Visual complexity in human-machine interaction

Visuelle Komplexität in der Mensch-Maschine Interaktion

zur Erlangung des akademischen Grades eines
DOKTORS DER
INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des

Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

M. Sc. Fabian Ries

Tag der mündlichen Prüfung: 20.11.2020

Hauptreferentin: Prof. Dr.-Ing. Barbara Deml

Korreferent: Prof. Dr. phil. Klaus Opwis

Abstract

Visual complexity is often defined as “the level of detail or intricacy contained within an image” (Forsythe, 2009, p. 158; Snodgrass & Vanderwart, 1980, p. 183). It affects many areas of everyday life, including those that rely on the interaction with technology. For example, effects of visual complexity have been demonstrated in road traffic (Edquist, Rudin-Brown, & Lenné, 2012; Mace & Pollack, 1983) or for the interaction with software (Alemerien & Magel, 2014) or websites (Deng & Poole, 2010; Tuch et al., 2011). Although research on visual complexity has already had its beginning with the Gestalt psychologists, who incorporated the meaning of simplicity and complexity in the perception process for example with the Gestalt principle of *Prägnanz* (Koffka, 1935; Wertheimer, 1923), neither the influencing factors of visual complexity nor the connections with eye movements or mental workload have yet been conclusively investigated. The present study addresses these points by means of four empirical studies.

Study 1 examines the significance of the construct in human-machine interaction on the basis of the complexity of videos in control rooms as well as their effects on subjective, physiological and performance measures of mental workload. Study 2 takes a closer look at the dimensional structure and the significance of influencing variables and factors of visual complexity, using different types of stimuli. Study 3 applies an experimental approach in order to investigate the effects of visual complexity on subjective ratings and a selection of ocular parameters with simple black and white shape patterns serving as stimuli. In addition, various computational and ocular parameters are used to predict complexity ratings. In study 4, this approach is transferred to screenshots of websites in order to investigate the validity of the conclusions within a field of application.

Findings from the studies extend the existing body of research. Associations with mental workload particularly suggest that visual complexity is a relevant construct within human-machine interaction. Quantitative and structural, but potentially also other aspects have an influence on the perception of visual complexity as well as the observer’s viewing behavior. The acquired results also allow for conclusions about the associations with computational measures, which in combination with ocular parameters are well suited for predicting complexity ratings.

The insights provided by the studies are finally discussed in the context of previous research, whereby an integrative research model of visual complexity in human-machine interaction is derived.

Zusammenfassung

Visuelle Komplexität wird oft als der Grad an Detail oder Verworrenheit in einem Bild definiert (Snodgrass & Vanderwart, 1980). Diese hat Einfluss auf viele Bereiche des menschlichen Lebens, darunter auch solche, die die Interaktion mit Technologie involvieren. So wurden Effekte visueller Komplexität etwa im Straßenverkehr (Edquist et al., 2012; Mace & Pollack, 1983) oder bei der Interaktion mit Software (Alemerien & Magel, 2014) oder Webseiten (Deng & Poole, 2010; Tuch et al., 2011) nachgewiesen. Obwohl die Erforschung visueller Komplexität bereits bis auf die Gestaltpsychologen zurückgeht, welche etwa mit dem Gestaltprinzip der Prägnanz die Bedeutung von Simplizität und Komplexität im Wahrnehmungsprozess verankerten (Koffka, 1935; Wertheimer, 1923), sind weder die Einflussfaktoren visueller Komplexität, noch die Zusammenhänge mit Blickbewegungen oder mentaler Beanspruchung bisher abschließend erforscht. Diese Punkte adressiert die vorliegende Arbeit mithilfe von vier empirischen Forschungsarbeiten.

In Studie 1 wird anhand der Komplexität von Videos in Leitwarten sowie der Effekte auf subjektive, physiologische und Leistungsparameter mentaler Beanspruchung die Bedeutung des Konstruktes im Bereich der Mensch-Maschine Interaktion untersucht. Studie 2 betrachtet die dimensionale Struktur und die Bedeutung verschiedener Einflussfaktoren visueller Komplexität genauer, wobei unterschiedliches Stimulusmaterial genutzt wird. In Studie 3 werden mithilfe eines experimentellen Ansatzes die Auswirkungen von Einflussfaktoren visueller Komplexität auf subjektive Bewertungen sowie eine Auswahl okularer Parameter untersucht. Als Stimuli dienen dabei einfache, schwarz-weiße Formenmuster. Zudem werden verschiedene computationale und okulare Parameter genutzt, um anhand dieser Komplexitätsbewertungen vorherzusagen. Dieser Ansatz wird in Studie 4 auf Screenshots von Webseiten übertragen, um die Aussagekraft in einem anwendungsnahen Bereich zu untersuchen.

Neben vorangegangenen Forschungsarbeiten legen insbesondere die gefundenen Zusammenhänge mit mentaler Beanspruchung nahe, dass visuelle Komplexität ein relevantes Konstrukt im Bereich der Mensch-Maschine Interaktion darstellt. Dabei haben insbesondere quantitative und strukturelle, aber potentiell auch weitere Aspekte Einfluss auf die Bewertung visueller Komplexität sowie auf das Blickverhalten der Betrachter. Die gewonnenen Ergebnisse erlauben darüber hinaus Rückschlüsse auf die

Zusammenhänge mit computationalen Maßen, welche in Kombination mit okularen Parametern gut für die Vorhersage von Komplexitätsbewertungen geeignet sind.
Die Erkenntnisse aus den durchgeführten Studien werden im Kontext vorheriger Forschungsarbeiten diskutiert. Daraus wird ein integratives Forschungsmodell visueller Komplexität in der Mensch-Maschine-Interaktion abgeleitet.

Acknowledgements

At this place, I would like to thank those persons, who have helped me along the journey of this work.

First of all, I would like to thank my main supervisor, Prof. Barbara Deml, who gave me the opportunity work at her institute and to write this thesis. She helped me to find a topic that did not become boring even after all the time I spent on it. Through her input and support, as well as the trust and freedom I was given, the work was able to take on the form in which it is now.

Moreover, many thanks are due to Prof. Klaus Opwis, who has taken over the secondary supervision of the work. His immediate commitment, the very friendly contact and his helpful comments greatly facilitated the completion of this work.

Furthermore, I would like to thank Prof. Andreas Groll, who provided me with a yet unpublished beta-version of the R-package glmmLasso as well as enormously helpful support with its use. Our conversations on ordinal regression models and his feedback were very enriching for me.

A very special thanks goes to my colleagues at ifab. I really enjoyed the good atmosphere, the common lunch breaks, the table football matches, the Christmas and summer celebrations and many other events. You made my four years at the institute a very special time for me. In particular, I would like to thank Jonas Imbsweiler for spreading a good mood and questionable music as well as uncountable fun facts and science news, Elena Wolf for supporting me with her expertise on physiological data and sharing both her office and her humour with me for quite a while, and Manuel Zaremski for his enormous helpfulness and for infecting me with his passion for good coffee.

Moreover, thanks to my research assistants Floria Krusche, Jascha Löbel, Dániel Kiss-Illés and Marco Wetter for their assistance with data collection and preparation as well as the fabulous collaboration in the further development and implementation of the many ocular and computational measures for visual complexity.

Thanks to the graduate workers Yang Xie, Jessica Waibel, Sibylle de Vandière and Yi Ding for the exciting exchange of ideas, the interesting discussions on visual complexity and your commitment for a topic that might appear rather exotic for most mechanical engineering students at first sight.

Thanks to my parents, who always supported me with love, trust, good food and many other things and therefore significantly contributed that I got to this point.

Especially, I would like to thank Wiebke for being by my side for quite some time now and for the beautiful moments we shared. Thanks for your consideration particularly within the last four years. Thanks for coming to Karlsruhe with me, for bearing patience, for motivating and supporting me to finish this work. Thanks for sacrificing on many evenings and vacations together without grumbling and for bravely enduring the uncertainty of what the future will bring.

Thanks to Karla for setting me a deadline (although I didn't quite meet it) and for being a giant part of my life since almost half a year. Thanks to both of you for the time we spent together. I am looking forward to our common future as a family and whatever adventures come next.

Karlsruhe, end of July 2020 Fabian Ries

Table of Contents

List of Figures

List of Tables

1. Introduction: Visual complexity in everyday life

Visual complexity can play an important role within various everyday activities. This can be illustrated by the following example of a trip to work, as it is common for many people: first of all, when leaving our home by car, we will find ourselves confronted with traffic scenarios. Meanwhile, the visual complexity of the road environment may strongly interfere with the driving task, since the complexity of road environments revealed to be positively associated with the driver's mental workload (Edquist et al., 2012). Additionally, Mace and Pollack (1983) found that the visual complexity of the road surrounding strongly affected the ability of drivers to detect and recognize traffic signs. In the inside of the car, we rely on using an instrument cluster for adapting our speed. Here, the visual complexity of this display interferes with visual search performance (Yoon, Lim & Yi, 2015). Finally, when we have arrived at the office, we probably spend most of the time working with a computer, thereby interacting with particular software interfaces (Alemerien & Magel, 2014) or browsing web pages (Deng & Poole, 2010; Tuch et al., 2011) of different visual complexity. As could be shown in previous research, the design of user interfaces and websites in terms of visual complexity can not only have an impact on the cognitive load of the user (Harper, Michailidou, & Stevens, 2009) but also affect the physiological responses (Tuch, Bargas-Avila, Opwis, & Wilhelm, 2009) as well as aesthetical appraisal (Tuch, Presslaber, Stöcklin, Opwis, & Bargas-Avila, 2012).

As could be seen in the previous example, visual complexity has an impact on persons within different contexts and affects them in multiple ways. One field where visual complexity is of special relevance is human-machine interaction. Incorporating the visual complexity of human-machine interfaces within this discipline may contribute to the creation of user-friendly products and systems that are designed for high situation-awareness (Endsley, 2016), optimized workload (Harper et al., 2009) and positive affect (Deng & Poole, 2010). This can be achieved by evaluating and subsequently adapting the design of a product or system with regard to the assessed visual complexity during the development process. As such, considering visual complexity within the user-centered design (UCD) process may have positive and direct consequences. This may not only affect the user as in the aforementioned scenarios, but in the middle and long term also reflect in the economic benefit of the manufacturer (Aquino Shluzas & Leifer, 2014).

Within the scope of this dissertation, I will investigate the effects of visual complexity within human information processing. Not only will I take a closer look at the theoretical background regarding theories and models of visual complexity as well as its influencing variables and factors. I will also focus the effects of visual complexity, for example with regard to mental workload and eye movements. Moreover, this work will also look at how computational as well as ocular measures can be helpful as indicators of and predictors for visual complexity. The gathered results of the four conducted studies will be integrated with previous findings within a research model of visual complexity. Thereby, this thesis will provide insights into the meaning and potential benefit of considering visual complexity within the context of human-machine interaction.

