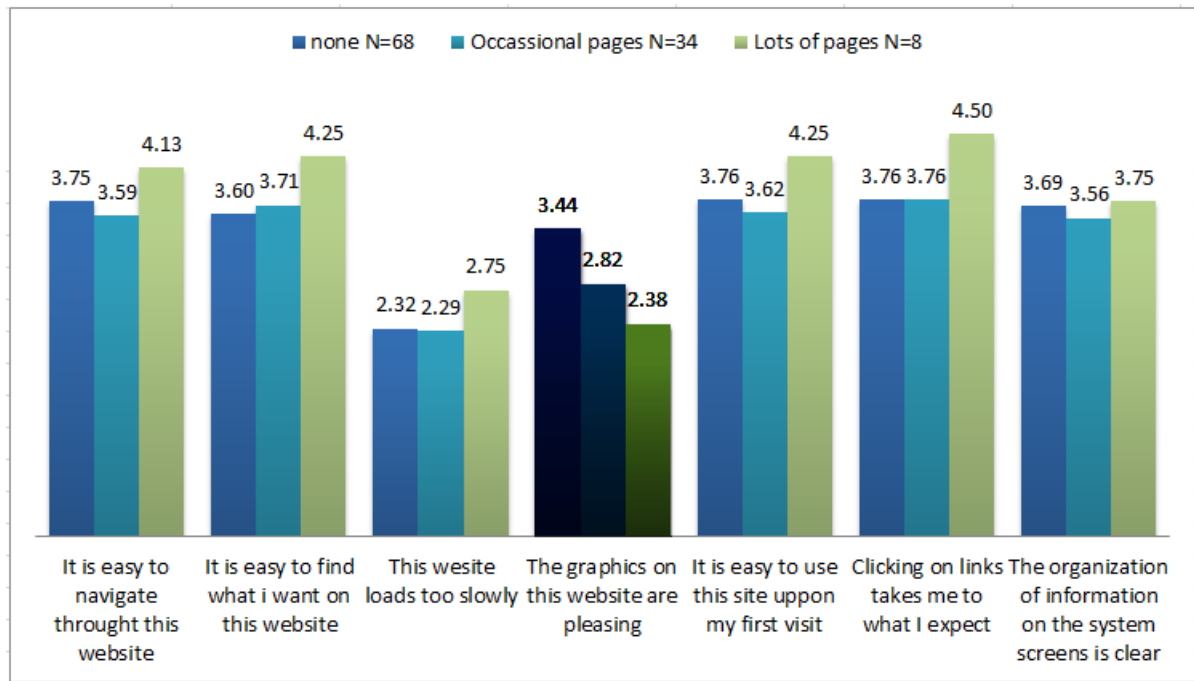
Figure 7.7: Means of web design experience, with statistically significant differences in bold.

We saw that the depending on *web design* experience, users rate the site incrementally more severely depending on how much experience they have. The more experience, the more severe they are. Thus we decided to select cases and see how users rate the two templates, low and high aesthetics, if we choose to test only users that had *no web design experience*. For both of these groups the question about how pleasing the graphics on the page were had a statistically important difference.

For users that have *no web design experience* the results for low and high aesthetics reveal one statistically significant difference and one trend. The results follows; more details are shown in Figure 7.8 (and Tables A.16 and A.17 on page 89 - 90):

It is easy to use upon first visit

   Low aesthetics ($M = 3.38$, $SD = 1.178$), High aesthetics ($M = 4.05$, $SD = 1.099$); (Mann - Whitney $U = 378$, $Z = -2.419$, $p = 0.016$).

It is easy to find what I want on this website

   Low aesthetics ($M = 3.31$, $SD = 1.198$), High aesthetics ($M = 3.82$, $SD = 1.189$); (Mann - Whitney $U = 422$, $Z = -1.839$, $p = 0.066$).
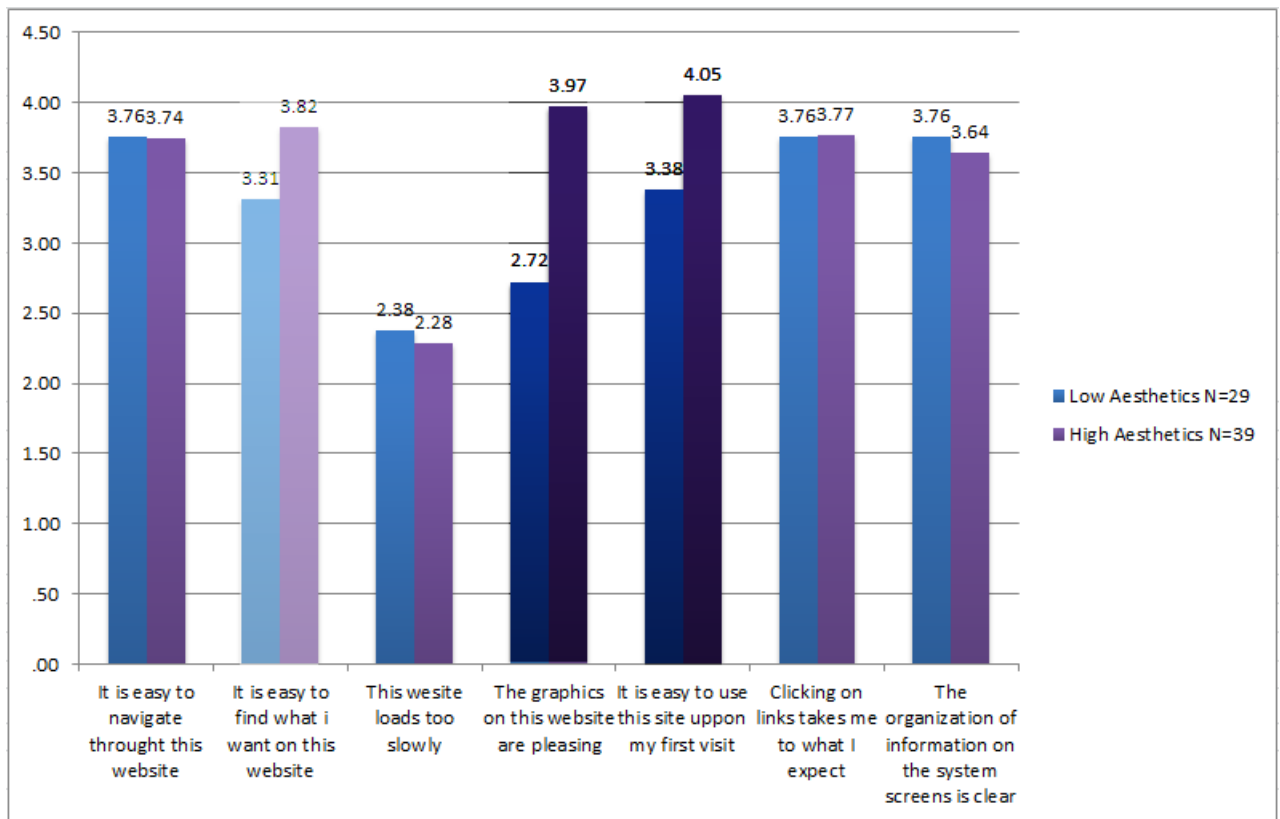
Figure 7.8: Means of users that have no web design experience. Bold signifies statistically important differences and lighter colour signifies trends.

For users that have *some or a lot of web design experience* the results for low and high aesthetics reveal, again, one statistically significant difference and one trend, but in different categories than those of users with *no web design experience*. The results are as follows, with more details shown in Figure 7.8 (and Tables A.18 and A.19 (page 90 - 90):

The organisation of this information on the system screens is clear

Low aesthetics ($M = 3.19$, $SD = 1.415$), High aesthetics ($M = 4.25$, $SD = 0.775$); (Mann - Whitney $U = 119.5$, $Z = -2.419$, $p = 0.016$)

It is easy to use upon first visit

Low aesthetics ($M = 3.42$, $SD = 1.419$), High aesthetics ($M = 4.25$, $SD = 0.856$); (Mann - Whitney $U = 136.5$, $Z = -1.951$, $p = 0.051$)
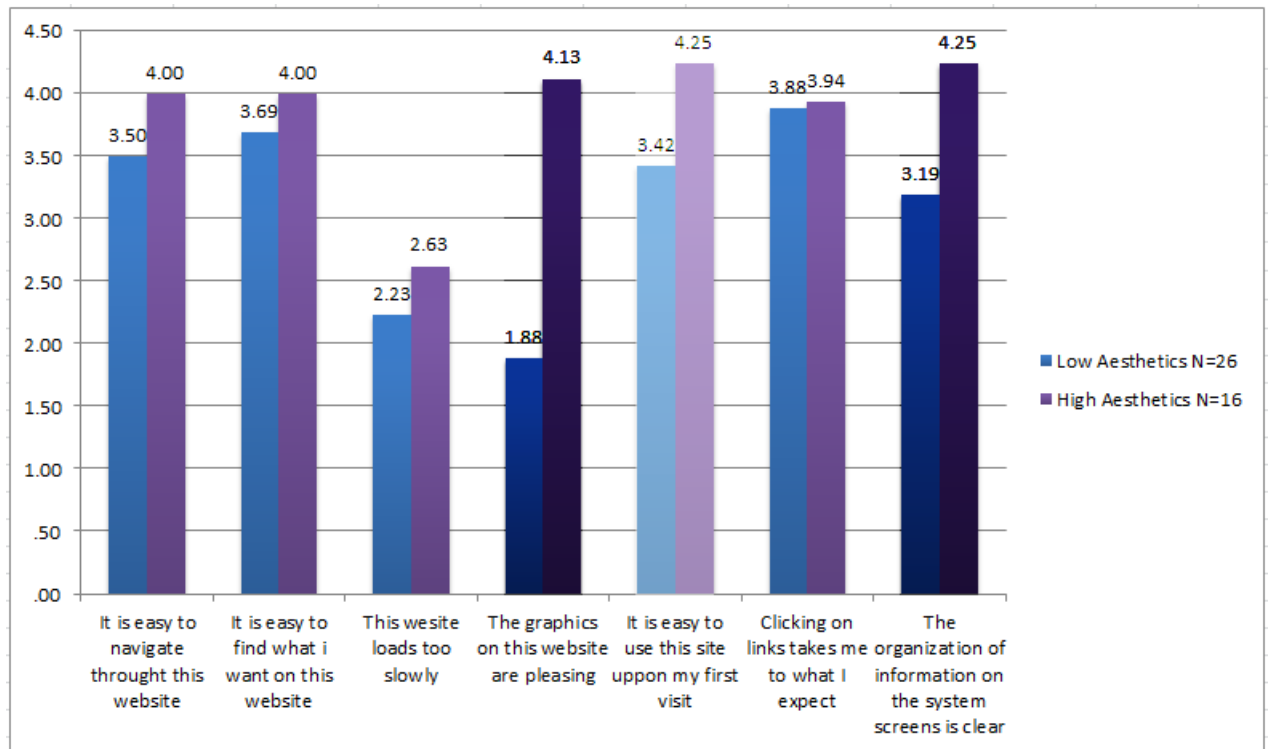
Figure 7.9: Means of users that have web design experience. Bold signifies statistically important differences and lighter colour signifies trends.

7.1.4.1   Short synopsis of results of open experiment and first questionnaire

As anticipated, users aesthetically preferred one template from the other. What we considered "*Low Aesthetics*" was perceived as *uglier* than the "*High Aesthetics*" template, as was anticipated. The second difference is in the question *It is easy to use this site upon my first visit.* The other questions have statistics with the same results.

With respect to gender there is a sufficient number of subjects that is more or less evenly distributed, 49 male and 61 female subjects. There seem to be no gender issues; by this we mean that there are no statistically significantly different results between the two genders.

Age groups reveal a statistically significant difference in the question *It is easy to use this site upon my first visit*, with older people ranking the site as easier than younger people.

Web design experience also reveals a statistical difference as people with more design experience ranked the sites lower than people with no experience.

By selecting cases of users that have some web design experience and users that no web design experience, we can see that there is a significant difference in how these two categories answered. Each category revealed, apart from the aesthetic category which was significantly different for both groups, one statistically different category and one category that shows a trend to differ. But the categories were not the same for the two groups.

Users with no web experience found the high aesthetics template significantly *easier upon first visit* and showed a trend in rating it easier *to find what they want on the website.*

Users that have web design experience rated significantly higher the *organisation of the information on the system* for hight aesthetics while they only showed a statistical trend for *easiness of use upon first visit.*

7.1.4.2 Statistical results of WAMMI questionnaire

The results of WAMMI, as previously mentioned, are given as reports. These reports are in the form of percentiles, as also stated previously, along with the procedure to get these percentiles, as explained in Section 7.1.2.4, (page 60).

According to the Shapiro - Wilk test, our data is not normally distributed and as such *non-parametric* statistical treatments are in order, (Tables A.20 and A.21 page 91).

According to Mann - Whitney U our results have one set of values that significantly differ and we can identify a *trend* in a second set. Values differ significantly as to how users perceived the attractiveness of the two different templates, as expected and was found also from the first questionnaire. *Low aesthetics* ($mean = 31.19$, $SD = 25.541$) is lower than *high aesthetics* ($M = 50.25$, $SD = 29.149$), (Mann-Whitney $U = 776.5$, $Z = -3.134$, $p = 0.002$).

In Global the results are not statistical significant, in the *low* and *high aesthetics* set

of values. *low aesthetics* ($M = 39.17$, $SD = 20.494$) is lower than *High Aesthetics* ($M = 48.06$, $SD = 23.385$); (Mann - Whitney $U = 966$, $Z = -1.807$, $p = 0.071$). Presumably a larger sample would lead to a statistically significant result. More details are shown in Figure 7.10 and Tables A.22 and A.23 (page 91 - 92).
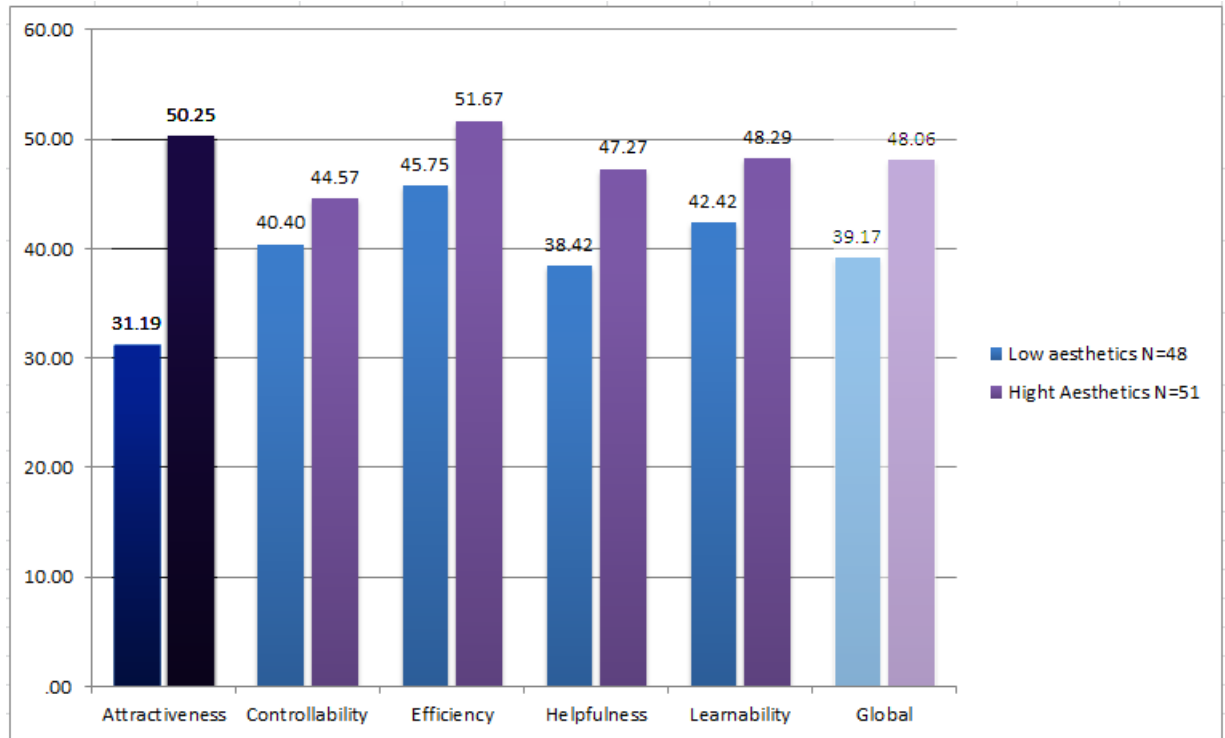
Figure 7.10: Means of low and high aesthetics are in bold because they are statistically significant. Means of Global are in lighter colour because there is a trend to difference.

7.1.4.3   Select cases according to interest in the site

The WAMMI questionnaire has two extra questions concerning the population:

- *How important for you is the kind of site you have just been rating?*

- *How would you rate your Internet skills and knowledge?*

We selected cases of people who showed a strong interest in the domain of the site, to eliminate users that were not in the target audience of the site. The users selected are the ones that answered the question *How important for you is the kind of site you have just been rating?* either *extremely important*, although none actually did, or that *they would need it sometimes.*

Normality tests indicate that the data is not normally distributed for all the categories. For more details see Tables A.25 and A.26 (page 94).