2. Theoretical background

Within the following paragraphs, I will provide the theoretical foundations for the empirical studies conducted within the range of this thesis. In order to introduce the reader to the topic, I will firstly begin with a rather short general perspective on complexity and simplicity before focussing on the key concept of this work, visual complexity, in more detail. Consequently, the relevance of visual complexity within human machine is examined more closely, before taking a look at relations with both computational measures as well as visual attention and eye tracking parameters. Finally, the theoretical background will conclude with the research agenda of this work.

2.1 Complexity in general

The definition of complexity is not trivial (Johnson, 2009), even though it plays an important role in many different domains of human life. Originating from the Latin word complexus, as past participle form of com- (“together”) and plectere (“to weave, braid”), complecti can mean “to entwine, encircle, compass, infold” (“complex - Wiktionary,” 2019). Collins Dictionary (2019a) suggests a definition of “the state of having many different parts connected or related to each other in a complicated way.”, which also agrees with the definitions by (Johnson, 2009) and the Cambridge Dictionary (2019). Thus, it may be summarized that on a semantic level, complexity relates both to a quantitative aspect of the existence of multiple different parts as well the aspect of the relation between these. This facet is also stressed by Standish (2008), who sees both the quantitative as well as the qualitative aspect as relevant for the definition of complexity. However, neither of both are necessary nor sufficient conditions for complexity. On the one hand, it is easy to understand that **quantity**, or being more specific, the *number of parts or objects* can be a valid measure for complexity, since for example a car would probably be seen as more complex than a bicycle, since it consists of many more parts (Standish, 2008). On the other hand, a higher number of sand grains increase the complexity of a pile of sand very little, if at all. However, the definition of complexity as a *number of distinct parts* partially avoid this issue, but raises other questions. Standish (2008) argues that in this case, a shopping list would have the same level of complexity as a Shakespearian play, since both consist of the same distinct letters. Moreover, when seen as related to a **quality**, complexity describes the amount of information that is needed to specify a system (Standish, 2008). Within regard to quality, the concept of emergence can also play an important role. According to Standish (2008, p. 117), it describes the "patterns arising out of the interactions of the components in a system" and therefore might be interpreted similar to the saying 'The whole is greater than the sum of its parts'.

Despite the ambiguity of the concept, complexity is a concept of interest for various different research disciplines such as physics (e.g. Bennett, 1990), information theory (e.g. Traub, 2003), anthropology (e.g. Kersten, 2013), computer science (e.g. Davis, Sigal, & Weyuker, 1994), sociology (e.g. Eve, Horsfall, & Lee, 2003), and economics (e.g. Durlauf, 2005).

Within the field of human-machine interaction, various kinds of complexity can have an impact (Endsley & Jones, 2012). Among these are especially the system complexity, which encompasses the overall complexity of a system that the user is dealing with. Coming back to the previous example, a car would be a more complex system than a bicycle. However, this does not necessarily cause a higher complexity for the user. For example, despite its high system complexity, a modern car is relatively easy to use. Thus, the operational complexity (Endsley & Jones, 2012) for the driver would be rather low. Referring specifically to the differentiation between system complexity and operational or observer complexity, authors have stressed that the latter is only meaningful when considered in relation to a certain observer (Casti, 1979; Edmonds, 1999). Another concept, termed apparent complexity on the other hand is related to the user's representation of the system. As such, it depends on the cognitive, display and task complexity (Endsley & Jones, 2012). While cognitive complexity (see also Rauterberg, 1996) refers to how a system works or the complexity of the logic used by the system, display complexity focusses on the aspects of how its information is presented to the user (Endsley & Jones, 2012). Aspects such as the overall and local density of items that are presented as well as their grouping and the layout complexity according to Tullis (1983) are the determinants of the display complexity. An overview of the different types of complexity and their effect on the user's mental model is depicted in Figure 1.

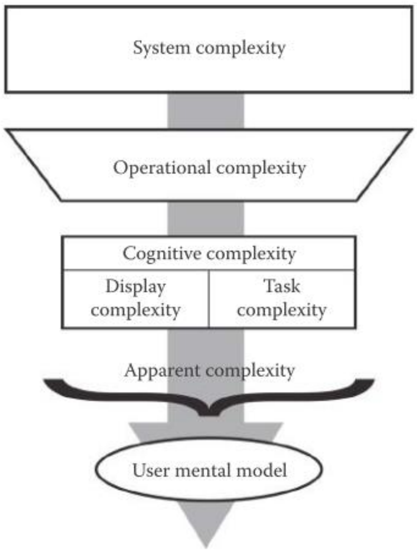


Figure 1. Layers of complexity from Endsley and Jones (2012)

Schlick, Winkelholz, Motz, Duckwitz, and Grandt (2010) for example proposed a measure for the assessment of complexity in human-computer interaction, which relies on interaction events generated by the user or computer.

Visual complexity, which this thesis will primarily focus on, has connecting points especially to the aforementioned aspects of apparent and especially display complexity. However, since not focussing on aspects of the interaction, the scope of visual complexity as a concept is also not restricted to the interaction with systems or products but can refer to any visual material, such as art for example (Leder, Belke, Oeberst, & Augustin, 2004). At the same time, it is still highly relevant within the context of human-machine interaction. Yet, in contrast to system or operational complexity, it does not include the perspective on the whole task, but takes into account only the visual presentation. Definitions and all further theoretical backgrounds of visual complexity will be addressed in the following paragraphs.

In summary, it is not possible to give a simple definition of what complexity is and which precisely describes what makes some objects or systems more complex than others. Instead, many different approaches have been taken towards complexity from various disciplines, making a comprehensive understanding of the concept a rather complex task itself. The last paragraph however provided a starting point for the further analysis and examination of complexity and subtypes of complexity such as visual complexity. The provided framework of different types of complexity pointed out their associations and should help to differentiate between these in order to facilitate a better understanding of the key concept of this work: visual complexity.

2.2 Visual complexity

In the following, visual complexity will be focused in more detail. Firstly, the concept is introduced by comparing different definitions, before taking a closer look at foundations of visual complexity research. Subsequently, I will address the typical dimensions found in literature as well as the theories and models of the construct. Finally, relations and effects between visual complexity and other variables are discussed.

2.2.1 Definitions

Commonly, visual complexity is defined as “the level of detail or intricacy contained within an image” (Forsythe, 2009, p. 158; Snodgrass & Vanderwart, 1980, p. 183). Other authors consider it from a different perspective and provide a definition of visual complexity originating from research on textures which describes it as „the degree of difficulty in providing a verbal description of an image” (Heaps & Handel, 1999, p. 301; Rao & Lohse, 1993). In some cases, additional aspect are focused such as the absence of a pattern, which especially stresses the randomness within a picture (Feldman, 2004), or the abstractness (versus concreteness) of a stimulus (García, Badre, & Stasko, 1994). Madan, Bayer, Gamer, Lonsdorf, and Sommer (2017, p. 1) moreover describe visual complexity as follows: “a picture of a few objects, colors, or structures would be less complex than a very colorful picture of many objects that is composed of several components”. Some authors even chose a very different approach by not defining visual complexity semantically but using a rather methodological approach instead, such as Tuch et al. (2009). They operationalized it as the JPEG file size due to the high correlations with subjective ratings of visual complexity, which of course allows for only minor insights regarding the semantic background of the construct.
Both Snodgrass and Vanderwart's (1980) definition as well as the one by Rao and Lohse (1993) were adopted by several researchers. While both seem to adopt quite different perspectives, they may also share some communalities. For instance, it seems likely that an image with a higher level of detail is also more difficult to describe,

since very detailed features also require a longer description. The enduring differences between definitions of visual complexity have not led to a commonly accepted solution yet. However, this issue is far from insignificant. A disparity within definitions may go along with a different understanding of the concept and thus eventually affect manipulations and measurements, which in turn result in different findings (Nadal, Munar, Marty, & Cela-Conde, 2010).

Since the definitions of visual complexity refer to the visual properties of the stimulus, this might be interpreted as if the perception of visual complexity directly emerged from these and could therefore be assessed completely objectively. This is however not the case, since perception is a constructive process that is also guided by top-down processes, which can influence the appearance of scenes (Machado et al., 2015). This has already been stressed by Berlyne (1974), who wrote that the perception of collative variables such as complexity depends both on properties of the stimulus as well as processes within the observer and can therefor differ between persons. At the same time, he suggested that the subjective ratings of complexity within many experiments has shown to vary according to the objective features of the stimulus.

Now that an overview of the definitions of visual complexity has been given, the subsequent paragraphs will present deeper insights into the foundations and more recent research findings of visual complexity. This should add more clarity regarding the theoretical background of the construct.

2.2.2 Foundations: Gestalt psychology and beyond

Visual complexity has a long history in psychology, although it has not always been explicitly been labelled as such. The roots for the investigation are often seen in Gestalt psychology, which focussed on perceptual mechanisms in general and thereby also provided the foundations for visual complexity research. Gestalt psychologists were dedicated to defining a connection between sensory input on the one and perceptual simplicity or complexity on the other hand (Donderi, 2006b). Within this regard, they defined the ability to perceive order and structure within visual stimuli as a key aspect of human perception (Koffka, 1935; Köhler, 1947). Building on this notion, Gestalt principles were postulated, which were assumed to describe processes associated with perceptual organization. One key aspect of perceptual organization is the idea of perceptual grouping. This means that “observers perceive some elements of the visual field as ‘going together’ more strongly than others” (Wagemans, Elder et al., 2012, p. 1181) and encompasses factors or principles that influence the grouping of perceptions (Wertheimer, 1923). Among these are for example the factors proximity or similarity (Wertheimer, 1923), which are described in more detail below.

A very essential one of these is the simplicity or minimum principle, which is also called the law of *Prägnanz* (Koffka, 1935; Wertheimer, 1923). This holds that perceptions tend towards simplicity and are structured into the simplest possible organization (Hatfield & Epstein, 1985; Wagemans, Feldman et al., 2012). In this regard, Koffka (1935) stated that “psychological organization will always be as ‘good’ as the prevailing conditions allow.” (p. 110). Similarly, Chater (1997) and Chater and Vitányi (2003) argued that many cognitive processes are designed to structure sensory input and find patterns within the data in order to find a most simple explanation for the data. From the various references, it can be seen that the notion of simplicity versus complexity and particularly the simplicity principle has been in the focus of research for a while, especially in the context of visual perception. This may be due to the appeal of perceptual economy. The visual system is often confronted with innumerable pieces of information with which it usually copes well. In this context, simple representations may also require fewer cognitive resources for processing (Hatfield & Epstein, 1985). Direct empirical testing of the simplicity principle is however confronted with some challenges, for example since it is still far from clear how exactly perceptual stimuli are mentally represented (Chater & Vitányi, 2003). Thus, even though evidence remains partly ambiguous, there are multiple lines of support (see Chater & Vitányi, 2003). For example, simple items are typically more easily detected in noise (Hochberg & McAlister, 1953; van der Helm & Leeuwenberg, 1996) and learned faster (Feldman, 2000).