We will start by presenting non parametric results. High and low aesthetics groups differ significantly according to Mann - Whitney U for *Attractiveness*, *Helpfulness*, and *Global*. Figure 7.11 shows the means of the categories that differ significantly between low and high aesthetics.
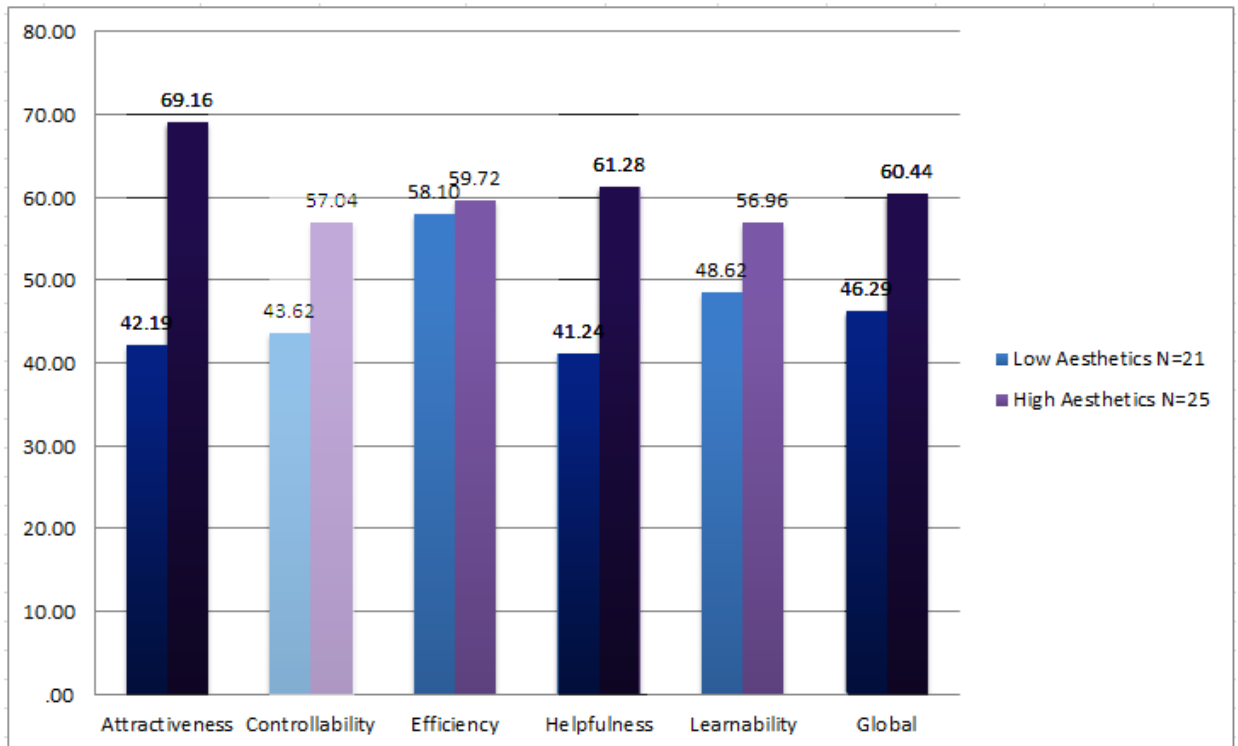
Figure 7.11: The means of WAMMI questionnaire after selecting cases that show interest in the site. Categories where there are statistically significant differences are in bold, lighter graphs show a trends towards differences.

The results are as follows:

Attractiveness,

Low aesthetics ($M = 42.19$, $SD = 24.94$), high aesthetics ($M = 69.16$, $SD = 23.79$); (Mann - Whitney $U = 114.5$, $Z = -3.27$, $p = 0.001$)

Helpfulness,

Low aesthetics($M = 41.24$, $SD = 25.24$), high aesthetics ($M = 61.28$, $SD = 26.84$); (Mann - Whitney $U = 345.5$, $Z = -3.266$, $p = 0.001$)

Global,

Low aesthetics($M = 46.29$, $SD = 19.11$), high aesthetics ($M = 60.44$, $SD = 22.59$);

(Mann - Whitney $U = 163.500$, $Z = -2.18$, $p = 0.029$).

Controllability differences between sets of values show a trend rather than being significantly different, low aesthetics ($M = 43.62$, $SD = 25.59$), high aesthetics($M = 57.04$, 28.734) (Mann - Whitney $U = 186$, $Z = -1.688$, $p = 0.091$)

For more details see Tables A.27 and A.28 (page 94 - 95).

7.1.4.4   Short synopsis of statistical WAMMI results

WAMMI statistical results indicate two points:

- As expected, the two templates are perceived as aesthetically different.

- There is a trend to differ in the Global category.

The WAMMI questionnaire describes its Global category as such: *"Global Usability centres round the concepts that a site must make it easy for respondents to access what they need or want from the site, that there is a good, understandable level of organisation and that the site 'speaks the respondents language' and meets their expectations"* see Kirakowski et al. [44].

And here the results indicate that the users' *trend* is to think that globally the *high aesthetics* category is more *usable* than the *low aesthetics* one.

We then selected the users that showed an interest in the domain of the site and thus are its target group. This made the results more powerful. Attractiveness, as expected, is significantly different. Two more results show that users rated significantly higher the categories of *Helpfulness* and *Global*:

- Users rated the site as being significantly more helpful when using the high aesthetics template.

- Results show that globally the site is perceived as more usable when using the high aesthetics template.

7.1.5   Qualitative results

The WAMMI questionnaire has two open questions:

- *What feature of this web site do you think should be improved and why?*

- *What do you think is the best feature of this website and why?*

Furthermore WAMMI compiled a list of strengths and weaknesses of the site along with a list of phrases that describes what the users thought. This last list is presented as answers to the question *Visitors think that: ...* in the Appendix A.1.4.2 (page 98), along with all the tables with the answers to WAMMI open questions.

To the question *What could be improved?* the answers between low and high aesthetics groups were very different. Low aesthetics group users in their majority proposed some form or feature of the design to be improved. They did that either by simply saying that the design needs improving or by asking for more colours, other fonts and so on. The high aesthetics template answers were far more varied. The main issues raised were: the introduction to the website was not as clear to them as it could have been, navigation was also mentioned a few times and, finally, some users did not specify any improvements. As best feature of the website, users from both groups, low and high aesthetics consistently pointed out the content. There were no substantial differences in what the users thought was the best feature. So the level of aesthetics did not seem to influence their answers.

The list of *Visitors think that...* answers was very similar for both templates. But the reports for the two templates show no major strengths as well as no major weaknesses for the website. By comparing the compiled list of what the visitors thought of the website (Table A.32 page 98) we become aware that there is almost no difference.

7.1.6   Quick Synopsis of the data results

In this Chapter we described the experiment that we performed. We reported the procedures given to the users and the questionnaires they had to fill. We explained

the manipulation done on the data for the WAMMI questionnaire to create the reports given. We saw that the results were given in the form of percentiles.

We then proceeded to give the results of the controlled population groups and the open population from the two questionnaires respectively. Furthermore, we reported the results from the open questions and outlined the main answers given to the afford mentioned questions.

Both questionnaire results showed a strong preference towards the high aesthetics template. This was crucial to the experiment. The experiment was designed around this first and main assumption, that the stripped-down version of a template would be found less aesthetically pleasing than the normal template. This first assumption was met and this result is statistically significant in the results of both questionnaires.

The results of the first questionnaire indicates that users found the high aesthetics template easier to use on their first visit.

By dividing users into two groups, one with users that have no web design experience and the other with users that have some web design experience, we got some extra results, as stated in 7.1.4.1:

- Users with no web experience found the high aesthetics template significantly *easier upon first visit* and showed a trend to rate it easier *to find what they want on the website.*

- Users that have web design experience rated significantly higher the *organisation of the information on the system* for high aesthetics, while they only found statistical trend for *easiness of use upon first visit.*

When looking at the entirety of the population, the WAMMI questionnaire gives us only a trend. It seems that users tend to think of the high aesthetics template as globally more usable but the result is only significant to $p \leqslant 0.1$ and thus only a trend. By removing the users that stated that they were not interested in the domain of the site, some of the results became statistically significant. Users rated the high aesthetics

template better for helpfulness and for global usability. Normality tests showed that the results of the users interested in the domain of the site were normal.

This suggests that people that were not interested in the website did not rate it in same manner as users that were interested. Maybe their lack of interest towards the domain of the site contributed their not answering questions as seriously and made them answer more randomly. In any case users that were not interested in the domain of the site answered differently that users that were.

So the three factors that we found to change significantly between low and high aesthetics templates are the following:

- Ease of use upon first visit.

- Helpfulness[4]

- Global Usability Score[5]

- Tendency to find things around the website more easily[6]

- Organisation of information on the system screens is clear[7]

The factors that were found to have an effect on these findings were:

- Age group

- Interest in the domain of the site

- Web design experience

---

[4]For people that were interested in the domain of the site. Helpfulness is used as defined on the WAMMI list found in Table 7.1 page 59.

[5]For people that were interested in the domain of the site. Global Usability Score is used as defined on the WAMMI list found in Table 7.1 page 59.

[6]For people that have no web designer experience, please see Section 7.1.4.

[7]For users that do have web experience, see Section 7.1.4.

7.1.7   Discussion of the results

In this subsection we critically discuss the results of the experiment and contrast them with our initial hypothesis.

The main hypothesis of this document, as stated in subsection 2.2, on page 3 is:

- For two identical systems in functionality and usability, differences in aesthetics may positively influence users perceived usability of the site.

We also theorised that previous work on the field needs to be revisited because a too-narrow focus on engineering aspects of aesthetics has influenced the scope and success of experiments to date.

In relation to our initial hypothesis, there are no negative results from our experiment. The results of our experiment showed statistically significant differences from high aesthetics to low aesthetics in the following categories: Thus, our initial hypothesis is supported, however the results are affected by factors that we discuss bellow.

- Ease of use upon first visit

- Global Usability Score[8]

- Helpfulness[9]

- Tendency to find more easily things around the website[10]

- Organisation of information on the system screens is clear [11]

The following factors that were found to have an effect on these findings were:

- Interest in the domain of the site

---

[8]For people who are interested in the domain of the site. Global Usability Score is used with as defined by WAMMI list found in Table 7.1 page 59

[9]For people who are interested in the domain of the site. Helpfulness is used as defined by WAMMI list found in Table 7.1 page 59

[10]For people who are have no web design experience, see subsection 7.1.7.5

[11]For users who do have web design experience, see subsection 7.1.7.5

- Web design experience

- Age group

From the within experiment the statistically significant results are the following:

- How easy the users thought the site was before use

- How easy the users thought the site was after use

- How easy the navigation was

- How easy they thought the site was at first visit after using the site

- How accurate the information was between the two sites

Of the statistically significant results the following are the most supportive of our initial hypothesis.

7.1.7.1   First visit

The first statistically significant result to consider is *Site is easy to use upon first visit*. This result is statistically significant considering the entirety of the user population used in our experiment. Our data shows that when first used, between the two templates, the high aesthetics one is perceived as easier to use. This result supports our main hypothesis and partially confirms the results of Hartmann et al. [89] and Lindgaard et al. [51]. This is understood as perceived usability *upon first visit* and it refers to the initial reaction when participants first used the website. First impressions and visual appeal of a webpage have been revealed to be very important for a variety of issues [51]. Lindgaard in the same paper concludes that, "*first impressions form quickly and are consistent*".

We theorise that there is a habituation effect that does not carry over to this perception as to the overall usability results. The habituation effect is well known across many fields of research: any sensory stimulus that triggers a response from a subject,

when repeated, decreases the subjects' behavioural response. Groves and Thompson state: "*Habituation, defined simply as decreased response to repeated stimulation, is perhaps the most elementary and ubiquitous form of behavioural plasticity*" [31]. In usability, habituation's effect can be studied with a longitudinal study; a study that is repeated over time with the same users in order to see changes of responses during the experiment. According to Medonza and Novick, the results of their longitudinal study suggest that "*conventional usability tests catch causes of frustration that represent entry barriers for novice users rather than fundamental problems with an application's usability*"[55]. These results partially confirm that high aesthetics do improve perceived usability when the users are first visiting the site. This is an important find because during a first visit, a suer will decide if he or she they is going to stay or leave. A higher perception of usability upon first visit would contribute to convincing new users in staying and using a site.

We are not studying the effects of aesthetics on usability over a long period, thus a limitation of this study is that it is not a longitudinal study. We do not know how long *first visit* really is: First glance? First few pages?

In that effect the second experiment, will use a within design, and ask questions to clarify what first visit is.

Thus further research in this topic, in relation with aesthetics and usability, would be needed to find out what happens happens with the relationship of aesthetics and usability during long periods of use (days, weeks, or even months).

7.1.7.2 Global usability score (GUS)

Global usability score is the most significant variable to look at in relation to our initial hypothesis. When considering the entirety of the population that took part in the experiment we can only see a trend ($p \leqslant 0.1$) in users rating the high aesthetics template as more usable than the low aesthetics one. User results indicate, in a statistically significant way, that the high aesthetics template is globally easier to use than the

low aesthetics one but only for users that are interested in the domain of the site. As previously stated in Table 7.1(*page 59*) Global Usability Score is defined by WAMMI as follows: "*Global Usability centres round the concepts that a site must make it easy for respondents to access what they need or want from the site, that there is a good, understandable level of organisation and that the site speaks the respondents' language and meets their expectations*". But these results are only significant for part of the population.

As mentioned previously in subsection 7.1.4.3 (page 67), we have defined as users that have an interest in the domain of the website those that have answered the question: *How important for you is the kind of website you have just been rating?* with "*important*" or "*extremely important*".

Users need to be interested in the domain of the website in order for us to see statistically significant differences. After surveying the literature, we have found nobody else describing the impact of user interest in the domain of the website tested as being a factor reinforcing the link between aesthetics and usability. Our novel result is that user interest in the domain seems to reinforce the link between aesthetics and usability beyond the first visit. We discuss this in more detail in the next subsection.

7.1.7.3 Population factors that influence the results

In this part of the Chapter we describe the population factors that affect the way users have answered the questionnaires. As mentioned in the intro of this Chapter our questionnaires have revealed a very interesting set of results; users having common characteristics provide results that are statistically significant in categories other than just *first visit*. GUS was the first category we described and the most important one.

7.1.7.4 Interest

Results suggest that users that who interested in the domain of the website used for the experiment rated the site differently than users who were not particularly interested

or not interested at all. In past research done in aesthetics and usability we were not aware of anyone taking into account interest in domain as being linked to the quality of responses. Herzog and Bachman, in a paper about questionnaire length, state: *"Item sets which seem to prompt high levels of stereotypical responding deal largely with attitudes toward general social issues, which are presumably of less personal interest and require more attention and thinking"* [36]. Bosnjak and Tuten have a categorisation of how users or responders to a web questionnaire respond. A user's motivation to fill out a questionnaire could be *"due to an interest in the topic or the desire to comply with a request"*[14], but in all cases for users to respond there must be a triad of *"motivation, opportunity and ability"*. Unfortunately the paper does not point out any qualitative differences in the way users answer. Instead the paper considers users that answer some and not all of the questions. In our case our users had to fill in all of the questions in order to be able to submit the online form. Spool, in a paper about loosely structured interviews in usability testing and choosing participants that are passionate about the subject of the evaluation holds that users or responders that have a passion for the subject of the evaluation, would have a completely different behaviour in the evaluation to that in real life [82]. Interest in the domain of a website by users in using it, was a well-known fact; however, in the interplay of usability and aesthetics research it is a novel concept.

This find furthermore confirms our criticism of Sutcliffe et al. [89], (see subsection 5.4 on page 36). By asserting that interest in the domain of the tested webpage is an important factor that influences user choices when answering questionnaires, we confirm that using students to rate fundamentally different university departments can have unexpected results, due to the interference of users' interests and fondness of one subject over another. One of the standard practices in usability is to ask the users to role play, trying to imitate the reactions of someone else. This is what Sutcliffe et al. do for their paper. However we concur with Spool when he gives an example of the difficulties that can arise from such a practice: *"imagining a trip to Disney isn't hard for many people.*

It's harder to get into the role of shopping for the ultimate retirement fund" [82].

7.1.7.5 Users' web design experience

Users that have some or a lot of web design experience were more severe in rating the aesthetics of the website across the two templates than users with no experience. The severeness was incremental with the experience (see Figure 7.7 page 64). By selecting cases we see that the two groups rate the templates differently, as stated in subsection 7.1.4.3: Users with no web experience found the high aesthetics template significantly *easier upon first visit* and showed a trend in rating it easier *to find what they want on the website.* Users that have web design experience rated significantly higher the *organisation of the information on the system* for high aesthetics, while they only found statistical trend for *ease of use upon first visit.*

We know that expert users react differently from novice users; we know that knowledge of the site itself, or having used it before, is important too [37, 60, 61]. In our case all our users were first-time users of the website, since it was created and launched for this experiment.

Users with web design experience have a double identity. On one hand, in order to have arrived at the point of designing websites they are experienced users. On the other hand they have participated in the creation process of a website, and know the mechanisms and technologies that are used to create a website. They have been on both sides of the fence, visited sites as simple users, have become expert users and now have created pages for users.

Since our results suggest that expert users react differently than novices, we theorise that the results suggest that designers, when rating a site, are not doing it as simple users but they do a sort of an *"expert review"*. They try to predict what users would think of the site.