The simplicity principle or law of *Prägnanz* may be also play a role within visual complexity. According to the aforementioned endeavour of finding patterns and structure within visual sensory input, a larger amount of perceived visual complexity may be attributed to a larger difficulty of finding patterns.

Other Gestalt factors or principles that were assumed to influence the perception of grouping of discrete elements include for example *proximity*, *similarity* and *common fate*, which encompasses that elements moving in the same direction tend to be grouped together (Wagemans, Elder et al., 2012; Wertheimer, 1923). Other factors

that are of relevance especially in more complex elements include *symmetry*, *parallelism* and *continuity* or *good continuation*. The latter describes that smooth edges more likely seen as continuous than edges with sharp angles (Wertheimer, 1923). In addition, more recent Gestalt principles including *synchrony*, *common region*, *element connectedness* and *uniform connectedness* are discussed by Wagemans, Elder et al. (2012) in more detail. Gestalt principles are still influential today and taken into account for the design of user interfaces (Chang, Dooley, & Tuovinen, 2002) for example.
Building on information-theoretical approaches such as Shannon's (1948, 1951) entropy, which was established as a way of quantifying the information amount in a variable and strongly refers to Gestalt principles, Attneave (1954) investigated visual perception from a new perspective. He considered perception as an information-handling process and emphasized the concept of redundancy within visual perception. In order to describe perceptions economically, he stated that a central function of the perceptual system was to diminish the redundancy within a stimulus. Therefore, several principles were postulated, which he assumed would reduce the necessary amount of information. For example, areas of homogeneous colour or texture could be more economically described by specifying the colour or the parameters of the texture as well as the boundaries of the area according to him. Another important role is attributed to corners and contours, which are supposed to contain a large amount of information. Accordingly, results from a later study (Attneave, 1957) showed that complexity ratings of shapes strongly depended on the number of turns (or corners) within these shapes. This is depicted in exemplary images within Figure 2.

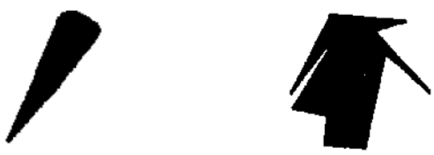


Figure 2. Shape of low (left) and high (right) complexity, from Attneave (1957)

Referring to the Gestalt principles, Attneave (1954) was convinced that many of these actually referred to information distribution and accordingly, a “*good gestalt* is a figure with some high degree of internal redundancy” (Attneave, 1954, p. 186). A similar idea

had already been represented by Musatti (1930), who suggested the principle of homogeneity, for example of colour or patterns, which could therefore be seen as a superordinate principle for example for Wertheimer's (1923) gestalt principles.
A foundation for the experimental investigation of visual complexity focussing particularly on objects was later provided by Snodgrass and Vanderwart (1980), who established a picture set with black-and-white line drawings and ratings of visual complexity (see for example Figure 3).

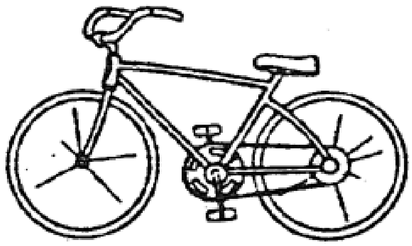


Figure 3. Example from Snodgrass and Vanderwart's (1980) picture set

Further researchers were also dedicated to work on stimuli for the investigation of visual perception, such as Rossion and Pourtois (2004), who added both a grey-level and a coloured version to Snodgrass and Vanderwart's (1980) original set. Within more recent research, authors often used photos of single objects (Brodeur, Guérard, & Bouras, 2014; Moreno-Martínez & Montoro, 2012; Paré & Cree, 2009) or scenes (Bradley, Hamby, Löw, & Lang, 2007; Bradley, Houbova, Miccoli, Costa, & Lang, 2011). Eventually, controlled picture sets with standardized ratings are essential for the investigation of visual complexity. These play a central role within research on the construct, since variations within the stimuli may also have a large impact on the findings (Nadal et al., 2010). After focussing on the rather historical foundations of visual complexity research, within the next paragraph, findings regarding the dimensional structure of the construct visual complexity will be discussed in detail.

2.2.3 Dimensions of visual complexity

As described by Standish (2008) (see also paragraph 2.1), two major aspects are seen as relevant for general complexity, both a quantitative as well as a qualitative aspect.

Similarly, researchers with a focus on visual complexity support the view that this construct is not a unidimensional one either. Instead, it is assumed to consist of both a quantitative dimension, which is related to the amount of elements and a qualitative or structural dimension, which is determined by the structural organisation of an image (e.g. Gartus & Leder, 2017). In the context of visual complexity, this goes back to Chipman (1977), who studied the determinants of complexity ratings for visual patterns and found that these could be grouped into the two types of features. From various experiments, she drew the conclusion that quantitative aspects such as the number of turns or corners set an upper bound on visual complexity while structural features such as symmetry subsequently reduces the perceived complexity. This notion was further supported by Ichikawa (1985), who argued that two separate cognitive processes are involved in complexity perception. According to him, a fast process is responsible for the evaluation of quantitative aspects in a stimulus, while another rather slow process is responsible for the detection of structure. These findings are based on experiments with dot patterns presented with different durations and further support the idea of seeing visual complexity as a two-dimensional construct. In the following, the dimensions and findings on the associated influencing factors are subsequently described in detail.

Quantitative Dimension

The quantitative dimension is often seen as the most influential dimension of visual complexity (e.g. Nadal et al., 2010). Many researchers have described quantitative aspects as relevant for visual complexity, going back to Attneave (1954). His assumption was that information within an image was largely concentrated at points of contour change. Thereby, drawings with a number of dots, which should indicate the outline of an object, were used in order to study perception. Building on this idea, the number of turns (or “points”, “angles”, “sides”, p. 222) were then identified as an important determinant of the perceived complexity of shapes (Attneave, 1957). Similarly referring to the complexity of forms, Arnoult (1960) showed that the number of independent sides of random polygon forms was a good predictor for the rating of visual complexity. Moreover, Thomas (1968) suggested, according to his findings, the number of angles within a polygon as the most important determinant of visual complexity.

Using visual patterns as stimuli, Berlyne, Ogilvie, and Parham (1968) identified information content as an important aspect of visual complexity, comprising for example the amount of elements within a picture. This aspect appears also in more recent literature, such as a work by Nadal et al. (2010), using various stimuli such as artworks or in Oliva, Mack, Shrestha, and Peeper (2004), who found that the quantity of objects was often mentioned as a criterion for the judgement of visual complexity within a hierarchical grouping task that subjects performed with pictures of indoor scenes.

Research focussing on user interfaces and websites revealed similar results regarding the influence of quantitative aspects on perceived complexity. As one of the earlier researchers within this area, Tullis (1983) identified overall density, denoting the "number of characters displayed, often expressed as a percentage of the total character spaces available" (p.662) as a central aspect of displays. Within more recent research, specifically visual complexity has been further related to quantitative aspects such as the number of graphics, links, and the home page size (Geissler, Zinkhan, & Watson, 2006), the amount and density of elements such as text, links and images (Harper et al., 2009; Michailidou, Harper, & Bechhofer, 2008) or, more generally, the amount of information (Miniukovich & Angeli, 2014; Miniukovich, Sulpizio, & Angeli, 2018). Similarly, Deng and Poole (2010) suggested visual richness or the "detail of information in a website, such as amount of text, number of graphics, links and layout" as one of the two main dimensions of website complexity.

In conclusion, various previous research works based on findings with stimuli such as basic patterns, polygons or displays and websites underline the relevance of quantitative aspects for the perception of visual complexity.

Qualitative / Structural Dimension

In addition to the quantitative dimension, many authors suggest structural or qualitative aspects as further central features of visual complexity. Organization or disorganization (Miniukovich et al., 2018; Miniukovich & Angeli, 2014; Oliva et al., 2004) and especially symmetry (Chipman, 1977; Day, 1968; Gartus & Leder, 2017; Nadal et al., 2010; Oliva et al., 2004; Riglis, 1998) are often mentioned as key aspects for the structural dimension within different kinds of pictorial stimuli. This can also be found for user interfaces and websites, where symmetry likewise plays an important role regarding the perception of visual complexity (Miniukovich et al., 2018; Miniukovich & Angeli, 2014; Tuch, Bargas-Avila, & Opwis, 2010). The special role of symmetry might not be very surprising since various findings indicate that the human visual system is extremely efficient in extracting symmetry in visual stimuli (Treder, 2010), even preattentively within very short presentation durations of 150 milliseconds and less (e.g. Wagemans, 1995). Thereby, mirror symmetry is usually more salient than other forms of symmetry such as translational or rotational symmetry (Wagemans, 1995). Within mirror symmetries, a vertical symmetry axis is particularly easy to detect compared to horizontal or diagonal symmetries with elements closer to the symmetry axis generally being more important than distant ones (Gartus & Leder, 2017). These findings are also considered in Bauerly and Liu's (2008) formula for calculating a measure of symmetry for example. Furthermore, symmetry effects can identified within neural correlates, for example using event-related potentials or functional magnetic resonance imaging (fMRI) on visual areas (Bertamini & Makin, 2014).

Another relevant structural aspect might be visual perceptual balance (Hübner & Fillinger, 2016; Lok, Feiner, & Ngai, 2004), which to the best of my knowledge has not been considered within the context of visual complexity yet. This designates “how well the elements in a picture are arranged” (Hübner & Fillinger, 2016, p. 1) and is popular within visual arts as well as research on aesthetic appraisal. Measures for perceptual balance such as the assessment of preference for balance (APB, Wilson & Chatterjee, 2005) or the deviation of centre of mass (DCM, Hübner & Fillinger, 2016) usually assume that the “mass” of dark (e.g. black) pixels in an image is higher than for bright (for example white) ones. Although it has been shown that these measures correlate with aesthetical preference (Hübner & Fillinger, 2016), the influence of perceptual balance on the perceived visual complexity has not yet been investigated. Based on the effects on aesthetics and the Gestalt principle of *Prägnanz* (Koffka, 1935; Wertheimer, 1923), it could however be hypothesized that perceptual balance might also affect visual complexity. Similar to symmetry, it could facilitate perceptual grouping and thus reduce the demand for cognitive resources since the identification of patterns may be easier.

In sum, the research findings suggest that structural or qualitative aspects such as organisation and especially symmetry are negatively associated with the perception of visual complexity.