However, the results that point to people with no web design experience were positively affected by high aesthetics in finding the site easier upon first visit. They also

tended to be positively affected in finding what they wanted on the website. People with web design experience perceived the organisation of the website to be positively affected by higher aesthetics. They also tended to find that the high aesthetics template was easier upon first visit. Both these results are positive results for our initial hypothesis; nonetheless, they highlight factors that have an impact on it.

7.1.7.6 Helpfulness

Helpfulness is another category that is positively influenced by higher aesthetics, but here again the factor of interest plays a major role. As stated in subsection 7.1.2.3 helpfulness is defined by WAMMI: "*A site which is high on Helpfulness corresponds with the respondents' expectations about its content and structure. A site low on Helpfulness can be misleading about its layout and content*". User results with an interest in the domain of the site find the high aesthetics template significantly more *helpful* than the low aesthetics template. Users that were interested in the content of the website are positively influenced by higher aesthetics; it makes them perceive the site in correspondence with their expectations in content but also instructure. This result suggests that aesthetics are a factor that users expect when visiting a website. Lack of aesthetics seems to have a negative effect on the perceived helpfulness of a website to the users.

Helpfulness is a concept that is part of the user experience (UE). It has to do with expectations, content and structure. UE stated in subsection 1.1 on page 1 is a more holistic manner at seeing beyond usability and considering user engagement [33]. The ISO 9241-210:2010 standards paper defines user experience (UE) a: "*person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service*[12]" [1]. All these contents are beyond the scope of pure usability.

---

[12]The definition comes with three notes:
—User experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments that occur before, during and after use.
—User experience is a consequence of brand image, presentation, functionality, system performance, interactive behaviour and assistive capabilities of the interactive system, the user's internal and physical state resulting from prior experiences, attitudes, skills and personality and the context of use.
—Usability, when interpreted from the perspective of the users' personal goals, can include the kind of perceptual and emotional aspects typically associated with user experience. Usability criteria can be

7.1.7.7 Age of participants

As described in the previous Chapter in subsection 7.1.4, page 61, across the two templates, users aged 39 and above have rated the sites with a statistically significant difference as easier to use upon first visit. Results suggest that users from 15 to 24 and 24 to 37 rated the site in the same way and significantly lower than 39 and above, for easy to use upon first visit. Our initial age categories were four not three. But since only two users were in the age group of 58 and older, we merged it with the previous age bracket. As such we do not have any meaningful data for senior citizens. Nonetheless the data seems to point that age is a factor in how users rate the site and could be a factor to consider in the relationship between aesthetics and perceived usability.

7.2 Second experiment

The first experiment revealed interesting results that needed to be further investigated. In particular the question of *upon first visit* needed further clarification. A second experiment was needed to further investigate these positive results. Furthermore a second experiment using a repeated measures design could validate the results of the first experiment. This chapter describes the design of a second experiment. It will report on the methodologies used, what led to these design decisions, how users were found to participate in the experiment.

7.2.1 Design of the experiment

The goal of this experiment is to explore the results of the previous experiment (chapter 7.1.1) and expand on them. As we previously mentioned in chapter 7.1.6 page 69 the main result of the previous experiment was *ease of use upon first visit*.

The "Bootstrap[13]" [7] framework was used to build two sites that would be easy to

---

used to assess aspects of user experience.

[13]Bootstrap 2.2.2 was used. Extensive information about Bootstrap can be found here: http://getbootstrap.com/about/

understand. The choice of building a site from scratch was made once more to keep control of the experimental procedure, rather than testing on already existing ones.

The main subject of the site was irrelevant to the testing, but it had to be understandable by all potential users. A fictional island as a holiday destination was chosen as a subject that would allow us to create a site that would need more than 10 minutes to go around if a user was to read all the texts and look at the pictures. The choice was made on account of its lack of previous knowledge required.

Questions in a multiple choice format were devised for the users to answer, rather than a compulsory protocol. This was to make sure that, the users did explore the site and were confronted with some minimal difficulties in use requiring some concentration on their part before they answered the usability questionnaire.

The user's task, as shown by the questions that appear on the right of the page (see figure 7.12 on page 75), is a task that does not require a specific skill set from the users. Furthermore it is a task that most users would be familiar with, its questions involve normal browsing, like online quizzes where the answer is somewhere in the text and lastly the task does not depend on the users having any specific interest.

Two versions of the same site were produced so the aesthetics on one of them could be made lower than the original one. For this the consistency of the fonts used were disrupted. More specifically a mix of serif and sans serif fonts were used for the low aesthetics site. We discussed lack of uniformity in section 4.3.2 on page 4.3.2. The heading of the page was made to be wider and take up more space than it needed. Instead of photographs, sketches created by the author, with no care as to the quality of the outcome were generated. The sketches were then treated in Photoshop with a filter to give them a unified look. (see screen grabs on figure 7.12 on page 75. A pilot test confirmed that this was the case. The results are also echoed in the results of the experiment as you can see further dow in section 7.2.1.3 on page 77.

A positive feature of building a custom site was that we could incorporate the questionnaire in a column on the right. The site was thus presented using a frame on the

right side of the website and a column where the questions remained visible to the users while they were looking for the answers in the site. When a question was answered it was greyed out and the next one appeared. This permitted the collection of data within the experience. The users did not log onto a different page or site to answer questions and so we could ask the questions about usability before use and while use of the site without disrupting the experience. The protocol for which questions to answer was also given to the user from within the site and thus the experience was not disrupted.

(a) High aesthetics site

(b) Low aesthetics site

Figure 7.12: Screen grabs of the two sites used in the experiment.

7.2.1.1 User collection

Users were found using Amazons' Mechanical Turk [9]. Randomness is assured since users were unknown to us. Groups were advertised as different Human Intelligent Tasks (HIT) and users were asked not to participate twice. For checking purposes their worker

ID number was requested to minimise double answers within the database. The incentive for filling up the questionnaire was the renumeration of the HIT through Amazon. Users recruited through Mechanical Turk, or "Turkers" as they are calling themselves, are found to be reliable test subjects and "*exhibit the classic heuristics and biases and pay attention to directions at least as much as subjects from traditional sources*"[71], furthermore the wages given for the task seem to have no effect on the quality of the data that is given [53].

7.2.1.2 The questionnaire

The questionnaire used in this experiment was designed to further explore the results that were revealed as significant from the previous experiment. The questionnaire comprises two stages; the first stage is to ask the users how easy they think the site is to use, then it asks them to use the site, then asks again what they now think of the site. After a more thorough usage of the site, guided by further questions, the users are asked to answer a final questionnaire. The list of questions users answered can be found in chapter A.2.1.1 page 103.

The experiment was involved 112 users all recruited through Amazon's Mechanical Turk. 52 subjects took part in the first group, that were presented with the site A(high aesthetics) and then with site B (low aesthetics), and 60 users took part in group b which saw site B and then site A. Group 1 had 25 female users and 35 male, and group 2 had 19 female users and 33 male. The age of the participants were, for group A, 12 users from 18 to 24, 38 users 25 to 38, 8 users 39 to 58 and 2 above the age of 59. For group 2, 16 users were 18 to 24, 28 were 25 to 38, 6 users were 39 to 58 and 6 users were above the age of 59. Data can be seen on the figure 7.13 on page 76.
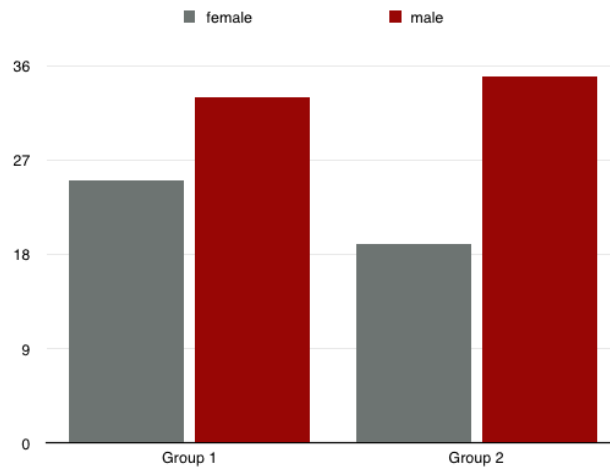
(a) Group 1 and 2 Male and female participants chart
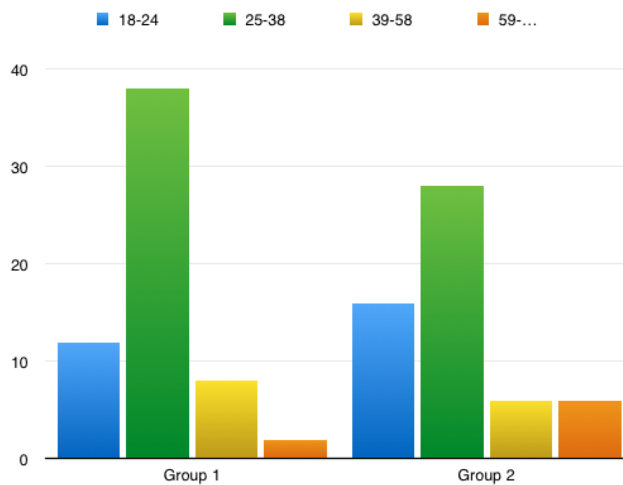
(b) Group 1 and 2 age categories of the participants

Figure 7.13: Charts visualising the gender of the participants and their age in each group that took part in the experiment.

The users are presented first with one site and then with the second. The pages load automatically on the right part of the page as users progress through the questionnaire and the tasks.

Worker ID numbers were collected only to ensure that no user took part in the experiment more than once. Along with the qualitative data collected from the questionnaire, the site also collects the number of mistaken answers that users give when responding to the guiding questions.

7.2.1.3   Piloting the experiment

The experiment was piloted by 32 users for group A and 29 users for group B (to counterbalance any ordering effect) who were students from Heriot Watt university. While it did uncover as most pilot tests do, some problems in the initial websites, it did confirm that *aesthetics* differed in a statistically significant way, between high and low aesthetics versions of the site. A misplaced title made the results of group B unusable as it was confusing the users as to which site they were rating.

Results of the pilot testing, according to Shapiro-Wilk were not normally distributed. Thus non parametric treatments were applied and showed that high aesthetics template was indeed rated as higher in a statistically significant way.

Aesthetics of the sites

Low aesthetics ($M = 2.41, SD = 1.13$), high aesthetics ($M = 5.72, SD = 1.25$); ($Z = -4.85, p < 0.01$)

Furthermore, for the pilot tests with the exception of the questions *How easy do you think this site would be to use?* and *How easy do you think this site is to use after having visited it?* questions about site A and B were asked at the end after the users had visited both sites. Asking users to rate site A and B after seeing both led to some confusion. Subsequently after the piloting of the experiment, the order of the questions was changed to avoid users getting confused between the two sites, low and high aesthetics. The questionnaire was placed just after the procedure that took the users around the sites. In that manner they were answering questions about the site they had just visited so that there was no confusion.

7.2.2   Data analysis

In this section we describe the statistical treatment of the data gathered. A short synopsis at the end of the chapter states the findings and the main points of this chapter.

A discussion about the results in relation to the hypothesis as those of chapter 7.1.2 will be presented in chapter 8

7.2.2.1   Results

This section will detail the results of the experiment. In each section there will be a short synopsis to explain the results given and what these results show.

According to the Shapiro - Wilk test, our data is not normally distributed and as such non parametric statistical treatments are in order (table A.38 on page 108).

Wilcoxon signed-ranks test indicated that there were statistically significant changes in the following questions (details see appendix tables A.42 on page 129).
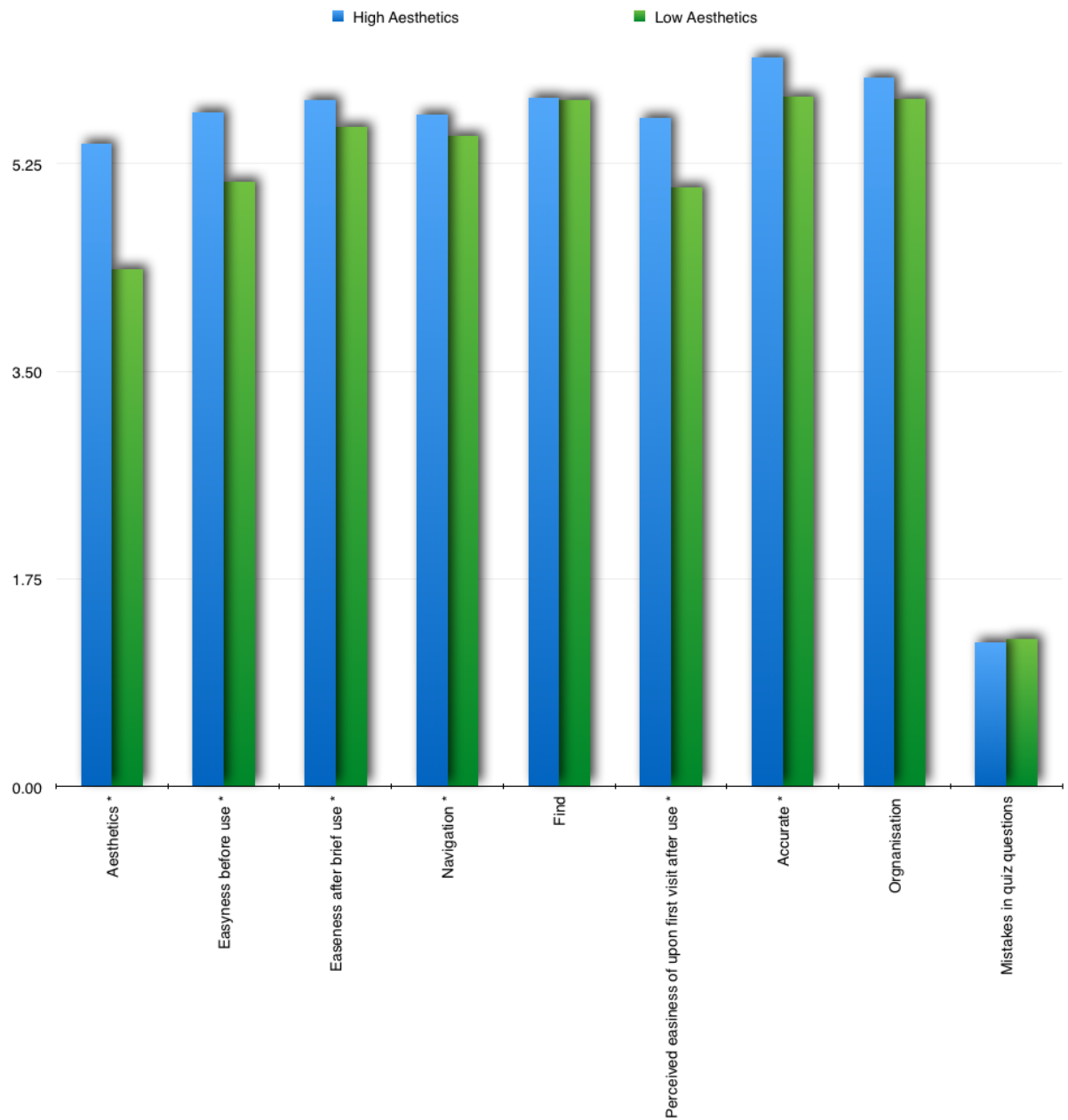
Figure 7.14: Means of second experiment. Bars with asterisks have statistically significant differences

Aesthetics of the sites

Low aesthetics $(M = 4, 63, SD = 1.8)$, high aesthetics $(M = 5.42, SD = 1.46)$; $(Z = -6, 270, p < 0.001)$

How easy the users thought he site was before use

Low aesthetics, $(M = 5\ SD = 1.41)$, high aesthetics $(M = 6,\ SD = 1.08)$;

$(Z = -4,581, p < 0.001)$

How easy the users thought the site was after use

Low aesthetics, $(M = 6, SD = 1.10)$, high aesthetics $(M = 6, SD = 1.04)$; $(Z = -2.442, p = 0.015)$

Navigation, how easy it is to navigate through the site

Low aesthetics, $(M = 6, SD = 1.34)$, high aesthetics $(M = 6, SD = 1.30)$; $(Z = -2.129, p = 0.033)$

Please rate how easy you thought it was to use site B when first opening it? (After having used it)

Low aesthetics, $(M = 5, SD = 1.46)$, high aesthetics $(M = 6, SD = 1.24)$; $(Z = -4.613, p < 0.001)$

Please rate how accurate the phrase is for website A or B: Clicking on links takes me to what I expect

Low aesthetics, $(M = 6, SD = 1.26)$, high aesthetics $(M = 6, SD = 1.01)$; $(Z = -3,204, p = 0.001)$

7.2.2.2 Comment on mistaken answers

As stated previously the site also collects also the mistaken answers to the multiple choice questions. The experiment did not produce any findings that real usability was being altered as user mistakes did not differ in a statistically significant way between the two versions of the site.

7.2.3 Synopsis

In this section we have reviewed the within experiment which has researched the results of the first experiment. Our first finding confirms the pilot test of the experiment. Aesthetics is indeed higher in a statistically significant way for the high aesthetics website version.

Several positive results were found in accordance with our primary hypothesis (see section 2.2 page 3). The experiment did not reveal any negative results. These results were, *perceived easiness before using the site, after using the site only briefly (after visiting two pages of the site), perceived ease of navigation, and perceived accuracy of links as well as users conserving an image that the site was easier to use on first visit after having visited the website..*

In this experiment the questionnaire was constructed so as to find out not only about our initial hypothesis: is a site perceived as easier to use of the aesthetics are high than when the the aesthetics are low, but also to give more detail on the findings of the the first experiment about ease of use on first visit. The positive results of this experiment are that when users log on to the site for the first time they do indeed perceive high aesthetics as easier to use.

The users were asked at their first log in for each site, low and high aesthetics, how easy they thought it was to use. They were then asked again after a brief tour around the site. Finally then after having visited the site and answered all questions they were asked again what they thought of of the site when they first visited it (the questions asked can be found in section 7.2.1.2 on page 76). All three questions gave statistically significant results that support our hypothesis.

Further more after using the site to answer two simple questions, they still find the high aesthetics pages easier to use than the low aesthetics ones.

Two further findings that support the hypothesis are that users find it easier to use a website with high aesthetics, that they find it easier to navigate through the high aesthetics site and that they also find that links take them where they expect, which echoes the previous finding also.

Discussion and Conclusion

8.1 Experimental novelty, limitations of current work and proposed further work

We have shown from our previous discussion that this is a challenging area in which to design experiments. Form, function and content are very hard to separate. Previous experiments had novel and interesting results which deserved further exploration; Kurosu and Kashimura were the first to publish a paper on the relationship between aesthetics and usability (as cited by Norman [67]) and Tractinsky's repeat experiment [91] is one of the founding experiments of the field. In this work we have pointed out some issues we thought were experimental weaknesses in past experiments, (see Section 5.6). In our experimental design we suggested and implemented methods to overcome those weaknesses.

The first issue that we addressed is *ergonomics*. In Section 5.1 we explained that we believed ergonomics to be misconstrued for aesthetics. As such in our experiment we overcame this problem by keeping all ergonomics the same. Buttons, links and content

are in the same position in both templates.

In Section 5.2 we explained why we believed that aesthetics cannot be deconstructed into heuristics. We overcame the issue of high and low aesthetics stripping a template of its embellishments, changing slightly the proportions of the different elements but keeping everything in the same place.

In Section 5.3 and Section 5.4 we stated that we thought that sites needed to be stylistically and ergonomically the same, but aesthetically different. In our experimental design, by stripping the embellishments of the website did not create major stylistic differences; the content was the same as are the ergonomics of the site.

A first limitation of our work was a forced choice. The two questionnaires used were not connected. This was due to the fact that the WAMMI questionnaire is handled by its owners and the two databases gathering the data were completely separate. Nevertheless under different circumstances it would have been interesting to be able to follow users' answers from one questionnaire to the other.

As mentioned before (section 7.1.4 page 61), the age of participants seemed to be important as a factor affecting the results of the study. There were very few older participants in our experiment, thus not allowing the collection of results for senior users and how they regard the relationship between usability and aesthetics. Our experiment had only two users aged 58 and above. It might prove to be an important factor for consideration, since eyesight and interest in new technologies, just to mention two, might be variables that affect the relationship between aesthetics and usability, further research targeting senior users would be recommended.

One limitation of this experiment, is the number of users participating, 110 users, 55 for the low aesthetics one and 55 for the high. Whoever has carried out some user testing knows that users willing to participate in such research are hard to come by. We have statistically significant results, but more users would refine the results and maybe reveal more trends. Small or more sensitive differences in the data can only be proven statistically significant when using a large number of users. A larger sample size could

prove that GUS (Global Usability Score) is not a statistical trend but a statistically significant difference. Other relationships could also arise.

There is a known trade-off between the number of questions asked and the accuracy of the answers in relation to the motivation that users have for answering [36]. In the first experiment we had restricted the number of questions in our first questionnaire, since the second, WAMMI, is not modifiable and a further two comments are asked of the users. We used a very quick usability questionnaire that has only 7 questions (listed in Table 7.2 page62). There is no clear answer as to how long a questionnaire can be before it starts putting people off, or start generating randomly answered questions, or a serious increase in the drop-out rate; what seems of more importance is the motivation of users to participate , [16]. Asking more questions posed a risk of demotivating users especially if they were not interested in the domain of the site. However it would be interesting to find out more about other factors that might have influenced the results, for example individual interest. It would also be interesting to see why web designers have answered questions differently than users that had no web design experience. It would be consequently interesting to ask questions concerning the users views in order to understand why these differences arise. A possible idea would be to split the questions between the users of the same group as suggested in paper by Maenpaa et al. [74]; to give a more concrete example, in group A give questions 1 to 20 to half the group and questions 21 to 40 to the other half.

Many experiments use objective measures; however our first experiment was not the kind of experiment that one could collect objective measures for, moreover we were interested in keeping perceived usability and real usability constant in both templates. The second experiment, with a different website was nevertheless not work oriented or information-intensive. Thus objective measures such as clicks, times would not have been relevant. Nevertheless this site did record users' mistakes in answering the quiz questions, however the results were not statistically significant.

While we believe that our first experimental design was sound, a between design, the

second experiment serves not only as a repeat experiment using a different experimental design (an inverse repeated measurements design,) it also that gathers further evidence and confirms our results.

As stated in the previous chapter (7.2 page 73) the second experiment used a repeated measures design. The experiment was split into two groups to counterbalance transfer effects.

8.1.1   Mental maps of the visited site

Our results from the first experiment, as discussed above, reveal user interest in a domain as an important factor in the research into aesthetics and usability. We hypothesise that interest is a motivator and a significant part of the user experience. It is important to do further research in the domain of user interest and questionnaire answering. Our results suggest that interest enables aesthetics to make a difference in how we perceive usability, but this needs to be further researched.

The results from the second experiment confirm the results of the first experiment and add new results. Our results as presented in the previous chapter, show that the high aesthetics template positively influences not only perceived usability before use, but also after a first very limited use, in this case answering one easy question from the multiple choice quiz questionnaire after browsing through two levels in the site and following two links. At the end of the manipulation the users still perceive that at first use the high aesthetics template is easier to use. These are positive results that do confirm findings from the first experiment and add more detail.

Firstly, the first experiment shows that a user's first look at the site is positively influenced by high aesthetics; they find it more usable.

However at this point, as confirmed by the second experiment, means that they still do not know the domain of the website and are only beginning to form a mental map of it.

The results of the second experiment suggest a cognitive interpretation is in order.

Higher aesthetics may influence users to have better expectations of the site and higher confidence about their future interactions. The perceived usability of websites does seem to cary halo effects that bear some resemblance to the ones described by Nisbett and Wilson [64]. The experiment has advanced the understanding on user first judgements and perceived usability when users log onto a site for the first time. Furthermore results are in accordance with Hartmann's findings [32].

We propose an encompassing hypothesis as a possible outcome of our results. We could formulate the theory that high and low aesthetics create a tipping point mechanism response in users. So a likely case scenario would be: a user logging for the first time onto a site depending on the aesthetics has different expectations. High aesthetics would generate higher user expectations. Conversely when a user logs onto a site with low aesthetics, thus would generate low user expectations. What we suggest is that users make different quality assessments depending on the aesthetics of the site. A quality assessment is in essence the expectations of the user for the site. High aesthetics would mean, for the user, better quality and higher expectations and as such a positive "frame of mind" towards the site. A user in a positive frame of mind is probably more inclined to excuse problems unconsciously or consciously when filing the questionnaire, than a user that is in a negative frame of mind. That would be the tipping point: users would, if in a positive frame of mind, excuse some problems while in a negative one might be actively looking for imperfections.

This theory would fit with the rest of the results of the second experiment in which the users perceived ease of navigation, accuracy of the website were higher.

The first experiment, results suggest that users "*upon first visit*" are trying to figure out what they are looking at, and whether it is of any interest to them. Subsequently users that are interested in the domain of the website might keep the same attitude while others lose interest. Losing interest would mean that they fill the questionnaires because they have to, the motivation here being that were asked to do it. However they would not appear to be influenced by the site's aesthetics because they would have no

interest in the site.

Our first experiment high aesthetics site template had a just above-average score (see section 7.1.4.2 page 66). It was designed on purpose to avoid fashion effects; that means that it was not groundbreaking or intended to try and draw users in even if they were not that interested in the site. As such users that had no interest to the domain of the site, rated it more or less the same between high and low aesthetics and there was no statistically significant result, because they had no more interest in doing it any differently. In other words, it did not concern them. To give an analogy, it would be like asking someone who knows about sports but is not interested in cricket and is absolutely not interested in learning anything about it, to rate an important cricket match.

Furthermore we theorise that a longitudinal study, as an experimental design, should only be used with participants that are actually *interested*. Interest here means not only real interest in the domain but also forced interest. Thus in working environment, for example, workers are often utterly uninterested in the work but have to do it anyway. Since there is no literature on "subject interest in a website" and in "real life", not a lab environment, we expect that the experimental design for such further research would be very challenging.

The question that we were left with from the first experiment was that: aesthetics does play a role *upon first visit*, but what happens when users have used the website for some time? It would mean that they have become familiar with the use of that particular site. But then does aesthetics become insignificant only content playing a role? Or does it continue being a variable in perceived usability?

The results of the second experiment give more detail. Perceived usability is not only higher when the users first look at site before they use it. It remains higher after they used it briefly (we explained above what 'briefly' means in this case). The perception of this first visit remains even after use of the website. Users remember having perceived the high aesthetics site as easier to use than the low aesthetics one.

Norman theorised that design and aesthetics are perceived emotionally. Helpfulness

is an affective engagement, as is interest. Interest is furthermore a motivator. Further work could be done in researching emotions in relation to design. Helpfulness could be related to calmness and stress emotions in the user experience. A possible experimental design might try to frustrate the users before the experiment and then make them use a website. Possibly a 2×2 experiment: the groups would split into users of high aesthetics and low aesthetics, frustrated and non-frustrated. The idea behind such an experimental design is that it is easier to aggravate users than actually relax them or make them happy. Some harmless aggravation technique could be used for example: make them four or five times repeat a logging-in procedure with wild error messages, or make them wait outside the lab for a long time, make them fill long and purposeless paperwork before they start the experiment; harmless, nonetheless irritating. This would give us an insight as to how emotions play a role in aesthetic appreciation and its interplay with usability perception.

8.2   Synopsis

This document presented research in aesthetics as an attribute of usability. In order to conduct this research we presented a brief review of philosophy of aesthetics, applied aesthetics and usability; we concluded this part by an account of our position concerning aesthetics. Following that, we critically reviewed past research to draw lessons for our experimental design. We briefly described our own first approaches to designing an experiment on this issue. Following that we conducted two experiments. We presented for each the experimental design, presented the data, its statistical analysis and the results. Finally we presented the limitations of this research as well as further research that we propose needs to be done.

There were several positive results were usability was significantly influenced by higher aesthetics. For the first experiment: *easier to use upon first visit, GUS, helpfulness, tendency to find more easily things around the website, Organisation of information on the system screens is clear.* For the second experiment: *perceived usability before use,*

perceived usably after brief use, ease of navigation, perceived usability upon first visit after having used the site, the links of the website take the user were he expects. Neither experiment revealed any negative results in relation to our hypothesis. And the results of the second experiment support the results of the first one.

We made a novel contribution to understanding the relationship of usability and aesthetics. Our first experiment had positive results which the second validated and explored further. Lastly we proposed an encompassing theory to interpret these results.

Apendix

A.1 Between Experiment

A.1.1 WAMMI questionaire

Scale	Reliability	
	Raw	Adj.
Attractiveness	0.640	0.899
Controllability	0.690	0.918
Efficiency	0.630	0.895
Helpfulness	0.700	0.921
Learnability	0.740	0.934
Global	0.900	-

Table A.1: WAMMI reliability table, cited on WAMMY webpage[6]

1	This web site has much that is of interest to me.
2	It is difficult to move around this web site.
3	I can quickly find what I want on this web site.
4	This web site seems logical to me.
5	This web site needs more introductory explanations.
6	The pages on this web site are very attractive.
7	I feel in control when I'm using this web site.
8	This web site is too slow.
9	This web site helps me find what I am looking for.
10	Learning to find my way around this web site is a problem.
11	I don't like using this web site.
12	I can easily contact the people I want to on this web site.
13	I feel efficient when I'm using this web site.
14	It is difficult to tell if this web site has what I want.
15	Using this web site for the first time is easy.
16	This web site has some annoying features.
17	Remembering where I am on this web site is difficult.
18	Using this web site is a waste of time.
19	I get what I expect when I click on things on this web site.
20	Everything on this web site is easy to understand
extra question 1	How important for you is the kind of site you have just been rating?
extra question 2	How would you rate your internet skills and knowledge?
open answer question 1	What feature of this web site do you think should be improved, and why?
open answer question 1	What do you think is the best feature of this website, and why?

Table A.2: WAMMI questions

A.1.2 WAMMI questionnaire controlled population, data

A.1.2.1 Statistical treatment equations

- Shapiro - Wilk[1] test:

---

[1]Shapiro - Wilk tests the null hypothesis that values of $x_1, x_2, x_3, ..., x_n$ are a normally distributed sample [80].

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n} (x_i - \bar{x})} \tag{A.1}$$

- Mann-Whitney U is calculated as follows:

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2} \tag{A.2}$$

Where n_1 is the sample size for sample 1, and R_1 is the sum of the ranks in sample 1.

- Wilcoxon signed ranked statistic is calculated as follows:

$$W_+ = \sum_{i=1}^{n} \phi_i R_i \tag{A.3}$$

- Kruskal - Wallis one-way analysis of variance by ranks for more than two samples:

$$K = (N - 1) \frac{\sum_{i=1}^{g} n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^{g} \sum_{i=1}^{n_i} (r_{ij} - \bar{r})^2} \tag{A.4}$$

A.1.2.2 Group B WAMMI results

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	.183	10	.200*	.901	10	.226
Controllability	.143	10	.200*	.952	10	.688
Efficiency	.282	10	.023	.833	10	.036
Helpfulness	.126	10	.200*	.946	10	.625
Learnability	.156	10	.200*	.957	10	.757
Global	.180	10	.200*	.889	10	.165
a. Lilliefors Significance Correction						
*. This is a lower bound of the true significance.						
b. .00=aesthetics low 1.00=aesthetics high = 1						

Table A.3: Group B low aesthetics normality test

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	.243	13	.034	.883	13	.078
Controllability	.140	13	.200*	.958	13	.719
Efficiency	.159	13	.200*	.902	13	.142
Helpfulness	.137	13	.200*	.940	13	.462
Learnability	.113	13	.200*	.950	13	.594
Global	.132	13	.200*	.967	13	.855
a. Lilliefors Significance Correction						
*. This is a lower bound of the true significance.						
b. .00=aesthetics low 1.00=aesthetics high = 1						

Table A.4: Group B high aesthetics normality test

Type	Attractiveness	Controllability	Efficiency	Helpufullness	Learnability	Global
Mean	26.3	37.7	42.6	31.7	45.9	36.5
Median	28	35	30	32.5	45	37
SSQ	1492.1	5056.1	4490.4	2446.1	4284.9	2360.5

Table A.5: Group B low aesthetics mean scores

Type	Attractiveness	Controllability	Efficiency	Helpufullness	Learnability	Global
Mean	38.23	43.46	32.08	42.54	51.54	41.08
Median	30.00	49.00	31.00	38.00	57.00	42.00
SSQ	9012.31	7209.23	6246.92	9583.23	10121.23	6110.92

Table A.6: Group B high aesthetics mean scores

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	.183	10	.200*	.901	10	.226
Controllability	.143	10	.200*	.952	10	.688
Efficiency	.282	10	.023	.833	10	.036
Helpfulness	.126	10	.200*	.946	10	.625
Learnability	.156	10	.200*	.957	10	.757
Global	.180	10	.200*	.889	10	.165

Table A.7: Group B Non-parametric Statistics, Man - Whitney U

A.1.3 Tables and data of open population questionnaires

A.1.3.1 First questionnaire open population

0='Low ae.,' 1='High ae.'		Qu. 1	Qu. 2	Qu. 3	Qu. 4	Qu. 5	Qu. 6	Qu. 7
0	Mean	3.64	3.49	2.31	2.33	3.40	3.82	3.49
	Median	4.00	4.00	2.00	2.00	4.00	4.00	4.00
	Std. Deviation	1.176	1.169	1.386	1.263	1.285	1.140	1.275
	N	55	55	55	55	55	55	55
1	Mean	3.82	3.87	2.38	4.02	4.11	3.82	3.82
	Median	4.00	4.00	2.00	4.00	4.00	4.00	4.00
	Std. Deviation	1.056	1.106	1.097	.913	1.031	1.203	1.140
	N	55	55	55	55	55	55	55
Total	Mean	3.73	3.68	2.35	3.17	3.75	3.82	3.65
	Median	4.00	4.00	2.00	3.00	4.00	4.00	4.00
	Std. Deviation	1.116	1.149	1.245	1.387	1.213	1.167	1.215
	N	110	110	110	110	110	110	110

Means of first questionnaire open population, 0= Low aesthetics, 1= High aesthetics

Table A.8: First questionnaire, open population, general statistics

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7
Mann-Whitney U	1383.500	1219.000	1372.000	478.000	1020.000	1494.000	1293.500
Wilcoxon W	2923.500	2759.000	2912.000	2018.000	2560.000	3034.000	2833.500
Z	-.801	-1.825	-.871	-6.347	-3.064	-.115	-1.361
Asymp. Sig. (2-tailed)	.423	.068	.384	.000	.002	.908	.174

Grouping Variable: 0='Low aesthetics', 1='High aesthetics'

Table A.9: First questionnaire open population,Test Statistics

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7
Mann-Whitney U	1409.500	1424.000	1458.000	1367.500	1493.000	1358.500	1432.000
Wilcoxon W	2634.500	3315.000	3349.000	2592.500	2718.000	3249.500	3323.000
Z	-.531	-.441	-.228	-.784	-.009	-.853	-.391
Asymp. Sig. (2-tailed)	.596	.659	.820	.433	.993	.394	.696

Grouping variable is Gender: 1 = Male 2 = Female

Table A.10: Open population, test statistics, gender as group variable

1='Male', 2='Female'		Qu. 1	Qu. 2	Qu. 3	Qu. 4	Qu. 5	Qu. 6	Qu. 7
1	Mean	3.67	3.69	2.37	3.06	3.76	3.92	3.73
	N	49	49	49	49	49	49	49
	Std. Deviation	1.088	1.245	1.236	1.360	1.217	1.152	1.114
2	Mean	3.77	3.67	2.33	3.26	3.75	3.74	3.59
	N	61	61	61	61	61	61	61
	Std. Deviation	1.146	1.076	1.261	1.413	1.220	1.182	1.296
Total	Mean	3.73	3.68	2.35	3.17	3.75	3.82	3.65
	N	110	110	110	110	110	110	110
	Std. Deviation	1.116	1.149	1.245	1.387	1.213	1.167	1.215

Means depending to Gender: 1= Male, 2 = Female

Table A.11: Open population, general statistics depending on gender

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7
Kruskal Wal. χ^2	3.810	1.837	3.415	1.750	6.277	1.903	1.101
df	2	2	2	2	2	2	2
Asymp. Sig.	.149	.399	.181	.417	.043	.386	.577

Grouping Variable Age: 1='15-24', 2='24-38', 3='39-...'