Other dimensions

Next to quantity and structure, further aspects have been found to affect the perception of visual complexity. In particular, the variety or diversity of elements in an image or interface (Deng & Poole, 2010; Harper et al., 2009; Heylighen, 1997; Miniukovich et al., 2018; Miniukovich & Angeli, 2014; Nadal et al., 2010) or their similarity (Riglis, 1998) appear to be of special relevance. The influence yet seems to be rather small compared to the quantitative and qualitative dimension (Nadal et al., 2010). However, in many scenarios the variety or similarity of elements can be related to the number of objects, where a larger number of elements likely goes along with a larger variety (as for example in Oliva et al., 2004).

Additionally, colour is often mentioned in the context of visual complexity within literature. In particular, the number or variety of colours was found to be related to visual complexity (Nadal et al., 2010; Oliva et al., 2004), however in some findings the influence of colours was rather small (Nadal et al., 2010) or did not represent an important aspect of visual complexity (Hall, 1969). These findings are in line with those of Rossion and Pourtois (2004). In their stimulus set containing object drawings in black and white, grey levels and colour, no significant difference in visual complexity ratings was found between the three categories. Similarly, Ciocca, Corchs, Gasparini, Bricolo, and Tebano (2015) could not show an influence of colour on complexity ratings either. Regarding websites however, Reinecke et al. (2013) demonstrated an influence of colourfulness on perceived visual complexity.

Further aspects related to visual complexity encompass for example the unintelligibility of the elements, which means the difficulty to identify the elements in the image, three-dimensional appearance (Nadal et al., 2010), the clutter and open space within an image (Oliva et al., 2004) as well as the familiarity with a visual stimulus (Riglis, 1998). Effects of these factors were however only examined within individual studies, which does not yet allow robust general conclusions to be drawn.

In summary, the findings regarding the influence of other possible variables such as variety and colour on visual complexity are rather unambiguous.

Concluding the previous findings on influencing factors of visual complexity, it can be stated that findings from various domains stress the relevance of both a quantitative and a qualitative or structural dimension. With regard to other aspects such as variety of elements or colour, findings are less clear. In general, very few of these studies relied on experimental methodology in order to directly investigate the effect of certain factors on visual complexity ratings but instead are based on correlational analyses. Therefore, many of the findings might be confounded by other features of the stimulus material. All in all, the dimensional structure of the construct visual complexity has not been consistently validated yet. While the findings of Ichikawa (1985) and Chipman (1977) strongly suggest a two-dimensional construct, Nadal et al. (2010) identified three factors with disorganization and symmetry loading on different factors. A different structure with three or four dimensions was proposed by Miniukovich and Angeli (2014) and Miniukovich et al. (2018) for graphical user interfaces. Thus, a conclusive investigation regarding the structure of visual complexity still remains to be realized. In sum, the existing research literature provides a solid theoretical ground regarding the dimensions of visual complexity while many issues still remain unresolved.

2.2.4 Theories and models concerning visual complexity

To this date, there is no comprehensive model of visual complexity that integrates the majority of relevant findings from current research. In particular, this holds true when focusing on aspects of human-machine interaction. However, some theories and models from different domains have integrated visual complexity within frameworks of information processing. This can still be of great interest for an overview of the various cognitive processes which visual complexity is intertwined with. One model, which deals with this aspect respecting the perception of art is the information-processing model by Leder et al. (2004), which is depicted in Figure 4. The authors stress the role of visual complexity for the processing of art within the first step, which they call perceptual analysis. Within the first stage, visual complexity is assumed to be processed next to other, rather basic perceptual variables such as order or symmetry (which are often also seen as influencing variables of visual complexity as described within the previous paragraph). The second step encompasses the implicit memory integration of the input from the perceptual analysis with the previous experience for example concerning familiarity and prototypicality. Further steps finally contribute to formation of an aesthetic judgement and the aesthetic emotion that the artwork elicits in the observer according to the model. In conclusion, according to Leder et al.'s (2004) Information-processing model, visual complexity is analyzed within a very early stage, before the integration with memory or evaluation processes occur.

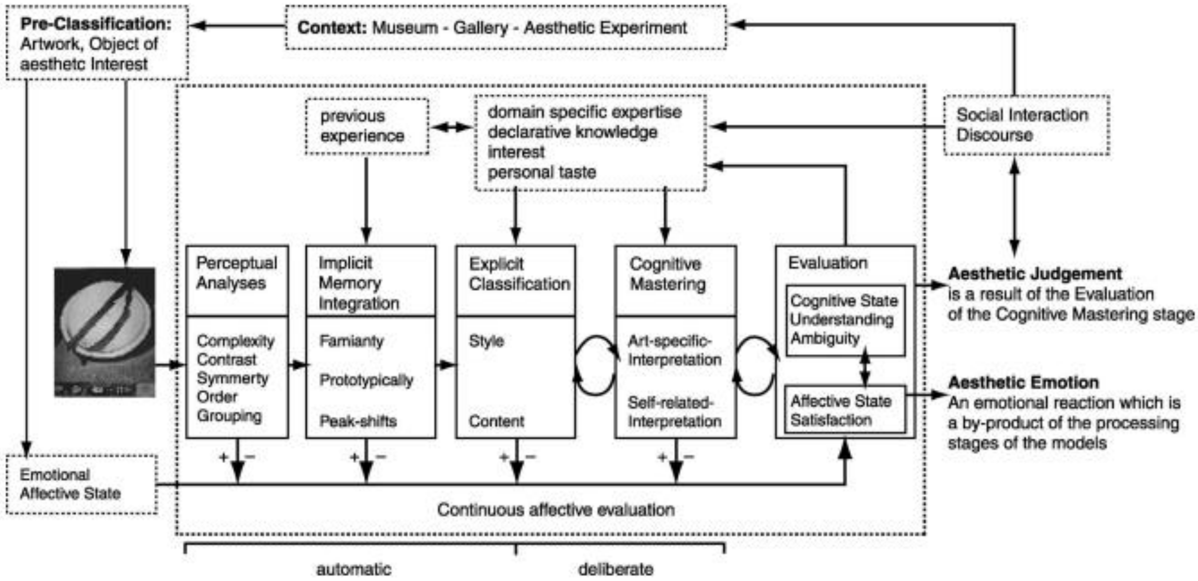


Figure 4. Information-processing model, taken from Leder et al. (2004)

Other models address the role of visual complexity within the domain of webpages. Deng and Poole (2010) proposed a research model in order to illuminate the relationship between visual complexity and design features of a webpage with regard to emotional responses and finally the approach-avoidance behaviour of users towards a website. Their model, which is depicted in Figure 5, builds on the framework of the environmental psychology model by Mehrabian and Russell (1974), the M-R model. This states that emotions mediate the effects of environmental stimuli on behaviour. In order to apply the model to the interaction with webpages, it is extended by considering findings from human-computer interaction, emotion and further relevant research disciplines in order to account for the approach-avoidance behaviour of users towards webpages. Within this model, visual complexity, which is considered separately from order, was shown to positively affect arousal and pleasantness, although the direction of the latter was partly different than expected by the authors. The metamotivational state of the user (goal-oriented vs. enjoyment-seeking) was shown to modulate the influence of visual complexity on pleasantness. Both emotional responses, arousal and pleasantness, were then shown to affect the approach-avoidance behaviour towards a webpage. All in all, empirical data support the research model as well as the suggested effects of visual complexity.

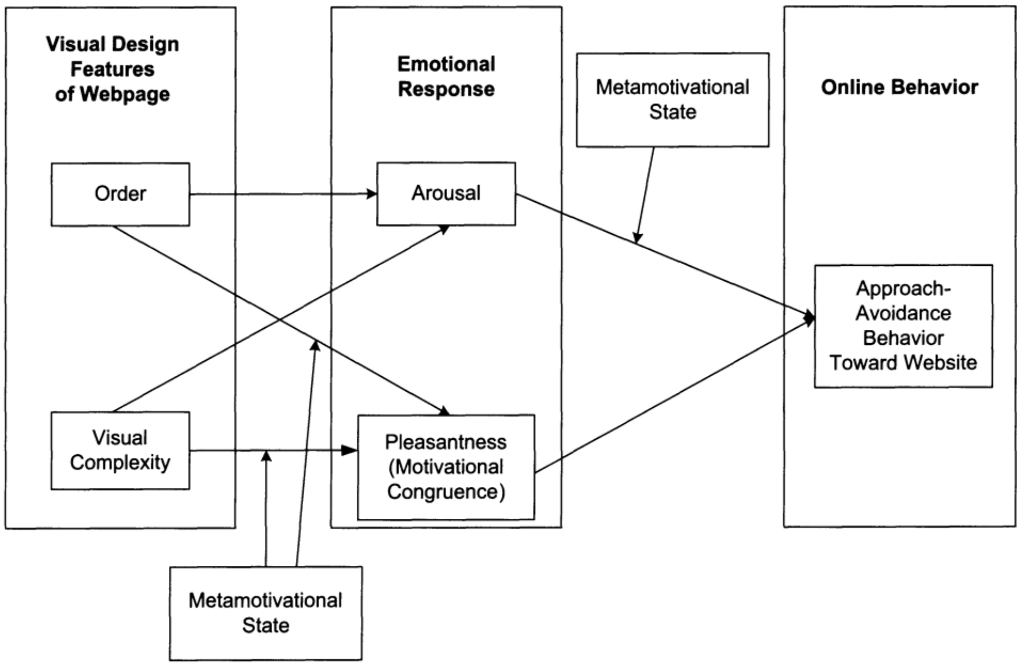


Figure 5. Research model of website visual complexity taken from Deng and Poole (2010)

Beyond the model of Deng and Poole (2010), Nadkarni and Gupta (2007) proposed a theoretical model of perceived website complexity (see Figure 6). Although visual complexity is not explicitly mentioned, this model is interesting primarily because it allows insights into the interplay between objective and perceived website complexity and user satisfaction. The authors define objective website complexity as “the number and configuration of information cues in the stimulus itself” (Nadkarni & Gupta, 2007, p. 503), thus it may be seen as related to the concept of visual complexity, although it is not necessarily restricted to a single page of a website. Results of their studies suggest that the positive relationship between objective complexity and perceived website complexity is moderated by user familiarity. This means that users with high familiarity experienced lower perceived complexity for a website of a certain objective complexity than users with low familiarity. This finding may also be transferred to other applied contexts, for example when users are already used to a certain software interface. Moreover, the authors found that perceived website complexity was related to user satisfaction and the shape of this relation depended on the user’s online task goals.