Table A.12: Kruskal-Wallis test

1='15-24', 2='24-38', 3='39-...'		Qu. 1	Qu. 2	Qu. 3	Qu. 4	Qu. 5	Qu. 6	Qu. 7
1	Mean	3.84	3.89	2.26	2.84	3.58	4.16	3.74
	Median	4.00	4.00	2.00	3.00	4.00	4.00	4.00
	Std. Deviation	.958	1.100	1.195	1.119	1.261	1.015	.872
	N	19	19	19	19	19	19	19
2	Mean	3.59	3.56	2.23	3.21	3.64	3.77	3.71
	Median	4.00	4.00	2.00	3.00	4.00	4.00	4.00
	Std. Deviation	1.177	1.225	1.185	1.462	1.240	1.173	1.230
	N	73	73	73	73	73	73	73
3	Mean	4.17	3.94	2.89	3.39	4.39	3.67	3.33
	Median	4.00	4.00	3.00	3.00	5.00	4.00	3.00
	Std. Deviation	.924	.802	1.451	1.335	.850	1.283	1.455
	N	18	18	18	18	18	18	18
Total	Mean	3.73	3.68	2.35	3.17	3.75	3.82	3.65
	Median	4.00	4.00	2.00	3.00	4.00	4.00	4.00
	Std. Deviation	1.116	1.149	1.245	1.387	1.213	1.167	1.215
	N	110	110	110	110	110	110	110

Table A.13: Open population, Means depending to Age

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7
Kruskal Wal. χ^2	1.275	2.154	.486	6.683	1.368	2.589	.436
df	2	2	2	2	2	2	2
Asymp. Sig.	.529	.341	.784	.035	.505	.274	.804

a. Kruskal Wallis Test b. Grouping Variable: 1='none' 2='occasional pages' 3='lots of pages'

Table A.14: Kruskal Wallis Test, Grouping Variable "Web Experience"

		Qu. 1	Qu. 2	Qu. 3	Qu. 4	Qu. 5	Qu. 6	Qu. 7
1	Mean	3.75	3.60	2.32	3.44	3.76	3.76	3.69
	Median	4.00	4.00	2.00	3.00	4.00	4.00	4.00
	Std. Deviation	1.084	1.211	1.263	1.274	1.173	1.186	1.162
	N	68	68	68	68	68	68	68
2	Mean	3.59	3.71	2.29	2.82	3.62	3.76	3.56
	Median	4.00	4.00	2.00	3.00	4.00	4.00	4.00
	Std. Deviation	1.209	1.060	1.115	1.527	1.349	1.208	1.260
	N	34	34	34	34	34	34	34
3	Mean	4.13	4.25	2.75	2.38	4.25	4.50	3.75
	Median	4.50	4.50	3.00	3.00	4.50	4.50	4.50
	Std. Deviation	.991	.886	1.669	1.188	.886	.535	1.581
	N	8	8	8	8	8	8	8
Total	Mean	3.73	3.68	2.35	3.17	3.75	3.82	3.65
	Median	4.00	4.00	2.00	3.00	4.00	4.00	4.00
	Std. Deviation	1.116	1.149	1.245	1.387	1.213	1.167	1.215
	N	110	110	110	110	110	110	110

1='none' 2='occasional pages' 3='lots of pages'

Table A.15: Means according to web experience

A.1.3.2 Select cases according to experience

		Qu. 1	Qu. 2	Qu. 3	Qu. 4	Qu. 5	Qu. 6	Qu. 7
0	Mean	3.76	3.31	2.38	2.72	3.38	3.76	3.76
	N	29	29	29	29	29	29	29
	Std. Deviation	1.091	1.198	1.498	1.306	1.178	1.154	1.091
1	Mean	3.74	3.82	2.28	3.97	4.05	3.77	3.64
	N	39	39	39	39	39	39	39
	Std. Deviation	1.093	1.189	1.075	.959	1.099	1.224	1.224
Total	Mean	3.75	3.60	2.32	3.44	3.76	3.76	3.69
	N	68	68	68	68	68	68	68
	Std. Deviation	1.084	1.211	1.263	1.274	1.173	1.186	1.162

0 = Low aesthetics and 1 = High aesthetics

Table A.16: Means of users that have no web development experience

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7
Mann-Whitney U	561.000	422.000	549.500	264.000	378.000	557.500	547.000
Wilcoxon W	996.000	857.000	984.500	699.000	813.000	992.500	1327.000
Z	-.058	-1.839	-.206	-3.883	-2.419	-.103	-.238
Asymp. Sig. (2-tailed)	.954	.066	.837	.000	.016	.918	.812

0 = Low aesthetics and 1 = High aesthetics

Table A.17: Mann Whitney significance tests on the results of users that have no web development experience

			Qu. 1	Qu. 2	Qu. 3	Qu. 4	Qu. 5	Qu. 6	Qu. 7
0		Mean	3.50	3.69	2.23	1.88	3.42	3.88	3.19
		N	26	26	26	26	26	26	26
		Std. Deviation	1.273	1.123	1.275	1.071	1.419	1.143	1.415
1		Mean	4.00	4.00	2.63	4.13	4.25	3.94	4.25
		N	16	16	16	16	16	16	16
		Std. Deviation	.966	.894	1.147	.806	.856	1.181	.775
Total		Mean	3.69	3.81	2.38	2.74	3.74	3.90	3.60
		N	42	42	42	42	42	42	42
		Std. Deviation	1.179	1.042	1.229	1.466	1.289	1.144	1.308

0 = Low aesthetics and 1 = High aesthetics

Table A.18: Means of users that have web development experience

Mann-Whitney U	163.500	179.000	157.500	29.000	136.500	199.000	119.500
Wilcoxon W	514.500	530.000	508.500	380.000	487.500	550.000	470.500
Z	-1.194	-.795	-1.359	-4.761	-1.951	-.245	-2.419
Asymp. Sig. (2-tailed)	.233	.427	.174	.000	.051	.806	.016

a. Grouping Variable: 0='Low aesthetics', 1='High aesthetics'

Table A.19: Mann Whitney significance tests on the results of users that have no web development experience

A.1.3.3   WAMMI questionnaire result tables

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	.203	48	.000	.903	48	.001
Controllability	.102	48	.200*	.931	48	.008
Efficiency	.108	48	.200*	.960	48	.105
Helpfulness	.107	48	.200*	.943	48	.022
Learnability	.078	48	.200*	.960	48	.098
Global	.082	48	.200*	.969	48	.242

a. Lilliefors Significance Correction

* This is a lower bound of the true significance

Table A.20: Normality test on low aesthetics group

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	.128	51	.036	.933	51	.006
Controllability	.107	51	.200*	.953	51	.043
Efficiency	.139	51	.015	.968	51	.179
Helpfulness	.143	51	.011	.921	51	.002
Learnability	.097	51	.200*	.940	51	.013
Global	.084	51	.200*	.962	51	.102

a. Lilliefors Significance Correction

* This is a lower bound of the true significance

Table A.21: Normality test on high aesthetics group

	Attractiveness	Controllability	Efficiency	Helpfulness	Learnability	Global
Mann-Whitney U	776.500	1116.000	1088.500	1018.000	1076.500	966.000
Wilcoxon W	1952.500	2292.000	2264.500	2194.000	2252.500	2142.000
Z	-3.134	-.756	-.950	-1.443	-1.033	-1.807
Asymp. Sig. (2-tailed)	.002	.449	.342	.149	.302	.071

Table A.22: Mann - Whitney test for WAMMI questionnaire

0= Low aesthetics, 1= High aesthetics		Attractiveness	Controllability	Efficiency	Helpfulness	Learnability	Global
0	Mean	31.19	40.40	45.75	38.42	42.42	39.17
	N	48	48	48	48	48	48
	Std. Deviation	25.541	26.635	25.860	24.816	24.758	20.494
1	Mean	50.25	44.57	51.67	47.27	48.29	48.06
	N	51	51	51	51	51	51
	Std. Deviation	29.149	27.232	23.390	28.867	27.517	23.385
Total	Mean	41.01	42.55	48.80	42.98	45.44	43.75
	N	99	99	99	99	99	99
	Std. Deviation	28.950	26.888	24.672	27.209	26.249	22.370

Table A.23: Mean results of WAMMI questionnaire

| | | Levene's Test for equal var. | | T-test for equality of means | | | | | 95% Conf. Inter. of the Diff. | |
		F	Sig.	t	df	Sig. (2-tailed)	Mean Dif.	Std. Error Dif.	Lower	Upper
Attractiveness	Equal var. assumed	1.016	.316	-3.453	97	.001	-19.067	5.522	-30.027	-8.107
	Equal var. not assumed			-3.467	96.519	.001	-19.067	5.500	-29.984	-8.151
Controllability	Equal var. assumed	.000	.989	-.770	97	.443	-4.173	5.419	-14.927	6.581
	Equal var. not assumed			-.771	96.852	.443	-4.173	5.415	-14.920	6.574
Efficiency	Equal var. assumed	1.210	.274	-1.195	97	.235	-5.917	4.951	-15.742	3.909
	Equal var. not assumed			-1.191	94.550	.236	-5.917	4.966	-15.776	3.942
Helpfulness	Equal var. assumed	1.448	.232	-1.633	97	.106	-8.858	5.426	-19.626	1.911
	Equal var. not assumed			-1.640	96.230	.104	-8.858	5.401	-19.578	1.862
Learnability	Equal var. assumed	1.010	.317	-1.115	97	.268	-5.877	5.272	-16.341	4.586
	Equal var. not assumed			-1.118	96.810	.266	-5.877	5.255	-16.308	4.553
Global	Equal var. assumed	1.609	.208	-2.007	97	.048	-8.892	4.431	-17.686	-.099
	Equal var. not assumed			-2.015	96.522	.047	-8.892	4.413	-17.651	-.133

Table A.24: Student T-tests on the percentile results of WAMMI questionnaire

A.1.3.4   WAMMI filtered

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	.124	21	.200*	.958	21	.474
Controllability	.132	21	.200*	.924	21	.106
Efficiency	.090	21	.200*	.982	21	.950
Helpfulness	.155	21	.200*	.933	21	.157
Learnability	.089	21	.200*	.972	21	.777
Global	.133	21	.200*	.947	21	.295

a. Lilliefors Significance Correction

* This is a lower bound of the true significance.

b. 0= Low aesthetics, 1= High aesthetics = 0

Table A.25: Normality tests, low aesthetics using users that were interested in the genre of site.

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	.189	25	.021	.869	25	.004
Controllability	.169	25	.063	.911	25	.032
Efficiency	.160	25	.099	.904	25	.022
Helpfulness	.139	25	.200*	.895	25	.014
Learnability	.201	25	.010	.886	25	.009
Global	.138	25	.200*	.925	25	.067

a. Lilliefors Significance Correction

* This is a lower bound of the true significance.

b. 0= Low aesthetics, 1= High aesthetics = 1

Table A.26: Normality tests, high aesthetics using users that were interested in the genre of site.

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Attractiveness	114.500	345.500	-3.266	.001
Controllability	186.000	417.000	-1.688	.091
Efficiency	258.000	489.000	-.099	.921
Helpfulness	151.500	382.500	-2.451	.014
Learnability	217.000	448.000	-1.005	.315
Global	163.500	394.500	-2.184	.029

a. Grouping Variable: 0= Low aesthetics, 1= High aesthetics

Table A.27: Mann-Whitney tests for WAMMI questionnaire on selected cases

0= Low Aesthetics, 1= High Aesthetics		Attractiveness	Controllability	Efficiency	Helpfulness	Learnability	Global
0	Mean	42.19	43.62	58.10	41.24	48.62	46.29
	N	21	21	21	21	21	21
	Std. Deviation	24.935	25.590	21.009	25.237	22.703	19.111
1	Mean	69.16	57.04	59.72	61.28	56.96	60.44
	N	25	25	25	25	25	25
	Std. Deviation	23.787	28.734	23.681	26.837	28.035	22.591
Total	Mean	56.85	50.91	58.98	52.13	53.15	53.98
	N	46	46	46	46	46	46
	Std. Deviation	27.615	27.876	22.270	27.732	25.805	22.030

Table A.28: Means of selected cases of WAMMI questionnaire

	Attractiveness	Controllability	Efficiency	Helpfulness	Learnability	Global
Mann-Whitney U	258.000	319.500	273.500	324.000	326.000	317.500
Wilcoxon W	636.000	670.500	651.500	675.000	704.000	695.500
Z	-1.656	-.561	-1.382	-.481	-.445	-.596
Asymp. Sig. (2-tailed)	.098	.575	.167	.631	.656	.551

Table A.29: Non-parametric statistics of selected cases of WAMMI questionnaire where users where not interested in the domain of the site.

A.1.4 Soft results

A.1.4.1 Groub B soft results

User No.	Is there anything you think is missing from this web site?
1	Not really
2	no
3	colors and a better construction
4	A few more categories could make it more appealing. It needs a bit of design.
5	n/a
6	nothing
7	No
8	Nothing
9	Attractive backgrounds and images to make it look appealing
10	No

Table A.30: Group B, low aesthetics template answers to open question: *Is there anything you think is missing from this web site?*

User No.	Is there anything you think is missing from this web site?
1	A point.
2	unkown
3	a
4	No
5	information about what is happening in the industry of design
6	no
7	No
8	No.
9	forum
10	no
11	don't know, you know your target
12	I think their should be a contact button so it is easy to get in touch for the webmaster. I also noticed there is no copyright symbol, or policy on the site which is an essential when creating a website.I feel the banner could do with some work, the text isn't in the center, it's more in the upper center. I think the text should stand out more and not have the .net" in the name. Where the background gradient goes from grey to a sort of pink you can see near the top on the banner a grey edge that looks out of place.
13	A comments section for blog posts

Table A.31: Group B, low aesthetics template answers to open question: *Is there anything you think is missing from this web site?*

A.1.4.2   WAMMI soft results

Low aesthetics template	High aesthetics template
it is difficult to move around this web site.	it is difficult to move around this web site.
it is difficult to tell if this web site has what they want.	it is difficult to tell if this web site has what they want.
learning to find their way around this web site is a problem.	learning to find their way around this web site is a problem.
remembering where they are on this web site is difficult.	remembering where they are on this web site is difficult.
the pages on this web site aren't very attractive.	the pages on this web site aren't very attractive.
they can't easily contact the people they want to on this web site.	they can't easily contact the people they want to on this web site.
they don't feel efficient when they're using this web site.	they don't feel efficient when they're using this web site.
they don't feel in control when they're using this web site.	they don't feel in control when they're using this web site.
they don't like using this web site.	they don't like using this web site.
this web site does not help them find what they are looking for.	this web site does not help them find what they are looking for.
this web site does not seem logical to them.	this web site does not seem logical to them.
this web site needs more introductory explanations.	this web site needs more introductory explanations.
this website does not have much that is of interest to them.	this website does not have much that is of interest to them.
using this web site is a waste of time.	using this web site is a waste of time.
not everything on this web site is easy to understand.	
they don't get what they expect when they click on some of the things on this web site.	
this web site has some annoying features.	

Table A.32: Main weaknesses as stated by the users. The statements are in the form of completing the phrase: *Visitors think that ...*

ID	Comment
1	Design, looks bland
2	esthetics - basic background and fonts
3	No contact link (unless i wasnt able to find any)
4	look and fonts
5	the categories should be better partitioned, the titles of the sections better chosen, and you shouldn't have to scroll down the page to get all the info. Plus, if I managed to find the info in the querry it was entirely thanks to the search machine for it would have taken ages otherwise
6	It's a little bare
7	Page design, to make it more attractive
8	The styling is very bad. Wrong background color and font.
9	needs to be more attractive
10	the graphics and the description of the featured items - it needs to be designed by a professional. there is a reason why people study graphics and web design...
11	speed
12	navigation could be improved with the use of colours
13	appearance and navigation to make it easier for the user
14	Grpahics and design. The website has no character.
15	different space on the page with the list of the content and the content itself,a space where it shows the path you clicked to get where you are
16	the appearance should be more attractive cause its the first impression and it is about art,which means for me something different from what i see everyday
17	Creativity
18	presentation for it to be more atractiv
19	Fonts-too classic, reminds me of google's this-page-cannot-be-found pages.../Different information should be more easy to distinguish
20	All, I didn't like it
21	style of pages
22	The design.So that we can see more easily the sections (by colour)
23	The graphics are too simple
24	The address is not relevant to the content and the presentation of the content, which is mainly artistic, lacks in design. Would expect bigger photos, better use of colours, fonts and editing from one subject to the next within a page.
25	compatibility with diferent operating systems such as linux
26	Is is very slowly.

27	look and feel because it feels unprefessional
28	New-Design , Maybe a coverflow effect..
29	the position of outter links and where they lead. i dont like links the pop me out of the main site
30	Videos loads to slowly
31	the appearance and the structure of the site
32	Eliminate 'irrelevant' things from main menu; give it a cleaner look; don't clog front page with posts - or at least make posts more 'clearly' separate from eachother
33	It's not very colourful :)
34	Apearance
35	colors/graphics/attractiveness
36	Design (Layout and Graphics), to make navigation more intuitive and website more attractive
37	Layout, slightly. It's a bit too uniform.
38	-
39	graphics, appearance, organisation/structure of content
40	Design - Q17
41	the organisation
42	Appearance and design. Content.
43	The organization and the overall appearance.
44	Interface and content info should get richer.
45	graphics
46	I found the site not user friendly (it's difficult to understand what I'm looking at, I don't know what to expact before watching a vidio)
47	The main menu. the bottom part is confusing as they don't seem to be categories. Maybe have a better font and some graphics in terms of web design-page presentation. + some colors!
48	The presentation because it could be more attractive and easy to explore.

Table A.33: Low aesthetics template answers on open text question: "What feature of this web site do you think should be improved, and why?"

ID	Comment
1	It's just a blog with some videos. If you want to make a cool video site out of it, you have to show your stuff (links/thumbnails) right on the home page. Don't expect your users they will click on every single link to find you what you have here for them. There are thousands of sites like this one, to catch the user for longer time, you have to really stand out. Ask a good web developer where to start, and focus on clarity and usability :) (and PS. most annoying part for ma was the questionnaire, that is, its organization, not the real questions...)
2	Explanation of the purpose for which the site was created
3	More introduction for whom this website is.
4	Content, Visual Appearance, Information Architecture
5	Left menu has some links that for no apparent reason pop new windows when clicked. Also links have names that make no sence (the last ones) *
6	-
7	syntax, spelling, use of commas.
8	as a blog everything looks well presentable
9	info
10	Promotion of content in a variety of ways (video/image slidebars, random picks, etc.)
11	more films
12	Maybe appearance.. it's fine, but it it could be more 'artistic', more original.
13	?
14	nothing
15	I think that the site doesn't need any improvement-except for the enrichment of art material presented
16	navigation, it's not always obvious to which category a film belongs
17	-
18	introductory explanations
19	more images, less (introductory) text. that would speed up the process of finding information.
20	Nothing
21	I think it's all good.
22	should be made more clear, more easy to navigate in it.
23	should be made more clear, more easy to navigate in it.
24	very slow

25	a better overview of the included content should be given. it still is blur to me about what this website wants to represent/show me. the main menu points aren?t ment for people with a 'regular' based knowledge or interest. it?s far out for those who don?t belong to this topic/field.
26	navigation
27	A LITTLE BETTER PRESENTATION
28	some windows opened in a new tab in Opera which confused me
29	Home page content needs to identify organization better
30	Syntax, spelling, use of commas. It would make the site easier to understand, plus these kinds of errors make the site look sloppy.
31	structure of knowledge
32	link section. Mention if a link takes you to another site
33	menu on the left
34	little more colour maybe
35	The content of the page should be more consistent and some links are off use
36	Navigation of the blog should be improved. Its seems like its been put together by some crazy puzzle maker..
37	design could be more tempting...
38	Searching
39	-
40	It depends to whom is adressed to. I can not answer that
41	It depends to whom is adressed to. I can not answer that
42	A presentation of what real it is and what we have to expect
43	I think it sould be easier to navigate
44	Some movies weren't running very well - bacause it is annoying. The navigation should be easier as well.
45	introductory explanations
46	contact data, how can I contact you? scrool bar, too many info per category
47	better organization of material
48	Background color.
49	i can't answer on this.
50	appearance
51	no improvements are needed

Table A.34: High aesthetics template answers on open text question: "What feature of this web site do you think should be improved, and why?"
* Translated from Greek

ID	Comment
1	Content
2	content - brilliant! Well selected, well exposed, interesting and beautiful
3	Search option presented exactly what i was expecting.
4	search option
5	the content. I like it!
6	Very fast to navigate
7	Simple, easy to follow structure (you find your way easily)
8	It s easy to navigate around.
9	the videos
10	it seems to contain a great collection of art material which could allow it to act as a point of reference for someone interested in arts
11	the subject
12	the content
13	the content - its something different and not common
14	Nice categorized menu. Is pretty useful and clear structured.
15	solid organised
16	loved the radiohead-nude video with the instruments
17	animation
18	different the content it s a friend s site
19	Content. Very interesting.
20	None
21	interesting content
22	The videos are very interesting.
23	The videos.Really great job
24	Inspires confidence to navigate through its pages, as the different categories are clearly stated, not too many and easy to spot.
25	content–interesting
26	Nice videos choice
27	the speed because it's efficient
28	The content :)
29	the best feature is its simplicity. it is easy to use and you know where you are while visiting the site
30	fotography and usufull information about this subject

31	its content is the best feature of it. Because the things it shows are interesting catch your attention and are friendly to the eye. I liked to see new and interesting videos.
32	It has some nice stuff - as a website, I don't think it has any particular featuer I'd like to single out.
33	It's simple and straightforward therefore navigation is easy
34	none
35	personal statements/ videos/ content
36	Couple of interesting pieces of content.
37	Content, excellent choices.
38	-
39	the content is too good for this website!
40	The content, because I like it, but the way how it is presented no.
41	the films , it was really good choices
42	My Room in a Box - The Casulo (VIDEO) , although i can't say i am that impressed by this video, it was nice to watch it.
43	The contents are very interesting.
44	Slick search features although I think it lacks any web 2.0 features such as content tagging and intuitive categorization.
45	its content
46	it was sugested to find vidios, and the search gave only one result each time (there were no need to choose and wonder whether it is the one ;))
47	the video choices are all very interesting!
48	The short films and videos because they are very interesting and of great importance

Table A.35: Low aesthetics template answers to the open text question: "What do you think is the best feature of this website, and why?"

ID	Comment
1	Don't know...
2	search engine, it helps a lot if I know what I'm looking for
3	The movies because they are very nice.
4	Search engine - helped me accomplish all tasks
5	You can see exeptionaly good works of art all gathered together. The selection of movies is amazing!! *
6	photos and videos
7	the layout, easy to navigate
8	Animation' section in 'Films
9	pictures
10	Specific and consistent content
11	the design is really good
12	In my opinion it's the content of the blog.
13	?
14	the videos
15	Very interesting and creative content. It can easily attract attention and is very inspiring
16	the content- I find the films, animation, design realy interesting
17	the site's layout is logical (I think i don't have to explain why being logical is good ;])
18	videos
19	the crow animation. i still was patient enough in the beginning then lost tolerance of completing questinnnaire.
20	Animations
21	I like the most the films, especially animations.
22	very interesting animations
23	very interesting animations
24	very good navigation (knowing where you are anytime)
25	sharp pics and short films. the pics are nicely shot and sharp even when they are intentionaly unsharp...the short films are interesting due to their 'lost highway' touch. at least some of them.the idea of gathering these ideas and videos is very posiitive the execution could use more effort.
26	design part
27	THE IT SHOWS THE VIDEOS
28	the content

29	Color and design I just like it
30	The layout, easy to navigate.
31	alternative content
32	Search function
33	videos because they are interesting
34	logical structure
35	The animations are really interesting
36	some of the videos are really cool. Did not know about these shorts.
37	it's really easy to find information on this web site that I want to find
38	the way that the information is organized
39	photos and videos
40	I liked the animation that i wached and 'screensaver'. But i don't think that i am the right person to answer or even more use such a site. Only for pleasure
41	I liked the animation that i wached and 'screensaver'. But i don't think that i am the right person to answer or even more use such a site. Only for pleasure
42	The possibility to find articles for design and live examples like films or photos
43	The videos. They are very pleasant.
44	contents: nearly every film is good and worth seeing.
45	videos
46	videos, too good!!
47	the 'room in a box' video
48	Main menu, easy to access.
49	the simplicity that it has to find what you want.
50	video content
51	pleasant aesthetic

* Translated from greek

Table A.36: High aesthetics template answers to the open text question: "What do you think is the best feature of this website, and why?"

A.2 Second experiment

A.2.1 Questionnaires

A.2.1.1 Questionnaire site A

Questions that were collecting data and where part of the procedural questions meant to guide the users in the site. have a title title in bold letters. The title is the one used

for the specific question in subsequent charts and tables.

Easy A1 This is site A. : How easy do you think this site would be to use? (1 being lowest value and 7 the highest)

 Find the page with the hotels. How many four star hotels does the Small Village have?

Easy A2 How easy do you think this site is to use after having visited it? (1 being lowest value and 7 the highest)

 Please find the mountain page in the site seeing page. What does the mountain offer?

 How many different types of beaches are there in the Small Village?

 Please read the text on the rocky beach and choose the answer that best qualifies it.

Aesthetics A Please rate the visual attractiveness of site A

Navigation A Please rate how easy it is to navigate through website A

Find A Please rate how easy it is to find what you want on website A

Easy A Please rate how easy you thought it was to use site A when first opening it?

Accurate A Please rate how accurate the phrase is for website A: Clicking on links takes me to what I expect

Organisation A Please rate how clear the organisation of information on website A is

A.2.1.2   Questionnaire site B

Easy B1 This is site B. : How easy do you think this site would be to use?

 If you wanted to try French cuisine to which establishment would you go?

Easy B2 How easy do you think this site is to use after having visited it?

 If you wanted to eat a sandwich to which take away would you go?

 Where in the Forest can a traveler enjoy a picnic?

 What kind of a beach is Pebbles beach?

Aesthetics B Please rate the visual attractiveness of site B

Navigation B Please rate how easy it is to navigate through website B

Find B Please rate how easy it is to find what you want on website B

Easy B Please rate how easy you thought it was to use site B when first opening it?

Accurate B Please rate how accurate the phrase is for website B: Clicking on links takes me to what I expect

Organisation B Please rate how clear is the organisation of information on website B

A.2.2   Personal data collected

At the end of the questionnaire some personal data where collected to measure how different groups react. The personal data collected were:

1. Genger

    (a) Male

    (b) Female

2. Age category

    (a) 18-24

    (b) 25-38

    (c) 39-58

    (d) 59-...

3. Web Design experience

    (a) None

    (b) Some

    (c) A lot

4. How interesting did you find the website?

    (a) Not at all

    (b) Somewhat

    (c) Very

5. Their Amazon Workers ID

Answer a short survey

I am conducting an academic survey about usability for the final stage of my Ph.D. Please visit the link bellow, and answer a few questions about the sites you are presented with. Please it is imperative that if you have answered any one of our surveys in past do not fill in this one, as it would be harmfull to the quality of the data that is gathered.

Please answer as truthfully as possible and please do not answer randomly.

Survey link: http://www.arktouros.net/_exp5/vlg05ba/raf1.html

Provide your worker ID also here:

Provide your worker ID in the questionaire box and here to have the HIT validated. This is to make sure that you have not done the experiment in a previous HIT.

Please do not repeat the experiment as you will not be payed.

Thank you, your help is apreciated!

Submit

Figure A.1: Screen capture of the the ad that was used to find users to take part in the experiment.

A.2.3 Statistic tables experiment 2

Group AB (1) and BA (2)								
Case Processing Summary								
Group		Cases						
		Valid		Missing		Total		
		N	Percent	N	Percent	N	Percent	
easyA1	1	52	100.0%	0	0.0%	52	100.0%	
	2	60	100.0%	0	0.0%	60	100.0%	
easyA2	1	52	100.0%	0	0.0%	52	100.0%	
	2	60	100.0%	0	0.0%	60	100.0%	
easyB1	1	52	100.0%	0	0.0%	52	100.0%	
	2	60	100.0%	0	0.0%	60	100.0%	
easyB2	1	52	100.0%	0	0.0%	52	100.0%	
	2	60	100.0%	0	0.0%	60	100.0%	
aestheticsA	1	52	100.0%	0	0.0%	52	100.0%	
	2	60	100.0%	0	0.0%	60	100.0%	
aestheticsB	1	52	100.0%	0	0.0%	52	100.0%	
	2	60	100.0%	0	0.0%	60	100.0%	
navigationA	1	52	100.0%	0	0.0%	52	100.0%	

	2	60	100.0%	0	0.0%	60	100.0%
navigationB	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
findA	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
findB	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
easyA	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
easyB	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
accurateA	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
accurateB	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
organisationA	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
organisationB	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
wrong1	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%
wrong2	1	52	100.0%	0	0.0%	52	100.0%
	2	60	100.0%	0	0.0%	60	100.0%

Table A.37: Case processing summary for within experiment

Tests of Normality			Kolmogorov-Sm.			Shapiro-Wilk		
gender								
			Statistic	df	Sig.	Statistic	df	Sig.
easyA1	1		.282	68	.000	.858	68	.000
	2		.249	44	.000	.859	44	.000
easyA2	1		.259	68	.000	.866	68	.000
	2		.253	44	.000	.844	44	.000
easyB1	1		.225	68	.000	.888	68	.000

	2	.261	44	.000	.874	44	.000
easyB2	1	.262	68	.000	.875	68	.000
	2	.285	44	.000	.861	44	.000
aestheticsA	1	.224	68	.000	.894	68	.000
	2	.270	44	.000	.796	44	.000
aestheticsB	1	.173	68	.000	.918	68	.000
	2	.178	44	.001	.916	44	.003
navigationA	1	.225	68	.000	.881	68	.000
	2	.241	44	.000	.834	44	.000
navigationB	1	.258	68	.000	.882	68	.000
	2	.258	44	.000	.812	44	.000
findA	1	.316	68	.000	.827	68	.000
	2	.246	44	.000	.848	44	.000
findB	1	.239	68	.000	.875	68	.000
	2	.264	44	.000	.771	44	.000
easyA	1	.247	68	.000	.878	68	.000
	2	.268	44	.000	.821	44	.000
easyB	1	.237	68	.000	.897	68	.000
	2	.217	44	.000	.876	44	.000
accurateA	1	.249	68	.000	.777	68	.000
	2	.258	44	.000	.750	44	.000
accurateB	1	.232	68	.000	.825	68	.000
	2	.216	44	.000	.767	44	.000
organisationA	1	.279	68	.000	.805	68	.000
	2	.243	44	.000	.815	44	.000
organisationB	1	.217	68	.000	.873	68	.000
	2	.249	44	.000	.821	44	.000
wrong1	1	.384	68	.000	.655	68	.000
	2	.301	44	.000	.729	44	.000
wrong2	1	.336	68	.000	.682	68	.000
	2	.333	44	.000	.652	44	.000

a. Lilliefors Sign. Correction

Table A.38: Normality tests for within experiment

Table: Descriptives				

			Stat	Std. Err.
aestheticsA	Mean		5.42	0.14
	95%Confidence interval for Mean	Lower bound	5.15	
		Upper bound	5.69	
	5% trimmed mean		5.53	
	Median		6	
	Variance		2.12	
	Std. Deviation		1.46	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		1.75	
	Skewness		-0.94	0.23
	Kurtosis		0.24	0.45
aestheticsB	Mean		4.36	0.17
	95% Confidence interval for Mean	Lower bound	4.02	
		Upper bound	4.69	
	5% trimmed mean		4.4	
	Median		5	
	Variance		3.24	
	Std. Deviation		1.8	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		3	
	Skewness		-0.27	0.23
	Kurtosis		-1.12	0.45
easyA1	Mean		5.68	0.1
	95% Confidence interval for Mean	Lower bound	5.48	
		Upper bound	5.88	
	5% trimmed mean		5.75	
	Median		6	
	Variance		1.16	
	Std. Deviation		1.08	
	Minimum		2	
	Maximum		7	
	Range		5	
	Interquartile range		1	