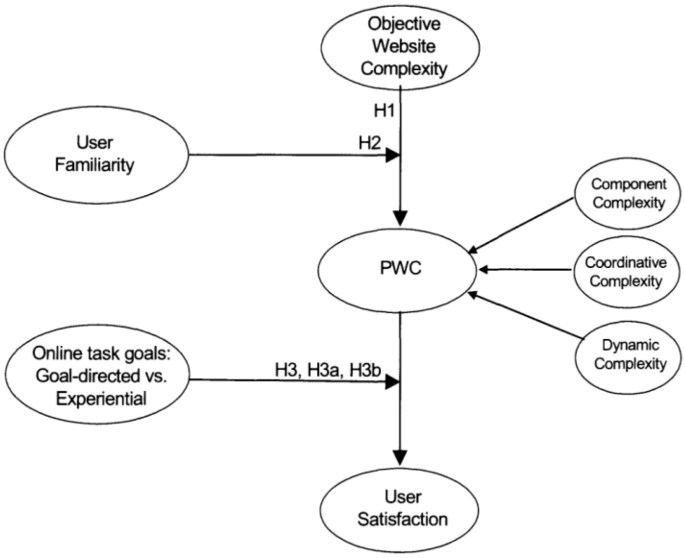


Figure 6. Theoretical model of perceived website complexity from Nadkarni and Gupta (2007)

In conclusion, the three models that were presented provide a better understanding of the relations and the interplay between (visual) complexity and other constructs. Furthermore, they reveal insights into the role of visual complexity in perception and information processing within different domains.

Next to the rather cognitive approaches described before, visual complexity was however also considered from an information-theoretical perspective. In this regard, Donderi (2006a) for example suggested that Shannon's (1948) information theory can also be applied to visual images. This means that images can be treated as messages whose complexity, which according to him is equivalent to their information content, can be measured by their compressed file sizes. This is consistent with the findings that compressed file sizes are correlated with ratings of visual complexity (e.g. Donderi & McFadden, 2005; Tuch et al., 2009). Next to the information-theoretical base by Shannon (1948), this can also be explained by the ideas of Kolmogorov complexity (Li & Vitányi, 2008) within the algorithmic information theory (AIT) (Chaitin, 1977). These incorporate probability and information theoretical ideas as well as philosophical notions of randomness. The basic idea of Kolmogorov complexity as a part of AIT is that the length of the shortest description of an object is a measure of this object's complexity. If there is a very short description of this object, it is less complex than an object, which requires very long descriptions (Li & Vitányi, 2008). This also implies that the description may consist of a computer program or script that produces the object, thus the Kolmogorov complexity is a measure of the computational resources necessary for the description. Within this context, compression is an important aspect. Some scripts can be strongly compressed, when there is enough regularity within the object they describe. Others however can hardly be compressed. This is where another important aspect of AIT and Kolmogorov complexity comes in: randomness. According to AIT, an absolutely random sequence is most complex, since it cannot be compressed. Therefore, its description would be very long. By contrast, a sequence with little randomness and high regularity can easily be compressed and would thus be simpler (Li & Vitányi, 2008). Using a similar approach as Kolmogorov complexity, Leeuwenberg (1968, 1969) introduced a measure for their complexity where each pattern could be described by a code, which he called structural information. The code's length, which relates to the number and regularity of certain operations, could be used as a measure for pattern complexity (see for example Figure 7). Within psychology, the according structural information theory (SIT) developed independently but in parallel with AIT, focussing specifically on human visual perception (Machado et al., 2015). While both AIT and SIT share the basic idea of using description length as an indicator of complexity, there are differences which are mostly related to the perceptual focus of SIT. For example, SIT differentiates between metrical and structural information which AIT does not. Moreover, SIT focusses only on perceptually relevant regularities while AIT incorporates any possible regularity (Machado et al., 2015).

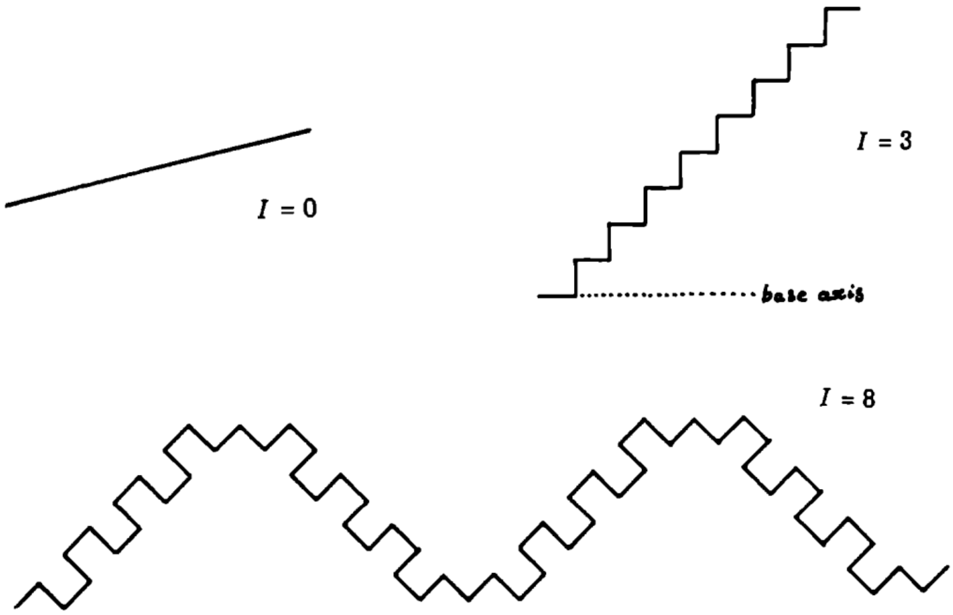


Figure 7. Examples for measure of structural information from Leeuwenberg (1968)

The basics of AIT and SIT are of special relevance with regard to the visual complexity of pictures. They laid the foundations for investigating the construct by using computational measures, such as compression methods like JPEG or GIF, in order to draw conclusions about the visual complexity of pictures. In paragraph 2.4, I will give a more detailed overview about the different methods that were developed on this basis.

2.2.5 Relations between visual complexity and other constructs

Why is it worth to take a closer look at visual complexity? The construct is of relevance for various different domains. In the following paragraph, I will focus on the relations between visual complexity and other constructs, which can play a role within and beyond human-machine interaction, such as aesthetical preference, familiarity, interest and physiological responses. Since mental workload is especially relevant within the context of human-machine interaction, this important aspect and its association with visual complexity is considered closely among aspects of human-machine interaction within paragraph 2.3.1.

2.2.5.1 Arousal and physiological responses

Complexity perception is often associated with arousal, which may encompass both subjective ratings, for example by means of the self-assessment manikin (SAM; Bradley & Lang, 1994) or be detected by physiological measures such as electrocardiogram (ECG) or electrodermal activity (EDA). One of the first researchers who took a closer look at the relation between complexity and arousal was Berlyne. According to his findings, complexity or specifically irregularity of patterns next to novelty and other stimulus properties affected arousal, manifesting for example in larger galvanic skin responses (GSRs) (Berlyne, Craw, Salapatek, & Lewis, 1963). This effect was later described as the *arousal potential* or the "'psychological strength' of a stimulus pattern, the degree to which it can take over control of behaviour and overcome the claims of competing stimuli" (Berlyne, 1971, p. 70). According to this notion, more complex stimuli are less easily recognized and associated with a response through learning. Therefore, they possess a higher arousal potential, so they are more likely to raise arousal. According to Berlyne (1971), further variables such as the smoothness of curves or the contrasts in brightness can also contribute to the arousal potential of a stimulus.

These findings have been supported several times. For example, Marin and Leder (2013) demonstrated strong positive correlations between perceived visual complexity and ratings of arousal within different kinds of stimuli such as environmental scenes and representational paintings. However, these associations appeared to become smaller when controlling for the influence of familiarity. In the context of user interfaces and websites, there is additional support for these findings. For example, Tuch et al. (2011) similarly showed significant correlations between the perceived visual complexity of websites and arousal ratings. Moreover, Tuch et al. (2009) and Tuch et al. (2011) identified effects using multiple physiological measurement methods such as ECG, electromyography and electrooculography. Tuch et al. (2009) for example found a significant correlation between visual complexity and change of the electrocardial interbeat interval as well as facial muscle tension. Additionally, Tuch et al. (2011) found a significantly larger heart rate decrease after stimulus onset for more complex webpages compared to less complex sites. Although the direction of the findings may appear surprising, they are in line with earlier findings with more basic stimuli (e.g. Fredrikson & Ohman, 1979), who similarly found a larger decrease of heart rate for

more complex stimuli. This may be due to an orienting response, which is more pronounced for complex stimuli that contain a larger amount of information.
Madan et al. (2017) furthermore revealed a strong relation between ratings of visual complexity and arousal, which the authors referred to as an *arousal-complexity bias*. They however argue that arousal (for example induced by emotionally arousing pictures) may also affect the perception of visual complexity. This could be due to effects of arousal on visual processing, which could thus influence complexity ratings.
In sum, an interrelation between visual complexity and arousal has been shown in several studies. Following Berlyne's approach (e.g. 1963) and supported by other researchers, it seems very likely that larger visual complexity in stimuli induces a higher level of arousal. On the other hand, according to Madan et al. (2017), arousing stimuli may also be rated as more complex, although a broader literature base exists for the former direction of the relation.

2.2.5.2 Aesthetical preference

Next to and partly associated with arousal, aesthetical preference is one construct that is often considered in the context of visual complexity. Many researchers have already investigated the relation between the two constructs as well as its shape.
Again, Berlyne (1971, 1974) was one of the first researchers who focussed on this issue in his psychobiological theory. Within this, he refers to the Wundt curve depicted in Figure 8, which has a long history in psychology, going back to Wundt (1874). As suggested by Berlyne (1971), it is based on a summation of activations of the aversion and the reward system, which goes along with a certain arousal potential. Above the threshold where stimuli are noticed, stimuli according to this approach are perceived as increasingly pleasant with larger arousal potential, until reaching a peak at a medium level. Beyond this peak, when arousal is further increased, hedonic value will decrease. Thus, it suggests that a medium level of arousal is preferred as can be seen by the inverted u-shape of the curve. According to Berlyne's (1970) findings, complexity next to novelty and other variables affects arousal (as also described in the previous paragraph), which again influences the hedonic value.