	Skewness		-0.96	0.23
	Kurtosis		1.18	0.45
easyA2	Mean		5.79	0.1
	95% Confidence interval for Mean	Lower bound	5.6	
		Upper bound	5.99	
	5% trimmed mean		5.86	
	Median		6	
	Variance		1.08	
	Std. Deviation		1.04	
	Minimum		2	
	Maximum		7	
	Range		5	
	Interquartile range		2	
	Skewness		-0.85	0.23
	Kurtosis		0.68	0.45
easyB1	Mean		5.1	0.13
	95% Confidence interval for Mean	Lower bound	4.83	
		Upper bound	5.36	
	5% trimmed mean		5.18	
	Median		5	
	Variance		2	
	Std. Deviation		1.41	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		1	
	Skewness		-0.94	0.23
	Kurtosis		0.51	0.45
easyB2	Mean		5.56	0.1
	95% Confidence interval for Mean	Lower bound	5.36	
		Upper bound	5.77	
	5% trimmed mean		5.64	
	Median		6	
	Variance		1.2	
	Std. Deviation		1.1	
	Minimum		2	
	Maximum		7	

	Range		5	
	Interquartile range		1	
	Skewness		-0.93	0.23
	Kurtosis		0.95	0.45
navigationA	Mean		5.67	0.12
	95% Confidence interval for Mean	Lower bound	5.43	
		Upper bound	5.91	
	5% trimmed mean		5.77	
	Median		6	
	Variance		1.7	
	Std. Deviation		1.3	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		2	
	Skewness		-1.03	0.23
	Kurtosis		0.79	0.45
navigationB	Mean		5.49	0.13
	95% Confidence interval for Mean	Lower bound	5.24	
		Upper bound	5.74	
	5% trimmed mean		5.59	
	Median		6	
	Variance		1.78	
	Std. Deviation		1.34	
	Minimum		2	
	Maximum		7	
	Range		5	
	Interquartile range		1	
	Skewness		-0.92	0.23
	Kurtosis		0.19	0.45
findA	Mean		5.81	0.1
	95% Confidence interval for Mean	Lower bound	5.61	
		Upper bound	6.02	
	5% trimmed mean		5.91	
	Median		6	
	Variance		1.18	
	Std. Deviation		1.09	
	Minimum		2	

	Maximum		7	
	Range		5	
	Interquartile range		2	
	Skewness		-1.12	0.23
	Kurtosis		1.19	0.45
findB	Mean		5.79	0.11
	95% Confidence interval for Mean	Lower bound	5.57	
		Upper bound	6	
	5% trimmed mean		5.88	
	Median		6	
	Variance		1.31	
	Std. Deviation		1.14	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		2	
	Skewness		-1.23	0.23
	Kurtosis		2.4	0.45
easyA	Mean		5.64	0.12
	95% Confidence interval for Mean	Lower bound	5.41	
		Upper bound	5.88	
	5% trimmed mean		5.74	
	Median		6	
	Variance		1.55	
	Std. Deviation		1.24	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		2	
	Skewness		-1.03	0.23
	Kurtosis		1.09	0.45
easyB	Mean		5.05	0.14
	95% Confidence interval for Mean	Lower bound	4.78	
		Upper bound	5.33	
	5% trimmed mean		5.14	
	Median		5	
	Variance		2.12	

	Std. Deviation		1.46	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		2	
	Skewness		-0.91	0.23
	Kurtosis		0.53	0.45
accurateA	Mean		6.15	0.1
	95% Confidence interval for Mean	Lower bound	5.96	
		Upper bound	6.34	
	5% trimmed mean		6.26	
	Median		6	
	Variance		1.01	
	Std. Deviation		1.01	
	Minimum		2	
	Maximum		7	
	Range		5	
	Interquartile range		1	
	Skewness		-1.66	0.23
	Kurtosis		3.88	0.45
accurateB	Mean		5.82	0.12
	95% Confidence interval for Mean	Lower bound	5.59	
		Upper bound	6.06	
	5% trimmed mean		5.97	
	Median		6	
	Variance		1.59	
	Std. Deviation		1.26	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		2	
	Skewness		-1.58	0.23
	Kurtosis		3.41	0.45
organisationA	Mean		5.98	0.1
	95% Confidence interval for Mean	Lower bound	5.78	
		Upper bound	6.18	
	5% trimmed mean		6.1	
	Median		6	

	Variance		1.17	
	Std. Deviation		1.08	
	Minimum		2	
	Maximum		7	
	Range		5	
	Interquartile range		1	
	Skewness		-1.44	0.23
	Kurtosis		2.55	0.45
organisationB	Mean		5.8	0.11
	95% Confidence interval for Mean	Lower bound	5.59	
		Upper bound	6.02	
	5% trimmed mean		5.89	
	Median		6	
	Variance		1.31	
	Std. Deviation		1.15	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile range		2	
	Skewness		-1.07	0.23
	Kurtosis		1.7	0.45
wrong1	Mean		1.22	0.18
	95% Confidence interval for Mean	Lower bound	0.86	
		Upper bound	1.59	
	5% trimmed mean		1	
	Median		0	
	Variance		3.74	
	Std. Deviation		1.93	
	Minimum		0	
	Maximum		7	
	Range		7	
	Interquartile range		2	
	Skewness		1.56	0.23
	Kurtosis		1.42	0.45
wrong2	Mean		1.24	0.2
	95% Confidence interval for Mean	Lower bound	0.85	
		Upper bound	1.63	

					Stat	
		5% trimmed mean			0.95	
		Median			0	
		Variance			4.29	
		Std. Deviation			2.07	
		Minimum			0	
		Maximum			11	
		Range			11	
		Interquartile range			2	
		Skewness			2.2	0.23
		Kurtosis			5.68	0.45

Table A.39: Descriptives for within experiment

	Gr	Descriptives			Stat	Std. Err
easyA1	1	Mean			5.37	.148
		95% Confidence Interval for Mean	Lower Bound		5.07	
			Upper Bound		5.66	
		5% Trimmed Mean			5.41	
		Median			5.50	
		Variance			1.138	
		Std. Deviation			1.067	
		Minimum			2	
		Maximum			7	
		Range			5	
		Interquartile Range			1	
		Skewness			-.691	.330
		Kurtosis			.794	.650
	2	Mean			5.95	.131
		95% Confidence Interval for Mean	Lower Bound		5.69	
			Upper Bound		6.21	
		5% Trimmed Mean			6.04	
		Median			6.00	
		Variance			1.031	
		Std. Deviation			1.016	
		Minimum			2	
		Maximum			7	

		Range		5	
		Interquartile Range		1	
		Skewness		-1.404	.309
		Kurtosis		2.897	.608
easyA2	1	Mean		5.73	.132
		95% Confidence Interval for Mean	Lower Bound	5.47	
			Upper Bound	6.00	
		5% Trimmed Mean		5.78	
		Median		6.00	
		Variance		.906	
		Std. Deviation		.952	
		Minimum		3	
		Maximum		7	
		Range		4	
		Interquartile Range		1	
		Skewness		-.559	.330
		Kurtosis		.148	.650
	2	Mean		5.85	.144
		95% Confidence Interval for Mean	Lower Bound	5.56	
			Upper Bound	6.14	
		5% Trimmed Mean		5.93	
		Median		6.00	
		Variance		1.248	
		Std. Deviation		1.117	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		2	
		Skewness		-1.052	.309
		Kurtosis		1.066	.608
easyB1	1	Mean		4.65	.194
		95% Confidence Interval for Mean	Lower Bound	4.26	
			Upper Bound	5.04	
		5% Trimmed Mean		4.77	
		Median		5.00	
		Variance		1.956	
		Std. Deviation		1.399	
		Minimum		1	

		Maximum		6	
		Range		5	
		Interquartile Range		2	
		Skewness		-1.045	.330
		Kurtosis		.316	.650
	2	Mean		5.48	.171
		95% Confidence Interval for Mean	Lower Bound	5.14	
			Upper Bound	5.82	
		5% Trimmed Mean		5.59	
		Median		6.00	
		Variance		1.745	
		Std. Deviation		1.321	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-1.012	.309
		Kurtosis		.824	.608
easyB2	1	Mean		5.35	.155
		95% Confidence Interval for Mean	Lower Bound	5.03	
			Upper Bound	5.66	
		5% Trimmed Mean		5.39	
		Median		5.50	
		Variance		1.250	
		Std. Deviation		1.118	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-.560	.330
		Kurtosis		.265	.650
	2	Mean		5.75	.136
		95% Confidence Interval for Mean	Lower Bound	5.48	
			Upper Bound	6.02	
		5% Trimmed Mean		5.85	
		Median		6.00	
		Variance		1.106	
		Std. Deviation		1.052	

		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-1.374	.309
		Kurtosis		2.566	.608
aestheticsA	1	Mean		5.04	.221
		95% Confidence Interval for Mean	Lower Bound	4.59	
			Upper Bound	5.48	
		5% Trimmed Mean		5.12	
		Median		5.00	
		Variance		2.548	
		Std. Deviation		1.596	
		Minimum		1	
		Maximum		7	
		Range		6	
		Interquartile Range		2	
		Skewness		-.547	.330
		Kurtosis		-.488	.650
	2	Mean		5.75	.161
		95% Confidence Interval for Mean	Lower Bound	5.43	
			Upper Bound	6.07	
		5% Trimmed Mean		5.87	
		Median		6.00	
		Variance		1.547	
		Std. Deviation		1.244	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		2	
		Skewness		-1.364	.309
		Kurtosis		1.790	.608
aestheticsB	1	Mean		3.92	.259
		95% Confidence Interval for Mean	Lower Bound	3.40	
			Upper Bound	4.44	
		5% Trimmed Mean		3.91	
		Median		4.00	
		Variance		3.484	

		Std. Deviation			1.867	
		Minimum			1	
		Maximum			7	
		Range			6	
		Interquartile Range			3	
		Skewness			.022	.330
		Kurtosis			-1.103	.650
	2	Mean			4.73	.215
		95% Confidence Interval for Mean	Lower Bound		4.30	
			Upper Bound		5.16	
		5% Trimmed Mean			4.76	
		Median			5.00	
		Variance			2.775	
		Std. Deviation			1.666	
		Minimum			2	
		Maximum			7	
		Range			5	
		Interquartile Range			3	
		Skewness			-.493	.309
		Kurtosis			-1.007	.608
navigationA	1	Mean			5.62	.160
		95% Confidence Interval for Mean	Lower Bound		5.29	
			Upper Bound		5.94	
		5% Trimmed Mean			5.67	
		Median			6.00	
		Variance			1.339	
		Std. Deviation			1.157	
		Minimum			3	
		Maximum			7	
		Range			4	
		Interquartile Range			2	
		Skewness			-.450	.330
		Kurtosis			-.706	.650
	2	Mean			5.72	.184
		95% Confidence Interval for Mean	Lower Bound		5.35	
			Upper Bound		6.09	
		5% Trimmed Mean			5.85	
		Median			6.00	

		Variance		2.037	
		Std. Deviation		1.427	
		Minimum		1	
		Maximum		7	
		Range		6	
		Interquartile Range		2	
		Skewness		-1.324	.309
		Kurtosis		1.390	.608
navigationB	1	Mean		5.35	.176
		95% Confidence Interval for Mean	Lower Bound	4.99	
			Upper Bound	5.70	
		5% Trimmed Mean		5.43	
		Median		6.00	
		Variance		1.603	
		Std. Deviation		1.266	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-.995	.330
		Kurtosis		.605	.650
	2	Mean		5.62	.180
		95% Confidence Interval for Mean	Lower Bound	5.26	
			Upper Bound	5.98	
		5% Trimmed Mean		5.72	
		Median		6.00	
		Variance		1.935	
		Std. Deviation		1.391	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		2	
		Skewness		-.953	.309
		Kurtosis		.091	.608
findA	1	Mean		5.77	.139
		95% Confidence Interval for Mean	Lower Bound	5.49	
			Upper Bound	6.05	
		5% Trimmed Mean		5.84	

		Median			6.00	
		Variance			1.005	
		Std. Deviation			1.002	
		Minimum			3	
		Maximum			7	
		Range			4	
		Interquartile Range			1	
		Skewness			-.970	.330
		Kurtosis			.865	.650
	2	Mean			5.85	.150
		95% Confidence Interval for Mean	Lower Bound		5.55	
			Upper Bound		6.15	
		5% Trimmed Mean			5.96	
		Median			6.00	
		Variance			1.350	
		Std. Deviation			1.162	
		Minimum			2	
		Maximum			7	
		Range			5	
		Interquartile Range			2	
		Skewness			-1.240	.309
		Kurtosis			1.450	.608
findB	1	Mean			5.48	.166
		95% Confidence Interval for Mean	Lower Bound		5.15	
			Upper Bound		5.81	
		5% Trimmed Mean			5.56	
		Median			5.50	
		Variance			1.431	
		Std. Deviation			1.196	
		Minimum			1	
		Maximum			7	
		Range			6	
		Interquartile Range			1	
		Skewness			-1.025	.330
		Kurtosis			2.529	.650
	2	Mean			6.05	.133
		95% Confidence Interval for Mean	Lower Bound		5.78	
			Upper Bound		6.32	

		5% Trimmed Mean		6.15	
		Median		6.00	
		Variance		1.065	
		Std. Deviation		1.032	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-1.538	.309
		Kurtosis		3.147	.608
easyA	1	Mean		5.44	.179
		95% Confidence Interval for Mean	Lower Bound	5.08	
			Upper Bound	5.80	
		5% Trimmed Mean		5.53	
		Median		6.00	
		Variance		1.663	
		Std. Deviation		1.290	
		Minimum		1	
		Maximum		7	
		Range		6	
		Interquartile Range		1	
		Skewness		-1.065	.330
		Kurtosis		1.501	.650
	2	Mean		5.82	.153
		95% Confidence Interval for Mean	Lower Bound	5.51	
			Upper Bound	6.12	
		5% Trimmed Mean		5.91	
		Median		6.00	
		Variance		1.406	
		Std. Deviation		1.186	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		2	
		Skewness		-1.020	.309
		Kurtosis		.683	.608
easyB	1	Mean		4.77	.222
		95% Confidence Interval for Mean	Lower Bound	4.32	

			Upper Bound	5.22	
		5% Trimmed Mean		4.85	
		Median		5.00	
		Variance		2.573	
		Std. Deviation		1.604	
		Minimum		1	
		Maximum		7	
		Range		6	
		Interquartile Range		2	
		Skewness		-1.000	.330
		Kurtosis		.296	.650
	2	Mean		5.30	.165
		95% Confidence Interval for Mean	Lower Bound	4.97	
			Upper Bound	5.63	
		5% Trimmed Mean		5.35	
		Median		5.00	
		Variance		1.637	
		Std. Deviation		1.280	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-.540	.309
		Kurtosis		-.339	.608
accurateA	1	Mean		6.12	.136
		95% Confidence Interval for Mean	Lower Bound	5.84	
			Upper Bound	6.39	
		5% Trimmed Mean		6.23	
		Median		6.00	
		Variance		.967	
		Std. Deviation		.983	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-1.784	.330
		Kurtosis		5.037	.650
	2	Mean		6.18	.133

		95% Confidence Interval for Mean	Lower Bound	5.92	
			Upper Bound	6.45	
		5% Trimmed Mean		6.30	
		Median		6.00	
		Variance		1.068	
		Std. Deviation		1.033	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-1.622	.309
		Kurtosis		3.477	.608
accurateB	1	Mean		5.62	.189
		95% Confidence Interval for Mean	Lower Bound	5.24	
			Upper Bound	5.99	
		5% Trimmed Mean		5.77	
		Median		6.00	
		Variance		1.849	
		Std. Deviation		1.360	
		Minimum		1	
		Maximum		7	
		Range		6	
		Interquartile Range		2	
		Skewness		-1.591	.330
		Kurtosis		3.493	.650
	2	Mean		6.00	.148
		95% Confidence Interval for Mean	Lower Bound	5.70	
			Upper Bound	6.30	
		5% Trimmed Mean		6.13	
		Median		6.00	
		Variance		1.322	
		Std. Deviation		1.150	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		2	
		Skewness		-1.523	.309
		Kurtosis		3.090	.608

organisationA	1	Mean		5.94	.149
		95% Confidence Interval for Mean	Lower Bound	5.64	
			Upper Bound	6.24	
		5% Trimmed Mean		6.06	
		Median		6.00	
		Variance		1.153	
		Std. Deviation		1.074	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		2	
		Skewness		-1.363	.330
		Kurtosis		2.828	.650
	2	Mean		6.02	.142
		95% Confidence Interval for Mean	Lower Bound	5.73	
			Upper Bound	6.30	
		5% Trimmed Mean		6.13	
		Median		6.00	
		Variance		1.203	
		Std. Deviation		1.097	
		Minimum		2	
		Maximum		7	
		Range		5	
		Interquartile Range		1	
		Skewness		-1.549	.309
		Kurtosis		2.683	.608
organisationB	1	Mean		5.63	.165
		95% Confidence Interval for Mean	Lower Bound	5.30	
			Upper Bound	5.97	
		5% Trimmed Mean		5.74	
		Median		6.00	
		Variance		1.413	
		Std. Deviation		1.189	
		Minimum		1	
		Maximum		7	
		Range		6	
		Interquartile Range		1	
		Skewness		-1.354	.330

		Kurtosis		3.328	.650
	2	Mean		5.95	.141
		95% Confidence Interval for Mean	Lower Bound	5.67	
			Upper Bound	6.23	
		5% Trimmed Mean		6.02	
		Median		6.00	
		Variance		1.201	
		Std. Deviation		1.096	
		Minimum		3	
		Maximum		7	
		Range		4	
		Interquartile Range		2	
		Skewness		-.778	.309
		Kurtosis		-.371	.608
wrong1	1	Mean		1.00	.230
		95% Confidence Interval for Mean	Lower Bound	.54	
			Upper Bound	1.46	
		5% Trimmed Mean		.79	
		Median		0.00	
		Variance		2.745	
		Std. Deviation		1.657	
		Minimum		0	
		Maximum		6	
		Range		6	
		Interquartile Range		2	
		Skewness		1.668	.330
		Kurtosis		2.002	.650
	2	Mean		1.42	.276
		95% Confidence Interval for Mean	Lower Bound	.86	
			Upper Bound	1.97	
		5% Trimmed Mean		1.19	
		Median		0.00	
		Variance		4.586	
		Std. Deviation		2.142	
		Minimum		0	
		Maximum		7	
		Range		7	
		Interquartile Range		3	

		Skewness			1.415	.309
		Kurtosis			.831	.608
wrong2	1	Mean			1.35	.249
		95% Confidence Interval for Mean	Lower Bound		.85	
			Upper Bound		1.85	
		5% Trimmed Mean			1.20	
		Median			0.00	
		Variance			3.211	
		Std. Deviation			1.792	
		Minimum			0	
		Maximum			6	
		Range			6	
		Interquartile Range			3	
		Skewness			.984	.330
		Kurtosis			-.355	.650
	2	Mean			1.15	.297
		95% Confidence Interval for Mean	Lower Bound		.56	
			Upper Bound		1.74	
		5% Trimmed Mean			.76	
		Median			0.00	
		Variance			5.282	
		Std. Deviation			2.298	
		Minimum			0	
		Maximum			11	
		Range			11	
		Interquartile Range			2	
		Skewness			2.722	.309
		Kurtosis			7.739	.608

Table A.40: Descriptive statistics by group factor for within experiment

Wilcoxon Signed Ranks Test				
Ranks		N	Mean Rank	Sum of Ranks
aestheticsB - aestheticsA	Negative Ranks	62.a	38.97	2416.00
	Positive Ranks	10.b	21.20	212.00
	Ties	40.c		

	Total	112		
easyB1 - easyA1	Negative Ranks	43.d	28.30	1217.00
	Positive Ranks	10.e	21.40	214.00
	Ties	59.f		
	Total	112		
easyB2 - easyA2	Negative Ranks	33.g	25.68	847.50
	Positive Ranks	16.h	23.59	377.50
	Ties	63.i		
	Total	112		
navigationB - navigationA	Negative Ranks	35.j	26.91	942.00
	Positive Ranks	18.k	27.17	489.00
	Ties	59.l		
	Total	112		
findB - findA	Negative Ranks	26.m	22.60	587.50
	Positive Ranks	21.n	25.74	540.50
	Ties	65.o		
	Total	112		
easyB - easyA	Negative Ranks	43.p	28.47	1224.00
	Positive Ranks	10.q	20.70	207.00
	Ties	59.r		
	Total	112		
accurateB - accurateA	Negative Ranks	37.s	25.70	951.00
	Positive Ranks	13.t	24.92	324.00
	Ties	62.u		
	Total	112		
organisationB - organisationA	Negative Ranks	32.v	22.38	716.00
	Positive Ranks	15.w	27.47	412.00
	Ties	65.x		
	Total	112		
wrong2 - wrong1	Negative Ranks	27.y	25.61	691.50
	Positive Ranks	26.z	28.44	739.50
	Ties	59.aa		
	Total	112		

a. aestheticsB < aestheticsA

b. aestheticsB > aestheticsA

c. aestheticsB = aestheticsA

d. easyB1 < easyA1

e. easyB1 > easyA1

f. easyB1 = easyA1

g. easyB2 < easyA2

h. easyB2 > easyA2

i. easyB2 = easyA2

j. navigationB < navigationA

k. navigationB > navigationA

l. navigationB = navigationA

m. findB < findA

n. findB > findA

o. findB = findA

p. easyB < easyA

q. easyB > easyA

r. easyB = easyA

s. accurateB < accurateA

t. accurateB > accurateA

u. accurateB = accurateA

v. organisationB < organisationA

w. organisationB > organisationA

x. organisationB = organisationA

y. wrong2 < wrong1

z. wrong2 > wrong1

aa. wrong2 = wrong1

Table A.41: Non parametric ranks for within experiment

Test Statistics .a

	aesthetics B - A	easy B1 - A1	easy B2 - A2	navigation B - A	find B - A	easy B - A	accurate B - A	organisation B - A	wrong 2 - 1
Z	-6,270b	-4,581b	-2,442b	-2,129b	-,262b	-4,613b	-3,204b	-1,677b	-,215c
Asymp. Sig. (2-tailed)	.000	.000	.015	.033	.793	.000	.001	.094	.830

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

Table A.42: Non parametric treatment results, Wilcozon Z significance.

Bibliography

[1] ISO 9241-210: 2010. *Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems.*

[2] National Gallery of Art - Teaching Art Nouveau: Glossary.

[3] 9241-11. *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)–Part II Guidance on Usability*, 1998.

[4] OFFF conference - Post-Digital creation culture, 2007.

[5] EN ISO 9241-151 Ergonomics of human-system interaction — Part 151: Guidance on World Wide Web user interfaces. BSi British Standards, SI Group Headquarters 389 Chiswick High Road, London W4 4AL, UK Tel +44 (0)20 8996 9001 Fax +44 (0)20 8996 7001 www.bsigroup.com/standards, 2008.

[6] WAMMI reliability results. http://www.wammi.com/reliability.html, 2008.

[7] Bootstrap 2.2.2, http://getbootstrap.com/, 2013.

[8] D. Adams. *The Hitch Hiker's Guide to the Galaxy.* Del Rey, 1997.

[9] Amazon. https://www.mturk.com/mturk/.

[10] Archer B. The need for design education. Royal College of Art, 1973.

[11] P.G. Bahn. *The Cambridge illustrated history of prehistoric art.* Cambridge University Press New York, 1998.

[12] W. Benjamin, B. Doherty, T.Y. Levin, and E. Jephcott. *The work of art in the age of its technological reproducibility, and other writings on media.* Belknap Pr, 2008.

[13] G.D. Birkhoff. Aesthetic measure. *Cambridge*, 1933.

[14] M. Bosnjak and T.L. Tuten. Classifying response behaviors in web-based surveys. *Journal of Computer-Mediated Communication*, 6(3):0–0, 2001.

[15] K. Boussora and S. Mazouz. The Use of the Golden Section in the Great Mosque at Kairouan. *Nexus Network Journal*, 6(1):7–16, 2004.

[16] N.M. Bradburn. Respondent burden. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, volume 1978, pages 35–40, 1978.

[17] J. Brooke. SUS A quick and dirty usability scale. *Usability evaluation in industry*, pages 189–194, 1996.

[18] BuiltWith. Google Analytics Usage Statistics, November 2010.

[19] Douglas Burnham. Kant's aesthetics.

[20] D. Dutton. *The art instinct: Beauty, pleasure, and human evolution*. Bloomsbury, 2008.

[21] A. Edgar. An Introduction to Adorno's Aesthetics. *The British Journal of Aesthetics*, 30:46–56, 1990.

[22] P. Ekman. Basic emotions. *Handbook of Cognition and Emotion*, pages 45–60, 1999.

[23] Gustav Theodor Fechner. *Vorschule der Ästhetik [Primer of aesthetics]*. Breitkopf & Hartel, 1876.

[24] A.P. Field and G. Hole. *How to design and report experiments*. Sage Publications Ltd, 2003.

[25] Gary Genosko. *Marshall McLuhan: Critical evaluations in cultural theory*. Series: Critical evaluations in cultural theory. Routledge, London; New York, 2005.

[26] Matila Ghyka. *The Geometry of Art and Life*, volume 2. Dover Publications, 1946.

[27] S.J. Gould. The mismeasure of man (Revised ed.). *New York: Norton*, 1986:365–66, 1981.

[28] Christopher D. Green. All that glitters: A review of psychological research on the aesthetics of the golden section. *Perception*, page 937, 1995.

[29] T. Grill and M. Scanlon. *Photographic composition*. Amphoto Books, 1990.

[30] J. Gronow. Taste and fashion: The social function of fashion and style. *Acta Sociologica*, 36(2):89, 1993.

[31] P.M. Groves and R.F. Thompson. Habituation: A dual-process theory. *Psychological Review*, 77(5):419, 1970.

[32] Jan Hartmann. *User Quality Judgements of Interactive Systems*. PhD thesis, University of Manchester, 2008.

[33] M. Hassenzahl. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4):319–349, 2004.

[34] Mark Hassenzahl and Mark Blythe. Interview with Don Norman. *Interactions*, 11(5):43–46, 2004.

[35] M. Heath. Pleasure in Aristotle's aesthetics. University of Leeds, 2005.

[36] A. Herzog and J.G. Bachman. Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45(4):549, 1981.

[37] Christoph Hölscher and Gerhard Strube. Web search behavior of internet experts and newbies. In *Proceedings of the 9th international World Wide Web conference on Computer networks: the International Journal of Computer and Telecommunications Networking*, pages 337–346, Amsterdam, The Netherlands, 2000. North-Holland Publishing Co.

[38] Cyril Jaquier and Kévin Drapel. Le nombre d'or: réalité ou interprétations douteuses? Projet STS. 2005.

[39] Patrick W. Jordan. *Designing Pleasurable Products, An introduction to the new Human Factors*. Taylor & Francis, London and New York, 2000.

[40] Immanuel Kant and James Creed Meredith. *The critique of judgement*. Clarendon Press, Oxford, 1952.

[41] John Karat and Alistair Sutcliffe. Can we measure quality in design and do we need to? In *Proceedings of the 6th ACM conference on Designing Interactive systems*, pages 119–121, University Park, PA, USA, 2006. K. Karat - IBM Development and Research and A. Sutcliffe - University of Manchester, ACM Press.

[42] Kristiina Karvonen. The beauty of simplicity. In *Proceedings of the Conference on Universal Usability ACM*, Arlington, 2000.

[43] S.B. Kim. Is fashion art? *Fashion Theory: The Journal of Dress*, 2(1):51–71, 1998.

[44] J. Kirakowski, N. Claridge, and R. Whitehand. Human centered-measures of success in web site design. In *Proceedings of the Fourth Conference on Human Factors & the Web*, 1998.

[45] Jurek Kirakowski and Nigel Claridge. WAMMI - publications. http://www.wammi.com/publications.html.

[46] R. Kominis. Aesthetics as an attribute to usability: A critique of (some) previous works. *Multimedia and E-Content Trends*, pages 137–145.

[47] Kirstin Krauss and Ralf Hoffmann. A critical evaluation of literature on visual aesthetics for the web. In *SAICSIT*, pages 205 – 209, 2004.

[48] Ray Kristof and Amy Satran. *Interactivity by design: Creating and communicating with new media*. Adobe Press, Mountain View, CA, 1995.

[49] M. Kurosu and K. Kashimura. Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. In *Conference on Human Factors in Computing Systems*, pages 292–293. ACM New York, NY, USA, 1995.

[50] W. Lidwell, K. Holden, and J. Butler. *Universal principles of design*. Rockport Publishers, 2003.

[51] Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and J. Brown. Behaviour and Information Technology; Attention web designers: You have 50 milliseconds to make a good first impression! 25(2):115–126, 2006.

[52] Peter Markie. Rationalism vs. empiricism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. 2008.

[53] Winter Mason and Duncan J Watts. Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter*, 11(2):100–108, 2010.

[54] Marshall McLuhan and Quentin Fiore. *The Medium is the Message*. Gingko Press, 2001.

[55] V. Mendoza and D.G. Novick. Usability over time. *University of Texas at El Paso (Departmental Paper)*, 2005.

[56] Kevin Mullet and Darrell Sano. *Designing visual interfaces: Communication oriented techniques*. SunSoft Press, Upper Saddle River, NJ, 1995.

[57] P. Murano. *Usability investigation of anthropomorphic user interface feedback*. PhD thesis, Heriot-Watt University, Computer Science, 2009.

[58] Nicholas Negroponte. *Being digital*. Random House, New York, 1996.

[59] D.C.L. Ngo, L.S. Teo, and J.G. Byrne. Modelling interface aesthetics. *Information Sciences*, 152:25–46, 2003.

[60] J. Nielsen. *Usability engineering*. Morgan Kaufmann, 1993.

[61] J. Nielsen. Novice vs. expert users. *Alertbox, February*, 6:2000, 2000.

[62] Jakob Nielsen. *Designing web usability*. New Riders, Indianapolis, 2000.

[63] Jakob Nielsen. Screen resolution and page layout. online, 2006.

[64] Richard E. Nisbett and Timothy D. Wilson. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35(4):250–256, 1977.

[65] I Niven. *Irrational numbers*. The Mathematical Association of America, 2005.

[66] Don Norman. Design as communication, 2004.

[67] Donald A. Norman. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, 2004.

[68] M. O'Brien and N. Sibley. The Photographic Eye: Learning to See with a Camera, 1995.

[69] A. Ortony and T.J. Turner. What's basic about basic emotions? *Psychological Review*, 97(3):315–331, 1990.

[70] Mark Packer. The Conditions of Aesthetic Feeling in Aristotle's Poetics. *British Journal of Aesthetics*, 24(2):138–148, 1984.

[71] Gabriele Paolacci, Jesse Chandler, and Panagiotis Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419, 2010.

[72] Plato. *The Republic*. Penguin Classics, 3rd edition, May 2007.

[73] Joshua Porter. Do canonical web designs exist? blog, November 2007.

[74] Susanne Rässler, Florian Koller, and Christine Mäenpää. A Split Questionnaire Survey Design applied to German Media and Consumer Surveys. Lehrstuhl für Statistik und Ökonometrie Universität Erlangen-Nürnberg, 2002.

[75] Mimi Recker. Website usability questionnaire. online, Utah University, Free Online Course Materials, (http://ocw.usu.edu/instructional-technology-learning-sciences/producing-distance-education-resources/resource30.html), May 2008.

[76] J. Rigau, M. Feixas, and M. Sbert. Conceptualizing Birkhoff's Aesthetic Measure Using Shannon Entropy and Kolmogorov Complexity. In *Proceedings of the Eurographics Workshop Computational Aesthetics in Graphics, Visualization and Imaging*, pages 105–112, 2007.

[77] M Rossenthal and P Yudin. *Litle Philosophic Dictionary*. G. Kaklamani (Greek edition), 1962.

[78] B. Russo and A. de Moraes. The lack of usability in design icons: An affective case study about Juicy Salif. In *Proceedings of the 2003 international conference on Designing Pleasurable Products and Interfaces*, pages 146–147. ACM, 2003.

[79] R. Scha and R. Bod. Computational esthetics. *Informatie en Informatiebeleid*, 11(1):54–63, 1993.

[80] S.S. Shapiro and M.B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[81] Georg Simmel. Fashion. *International Quarterly*, 62(6):541–558, May 1957.

[82] J. Spool. Interview-based tasks: Learning from Leonardo DiCaprio, March 2006.

[83] T. Staudek. On Birkhoff's aesthetic measure of vases. *Faculty of Information, Masaryk University Report Series (FIMU-RS-99-06)*, 1999.

[84] Daniel Stoljar. Physicalism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2009.

[85] A. Sutcliffe and A. De Angeli. Assessing interaction styles in web user interfaces. *Lecture notes in computer science, University of Manchester*, 3585:405, 2005.

[86] A. Sutcliffe and A. De Angeli. Interaction, usability and aesthetics: What influences users' preferences? In *Proceedings of the 6th ACM conference on Designing Interactive systems, Symposium on Designing Interactive Systems*, pages 271 – 280, University Park, PA, USA, 2006. ACM Press New York, NY, USA.

[87] Alistair Sutcliffe. Assessing the reliability of heuristic evaluation for website attractiveness and usability. In *Proceedings of the 35th Hawaii International Conference on System Sciences*, Hawaii, 2002.

[88] Alistair Sutcliffe. Is beautiful usable? 2007.

[89] Alistair Sutcliffe, Antonella De Angeli, and Jan Hartmann. Investigating attractiveness in web user interfaces. In *CHI Proceedings Empirical Studies of Web Interaction*, pages 287–396, San Jose, CA, USA, 2007. ACM.

[90] Leo Tolstoy and Aylmer Maude. *What is art?* Hackett Publishing, Indianapolis, 1996.

[91] N Tractinsky, A S Katz, and D Ikar. What is beautiful is usable. *Elsevier*, 13(2):127 – 145, 2000.

[92] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453, 1981.

[93] Armin Vit. Landmark web sites, where art thou? blog, October 2007.

[94] J. Wolff. *The social production of art.* Macmillan, 1981.

[95] Luke Wroblewski, Nan Goggin, and Jennifer Gunji. *Site-seeing: A Visual Approach to Web Usability.* Hungry Minds, 2002.

[96] J.M. Zain, M. Tey, and G.Y. Soon. Using Aesthetic Measurement Application (AMA) to Measure Aesthetics of Web Page Interfaces. In *Natural Computation*, volume 6. ICNC'08 Fourth International Conference on Natural Computation, 2008.

[97] Nick Zangwill. Aesthetic judgment. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Fall 2008 edition, 2008.

[98] Herbert Zettl. *Sight, Sound, Motion: applied media aesthetics*, volume 3. Wadsworth Publishing Company, 1998.

[99] D.W. Zimmerman and B.D. Zumbo. Can percentiles replace raw scores in the statistical analysis of test data? *Educational and psychological measurement*, 65(4):616, 2005.