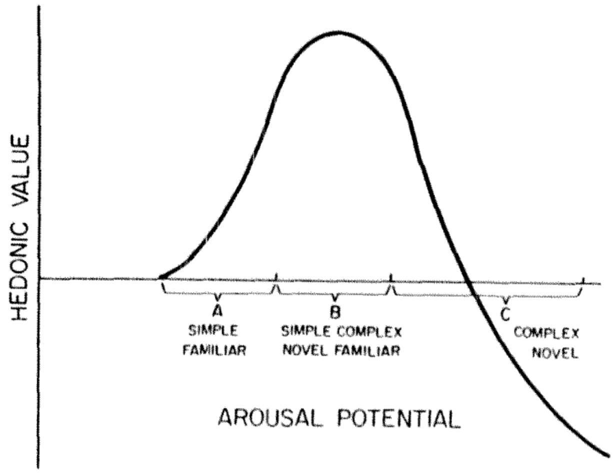


Figure 8. The Wundt curve (from Berlyne, 1971)

These findings and theory have subsequently been supported by a number of researchers (e.g. Farley & Weinstock, 1980; Imamoglu, 2000; Saklofske, 1975; Vitz, 1966). Within the field of human-computer interaction, the inverted U-shape was also partly underlined (Chassy, Lindell, Jones, & Paramei, 2015; Geissler et al., 2006; Güçlütürk, Jacobs, & van Lier, 2016). Other findings were however not in line with this shape of the relation. These instead suggested either a negative linear relation between complexity and preference (e.g. Marin & Leder, 2013), which could also be found using websites as stimuli (Michailidou et al., 2008; Reinecke et al., 2013; Tuch et al., 2009; Tuch et al., 2011; Tuch et al., 2012). Moreover, a generally positive linear relation with aesthetical appraisal was suggested (Nadal et al., 2010) particularly when visual complexity was operationalized by the number or variety of elements while other studies found no relation at all (Pandir & Knight, 2006). The difference in results may be explained by differences in the stimuli that were used or the fact that only a part of the whole complexity range could be depicted with these. For example, it might be that stimuli represented only the lower or the upper part, thus a positive respectively negative linear relation was found instead of the inverted U-shape, which could have been depicted if the whole variation of complexity had been included within the stimuli, as also argued by Tuch et al. (2012). Another option is of course, that the inverted U-shape is not applicable to all types of stimuli but that linear relations instead describe the relation between visual complexity and aesthetical appraisal or pleasure better.

2.2.5.3 Other factors: Familiarity, Interest, Prototypicality

Familiarity

The familiarity of stimuli can also significantly influence the perception of visual complexity. Snodgrass and Vanderwart (1980) define familiarity as „the degree to which you come in contact with or think about the concept“ (p. 183). Within their study, subjects were thus instructed to rate how usual or unusual something is within their life. They found a significant negative correlation of -.466 between visual complexity and familiarity. According to the authors, this may either be due to the style of drawing, so that complex drawing may appear more novel than simple drawings. Another explanation may be tied to the complexity of the object. More familiar objects, which are more present within the everyday life, may thus be perceived as less visually complex (Snodgrass & Vanderwart, 1980). Similar relations between familiarity and complexity were also shown by other authors (Alario & Ferrand, 1999; McDougall, Curry, & Bruijn, 1999). In this context, Forsythe, Mulhern, and Sawey (2008) point out that norms and results focussing on visual complexity may be biased by familiarity due to the high negative correlations such as for example found in Snodgrass and Vanderwart's (1980) picture set, which they call a *familiarity interference effect*. Within a controlled study, Forsythe et al. (2008) investigated the influence of learning and familiarity on complexity ratings and discovered that participants who were trained with a number of images perceived these as less complex than untrained participants. Additionally, both training and familiarity influenced the rating of visual complexity for nonsense shapes within a significant interaction.

Prototypicality

Next to familiarity, prototypicality was shown to affect different aspects of visual perception (e.g. Kayaert, Beeck, & Wagemans, 2011) and may thus also have an impact on the perception of visual complexity. Prototypicality is often defined as “the amount to which an object is representative of a class of objects’’ (Leder et al., 2004, p. 496). With regard to websites for example, Roth, Schmutz, Pauwels, Bargas-Avila, and Opwis (2010) investigated the mental models of users and found that these determined to a large degree where they expected certain objects for a specific type of website. For example, most users expected the navigation area on the left side for company, news and shopping sites. If these elements are not found at the expected locations, this may contradict the mental model of the user. Thus, it might be assumed that the page is perceived as more visually complex. Moreover, Tuch et al. (2012) found effects of prototypicality on aesthetical judgements about websites. While to the best of my knowledge no direct investigations exist on the association between prototypicality and visual complexity to this date, it could thus be hypothesized that the prototypicality of the arrangement of elements may also affect visual complexity. This might particularly apply to types of stimuli that users are familiar with and thus have a mental model for. This will be investigated more closely within study 4.

Interestingness

Interestingness is another construct that may relate to the perception of visual complexity. Within earlier investigations, Berlyne and Boudewijns (1971) showed a positive relation with a levelling off at high complexity levels. The shape of this relation thus differs from the one found for pleasingness and liking, which showed a decline for higher complexity ratings. Findings from their work are depicted in Figure 9.

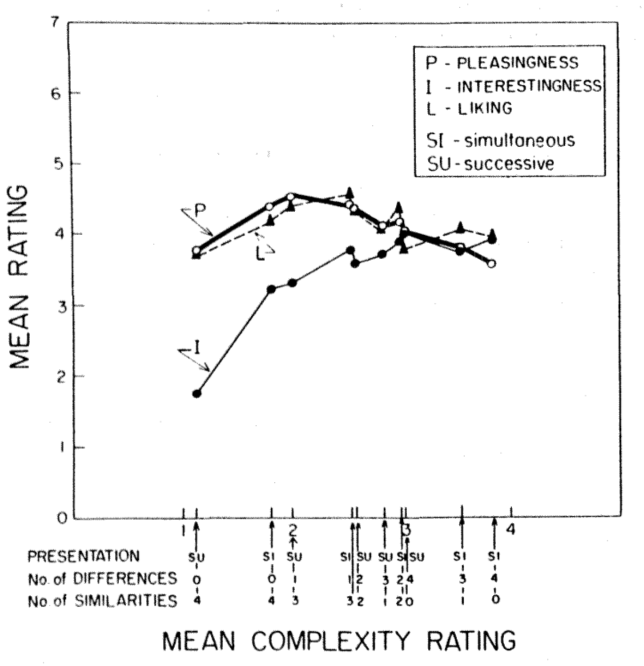


Figure 9. Findings by Berlyne and Boudewijns (1971) on the relation between complexity and interestingness

Similar results were also revealed by Aitken (1974) and Day (1968), who found that interestingness for random polygons used as stimuli increased with visual complexity until peaking at high complexity levels. In sum, these findings suggest a positive relation between visual complexity and interestingness.

All in all, within this paragraph it could be shown that a number of constructs are associated with visual complexity. This overview is far from complete, since depending on the domain, many further concepts may also play a role. However, it provides an idea of the relevance of the construct visual complexity as well as the interrelatedness of different constructs within the processing of visual information.

2.3 Visual Complexity in human-machine interaction

According to the German Federal Institute for Occupational Safety and Health, for 80 percent of all office workers, a visual display unit (VDU) such as a desktop PC, laptop or mobile device is the most important work equipment (BAuA, 2019). The visual interaction with software and user interfaces is thus a key aspect of work for many persons. Work may consist of many different tasks and differ with regard to the organisational and environmental circumstances, which may lead to different demands towards the user. Tasks can be for example encompass monitoring or surveillance tasks with CCTV systems, which are common in control rooms (Pikaar, Lenior, Schreibers, & Bruijn, 2015) but also consist in the use of specific software or standard programs as well as the interaction with websites.

However, since a big part of the interactions described above relies on graphical information displays, visual perception processes are of great relevance. Therefore, particularly visual complexity can have an impact for many persons within different scenarios of work and human-machine interaction. Within the next paragraphs, the relevance of visual complexity in this context will be considered in detail. A special focus will be on the key concept of mental workload as well as the implications for user interface design.

2.3.1 Mental workload and visual complexity

Mental workload can play an important role within work in general and particularly within human-machine interaction. For the user-centered design of human-machine systems and their optimization with regard to demands towards the user, particularly within the context of increasing automation, the consideration of the concept is central

(Manzey, 1998). Mental workload refers to the processing of information, making of decisions and the demands imposed on mental resources by these tasks (Moray, 2013), while it is mainly characterised by an intensity aspect (Kahneman, 1973). It can thus be seen as “difference between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time” (Gopher & Donchin, 1986, p. 41). It differs from physical and muscular workload by its focus on cognitive processes, however it is not always easy to distinguish from emotional load (Manzey, 1998). It is important to differentiate between strains on the one and stresses on the other hand. While the former describe external factors that act upon a person such as work tasks and environmental aspects, the latter also depend on individual properties and abilities of a person (Rohmert, 1984). Consequently, effects of external stresses can produce a different amount of mental workload for different persons.

There are several arguments for hypothesizing that there is a link between the two constructs visual complexity and mental workload, even though the actual body of research is still far from sufficient for a final conclusion. First of all, based on the interpretation of their results, Harper et al. (2009) suggest in the context of websites that visual complexity is implicitly linked to the perception of cognitive complexity. Consequently, they propose that visual complexity might serve as an implicit measure of cognitive load, although this has not yet been backed by empirical results. This notion might also reflect in the concept of visual load or the related perceptual load (Lavie, 1995; Lavie & Tsal, 1994). While not explicitly defined, the authors describe it as related to the processing demands of a stimulus. This already indicates the relation between perceptual stimulus features and cognitive processes and demands, which can also affect the selection in visual attention (Lavie & Tsal, 1994). While the terms visual or perceptual load have not been widely adopted in human-machine interaction, Perrott, Sadralodabai, Saberi, and Strybel (1991) for example referred to the former in order to describe that within a visual search task the number of distractors positively affected search latencies. Pierno, Caria, Glover, and Castiello (2005) moreover used a secondary visual task in order to induce and increase visual load within a virtual environment, which increased the time that was necessary to locate a visual target. These findings may partly relate to more basic aspects of information processing. Alvarez and Cavanagh (2004) in this regard for example showed that the capacity of the visual short-term memory varies both according to the number of objects presented as well as their visual information load, with an upper limit of four to five objects. This limit of processing capacity for example also appears in Miller's (1956) magical number seven. He found that the short-term memory of most people can only hold up to 7 ± 2 information chunks. This of course implies that by grouping or ‘chunking’ of information into units, it is possible to remember more information. The finding of a limited capacity of the visual working memory has also been supported by more recent findings (Rouder et al., 2008), where the participant’s task was to remember squares of different colours. A summary of multiple contemporary studies however suggests that the capacity limit may rather consist of three to five instead of seven chunks (Cowan, 2001). With regard to visual complexity, it can hence be hypothesized that more visually complex stimuli also demand more of the limited capacities within the visual working memory, for example because these contain more elements or information chunks, which can then contribute to an increased mental workload.

This relation between (limited) information processing capacities and workload is also addressed in more detail from a human factors perspective within Wickens' (1984, 2008) multiple resource theory (MRT). As the aforementioned theories, it suggests that the processing capacities are limited, however it states that there are multiple pools of resources for modalities, stages of processing and responses (see Figure 10).

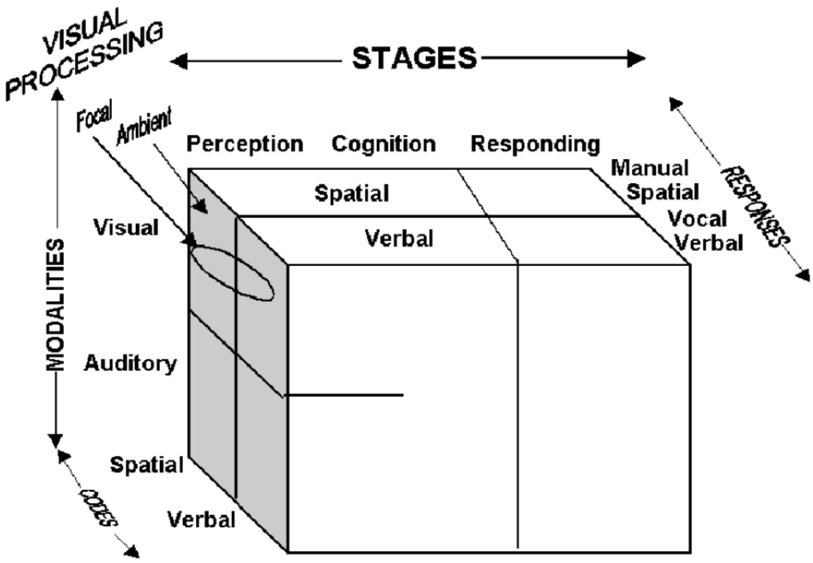


Figure 10. Multiple resource theory, taken from Wickens (2008)

If tasks performed by an individual produce high demand for resources within the same regions or pools, this may cause an increase in workload and finally produce errors or a decrease of performance (Wickens, 2002). With regard to visual complexity, it could be argued in line with MRT that visually complex stimuli use up more resources within the visual modality than simple stimuli. Since these resources are not available any-more, this can contribute to an increased mental workload, particularly in the case of multiple tasks with additional visual demands.

Several findings from applied human-machine contexts are in line with the assumption of a relation between visual complexity and mental workload. Primarily, findings from a small number of driving studies have addressed this question, focussing especially on the visual complexity of the road environment. For example, Edquist et al. (2012) revealed by means of a driving simulator study that within less visually complex road environments, participants' speed was closer to the limit and less variable than in more complex environments. Moreover, they found that mental workload was rated higher for visually complex environments. Visual complexity was operationalised by the road-side environment as well as the amount of on-street parking. Similarly, Horberry, An-derson, Regan, Triggs, and Brown (2006) found that especially old drivers drove more slowly within complex road environments with a high number of billboards, advertise-ments, buildings and oncoming vehicles while perceived workload rating was not af-fected. Engström, Johansson, and Östlund (2005) on the other hand more specifically investigated the effects of visual and cognitive load within a driving simulation study. They found that a visually demanding secondary task negatively affected speed and lane keeping in contrast to a cognitive load condition, operationalized by an auditory task. This effect is assumed to be related to the sharing of resources between the driving and the secondary task and can thus be related to workload.

Referring to Wickens' (1984, 2008) multiple resource theory, Verwey (2000) similarly assumed that both the number and complexity of visual information sources is associ-ated with visual workload. Moreover, he states that visual workload in driving is usually high, which may relate to the fact that drivers can rarely take their eyes off the road. Consequently, he found that road situation, which may be of different complexity, is an important determinant of both visual and also mental workload of drivers as measured by secondary task performance.

Next to the driving context, some indicators for a relation between visual complexity and mental workload also come from the area of aviation. This can encompass both the viewpoint of pilots but also be relevant for the ground crew, for example within the air traffic control (ATC). Within the former, Svensson et al. (1997) for example showed that information complexity on the tactical situation display (TSD) of a flight simulator affected both pilots' mental workload as well as their flight performance, for example regarding the correct altitude, as depicted in Figure 11. Moreover, complexity, which was operationalised by the number of objects presented in the TSD, had an effect on information handling as well as physiological measures such as heartrate.

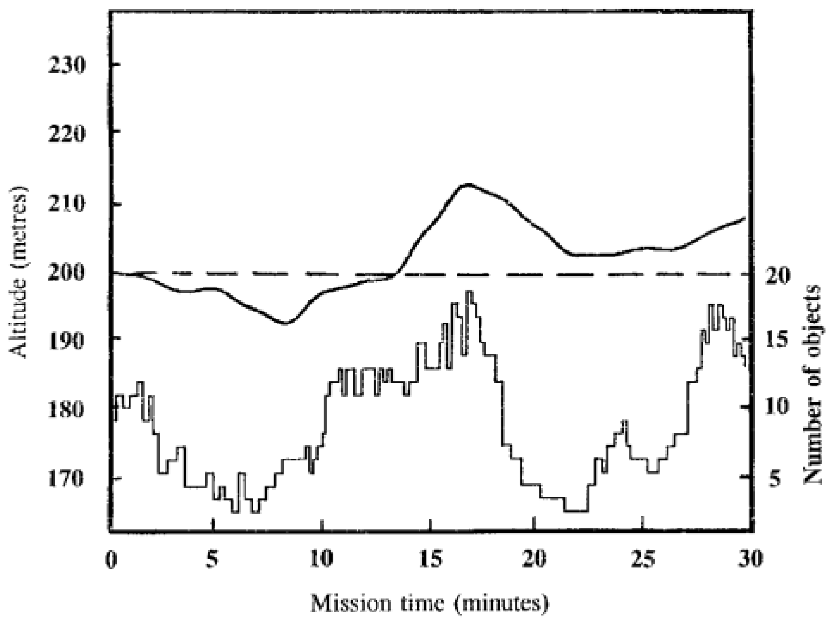


Figure 11. Complexity (operationalised by number of objects) and deviation of prescribed altitude of 200m, taken from Svensson, Angelborg-Thanderz, Sjoberg, and Olsson (1997)

Similarly, complexity is an essential construct within ATC (Athènes, Averty, Puechmorel, Delahaye, & Collet, 2002; Djokic, Lorenz, & Fricke, 2010; Mogford, Guttman, Morrow, & Kopardekar, 1995; Xing, 2007). For example, Djokic et al. (2010) could identify 24 ATC complexity factors, for example the number of aircrafts. After conducting a principle components analysis (PCA), it was found that all resulting components were significantly related to subjective workload ratings. Although ATC complexity is not the same as visual complexity, important aspects such as the number of aircrafts are likely to depict an overlap between both concepts. Furthermore, Xing and Manning (2005) stated the importance of ATC complexity with regard to controller workload. Consequently, Xing (2007) more closely investigated the information complexity of ATC displays within a literature review. By trying to combine the reviewed definitions and measures, he found that quantity and variety of elements as well as the relation between elements were the three factors that all references converged to. Accordingly, the concept of information complexity might be seen as strongly related to visual complexity (see 2.2.3). A graphical illustration of the quantity factor within ATCs can be found in Figure 12. Xing (2007) argued that the search time for a specific target increases with the number of visual elements within a display due to the serial processing of visual details, which should also affect the number of fixations.

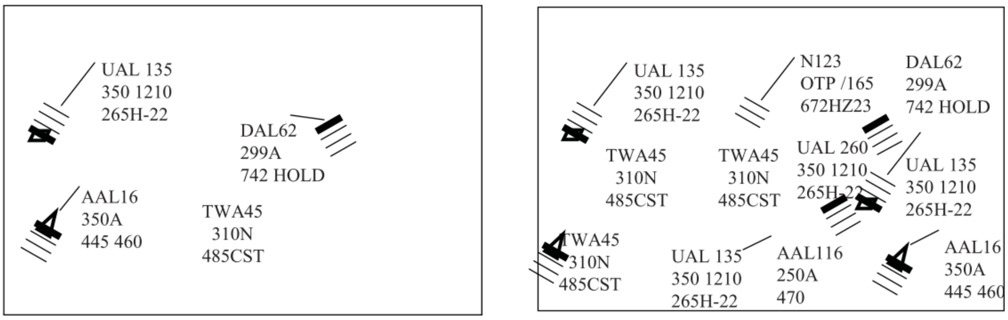


Figure 12. Quantity factor within ATC depictions (taken from Xing, 2007)

Beyond ATCs, the importance of visual complexity with regard to mental workload may also be extended to other types of control rooms. Within the context of nuclear power plants for example, Hugo and Gertman (2013) developed a method for the estimation of display complexity. Within nuclear power control rooms, human errors can have fatal consequences, while complexity is a key factor for human error as well as reliability in these (Cummings, Sasangohar, Thornburg, Xing, & D'Agostino, 2010). Other authors similarly suggested effects of information complexity on the operators' mental workload within this context, with information amount as a central aspect (e.g. Jones, Ma, Starkey, & Ma, 2007).

Moreover, the monitoring of closed-circuit television (CCTV) is used not only in security and surveillance control rooms but also within traffic supervision, tunnel safety and remote process control (Pikaar et al., 2015). The authors conclude that with regard to the human factors design of CCTV systems, both task complexity and image complexity can be highly relevant, although particularly experimental evidence is largely missing. Due to the mostly visual demands within CCTV monitoring and surveillance, visual complexity may be a highly relevant concept within this context and potentially also affect operator workload next to for example the number of screens per operator (Pikaar et al., 2015). The potential relevance of visual complexity on performance during CCTV monitoring is additionally stressed by Howard, Troscianko, Gilchrist, Behera, and Hogg (2009).

Next to mental workload, complexity can also affect situation awareness (Endsley, 2016). For the sake of brevity, this will not be discussed in more detail.

Within this paragraph, the association between visual complexity and mental workload was firstly reasoned at a theoretical level, before underlining it with findings from previous research in human factors. In particular, visual complexity can have an impact on the user's mental workload for user interfaces such as websites, as suggested for example by Buettner (2017). Therefore, the context of user interface design will be addressed in detail within the next paragraph.

2.3.2 Visual complexity and user interface design

User interfaces (UIs) describe "the software and input devices by means of which a computer and its user communicate" (Collins Dictionary, 2019b). Consequently, a UI can be almost anything that allows the use of or interaction with a technical device. One of the most common types of interfaces is the graphical user interface (GUI), which allows the interaction by means of a graphical display.

As general criteria for the design of graphical user interfaces in (alphanumeric) displays, overall density, local density, grouping and layout complexity were proposed in an older research paper by Tullis (1983), based on a literature search. For the first three, empirical evidence pointed towards effects of these on human performance while for layout complexity no direct empirical support was found by the author. Despite the age of the research, Tullis' (1983) findings may be seen as foundations for research on visual complexity of graphical user interfaces and as such still be significant today, although the author did not explicitly refer to the concept of visual complexity.

With the aim of providing a structure of the construct, Miniukovich and Angeli (2014) classified a number of visual complexity determinants of graphical user interfaces into three main determinants. These are amount, organisation and discriminability of information (see Figure 13). This structure however goes back to a literature search on findings for visual complexity determinants and is not further supported by empirical findings.

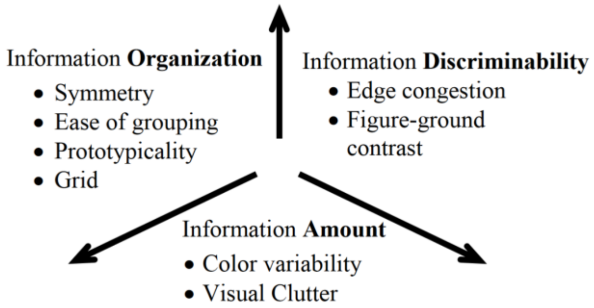


Figure 13. Classification of visual complexity determinants according to Miniukovich and Angeli (2014)

However, in a more recent work, Miniukovich et al. (2018) suggested that this categorization misses the diversity of visual appearance. Hence, the authors complemented the earlier structure, resulting in four visual complexity facets. These are quantity of information, variety of visual form, spatial organization and perceivability of detail. According to the authors, there are nine visual aspects, which describe the four facets (see Table 1).

Table 1.

Facets of visual complexity and GUI aspects representing them according to Miniukovich et al. (2018)

Facet	Visual aspect
Quantity of information	Number of distinct units of information
	Number of groups of units of information
Variety of visual form	Variety of colours
	Variety of sizes
Spatial organization	Vertical symmetry
	Content alignment point
Perceivability of detail	Congestion
	Figure-ground contrast
	Amount of white space

For example, quantity of information represents the number of elements and was described both by the number of individual and grouped units of information according to the authors. This structure, too, was not examined for the validity of dimensionality for example using factor-analytical methods. In the context of in-vehicle instrument clusters, Yoon, Lim, and Ji (2015) for example showed that the perception of visual complexity was related to the two dimensions quantity and structure. Thus, no final judgement of the dimensional structure of the construct visual complexity in the context of human machine interaction is possible.

Instead of considering graphical user interfaces in general, other researchers more closely focussed on the visual complexity of websites, depicting one type of graphical user interfaces in particular. In this context, Nadkarni and Gupta (2007) adopt Wood's (1986) definition of task complexity as a combination of the three dimensions component, coordinative and dynamic complexity in order to define perceived website complexity. Component complexity thereby refers to the density and dissimilarity of elements in a stimulus and might therefore relate to visual complexity (see Miniukovich et al., 2018). Deng and Poole (2010) focussed more specifically on the dimensions of visual complexity in webpages and proposed a structure consisting of two dimensions. The first of their proposed dimension is visual diversity, which refers to the different types of elements, such as graphics, links and text that appear in a webpage. Their second dimension is visual richness. This covers the amount of information or elements such as graphics, links and text. Similarly, Michailidou et al. (2008) propose density and diversity of the presented elements within a website as main influencing variables. Within their empirical work, the authors could show a positive relation between the number of images, links, words and sections of a page and visual complexity.

With regard to the possible impact of the visual complexity of user interfaces, there are multiple aspects worth considering. First of all, relations between visual complexity and aesthetical preference have been investigated within previous research as discussed within paragraph 2.2.5.2. Focussing on user interfaces, findings are rather inconsistent with some results pointing towards a negative relation (Reinecke et al., 2013; Tuch et al., 2009; Tuch et al., 2011; Tuch et al., 2012) while others supported the assumption of an inverted U-shape (Geissler et al., 2006) or found no relation (Pandir & Knight, 2006).

Importantly within the context of human-machine interaction, the visual complexity of user interfaces also affected different performance measures. Lee, Kim, and Ji (2019) for example found that the visual complexity of an in-vehicle information display negatively affected the performance within a visual search task as well as the driving performance of older drivers. With regard to websites, visual complexity was positively related to reaction times in a visual search task and negatively affected memory, with a higher recognition rate for less complex websites (Tuch et al., 2009). Similarly, Wang, Q., Yang, S., Liu, M., Cao, Z., and Ma, Q. (2014) revealed that task completion times in an online shopping task were higher for websites of high or medium complexity compared to those of low complexity. Next to performance, visual complexity was also shown to affect physiological measures such as cardiovascular or muscular activity (Tuch et al., 2009; Tuch et al., 2011). Finally, Harper et al. (2009) proposed that visual complexity of websites can serve as an implicit measure of cognitive load. According to them, especially visually impaired users could benefit from the consideration of visual complexity within the design process, for example by simplifying parts of a webpage.

In conclusion, visual complexity of graphical user interfaces such as webpages can affect users' cognition and emotion in many regards. However, the dimensional structure of the construct is still not consistently agreed on with many findings based on literature research and correlational studies rather than experiments. Particularly, this holds true within applied domains such as human-machine interaction. In sum, considering and investigating visual complexity using experimental approaches can allow for more reliable insights and a better understanding, which can benefit both researchers as well as designers. In this context however, next to a better theoretical understanding, the use of computational measures of visual complexity may be of utility. These could, among others, allow to predict or anticipate users' responses to the design of an interface, which could facilitate the improvement of a design solution by saving time and costs compared to testing and using surveys (Machado et al., 2015). Within the next paragraph, these will be described in more detail.

2.4 Computational measures of visual complexity

Computational measures in general offer the advantage of relative objectivity compared to ratings, which may be confounded by the rater's interest (Aitken, 1974; Day, 1967), familiarity (Forsythe et al., 2008) and novelty (Berlyne, 1970). Counting the numbers of elements within a stimulus such as the number of turns or corners (Attneave, 1957) or the number of lines and letters (García et al., 1994; McDougall et al., 1999) were used among the first approaches to quantify the visual complexity of icons or figures. These showed a correlation with visual complexity ratings, however their calculation was rather time consuming and their use therefore restricted to relatively simple stimuli such as symbols or icons (Machado et al., 2015).

Subsequently, a large number of computational measures were established, which might serve as indicators of visual complexity. Many of these are based on algorithmic information theory (AIT), Kolmogorov complexity or structural information theory (SIT), which were discussed in paragraph 2.2.4. In short, they state that the minimum length of a script describing a visual stimulus can be used as a measure for its complexity (e.g. Chaitin, 1977; Leeuwenberg, 1968; Li & Vitányi, 2008). Within the following, various computational measures are pictured, which were used for the prediction of mean and single complexity ratings within studies 3 and 4. For the reported measures, also subcomponents from their calculation, which might reveal as informative with regard to visual complexity, standard deviations and mean values and their product as well as combinations between types of measures, such as edge and compression measures, were included within the explorative investigations.

2.4.1 Compression measures

Based on the notions of AIT and Kolmogorov complexity, a group of very common computational measures of visual complexity is based on compression algorithms. These analyse the visual information of an image, as described by a bit string, in order to create a reproduction of it that is as true to the original as possible. Since simple images contain more redundant information than complex images, these can be described by a shorter bit string. Consequently, simple images can be compressed to a larger degree than complex images which results in a smaller file size (Donderi, 2006b; Madan et al., 2017; Marin & Leder, 2013). JPEG (Joint Photographic Expert Group) or ZIP compressed files were used within several studies with file sizes revealing to be a good indicator of visual complexity with correlations between file size and subjective visual complexity ratings of up to .80 (e.g. Tuch et al., 2009). File sizes were also shown to predict errors and search time within chart diagrams (Donderi & McFadden, 2005). Other image formats such as GIF (Graphics Interchange Format), PNG (Portable Network Graphics) or TIFF (Tagged Image File Format) similarly revealed solid correlations with ratings (e.g. Gartus & Leder, 2017). Compressed file sizes have already been used in a variety of domains with findings based for example on artistic paintings (Forsythe, Nadal, Sheehy, Cela-Conde, & Sawey, 2011; Marin & Leder, 2013), environmental scenes (Cavalcante et al., 2014; Marin & Leder, 2013), icons (Forsythe, Sheehy, & Sawey, 2003) and technical displays (Donderi, 2006a; Donderi & McFadden, 2005).

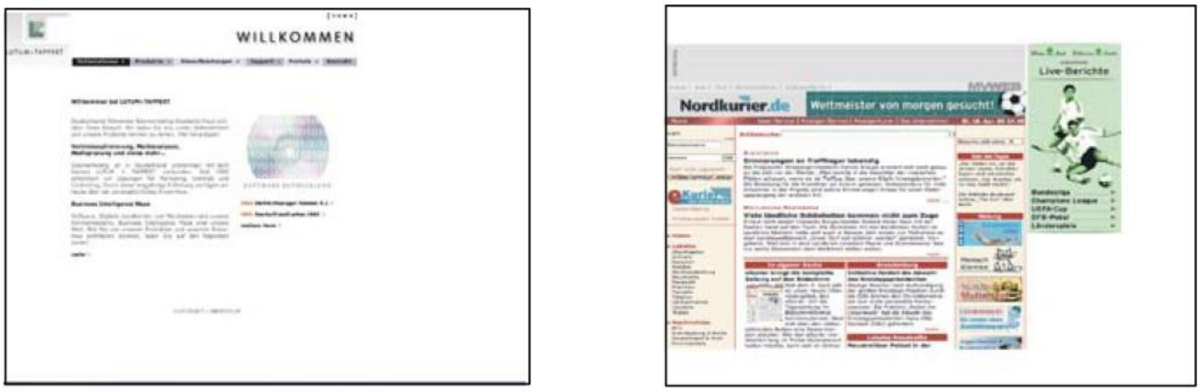


Figure 14. Examples for website screenshots with a smaller JPEG filesize on the left (312kb) and a larger filesize on the right (800kb), from Tuch et al. (2009)

2.4.2 Edge measures

Next to compression, edge detection methods represent another popular group of computational measures. These detect intensity changes at edges within an image. Typically, the edge intensity values can be both visualized as an image (see for example Figure 15) but also quantified as a single value. For example, the proportion and intensity of edge pixels within the full image represent the edge density (Madan et al., 2017). A larger percentage of edges can point towards a higher level of visual complexity (Forsythe et al., 2003). Among the methods used for edge detection are Canny and Sobel filters (Forsythe et al., 2008; Machado et al., 2015; Rosenholtz, Li, & Nakano, 2007) as well as Perimeter detection (Forsythe et al., 2011; Marin & Leder, 2013). Another approach for the detection of edges is the root mean square (RMS) contrast, which is calculated as the standard deviation of pixel intensities. This method also revealed significant correlations with ratings of visual complexity (Cavalcante et al., 2014; Marin & Leder, 2013). Finally, phase congruency is another method for detecting image features (Kovesi, 2000), which can be useful within the context of visual complexity. For example, Marin and Leder (2013) as well as Gartus and Leder (2017) could find a positive relation between phase congruency and visual complexity. Examples for edge images of a website are visualized in Figure 15.

Figure 15. Original website screenshot of Airgas (2020) and examples for Canny, Sobel, Perimeter, RMS and Phase Congruency images (from top left to bottom right)

2.4.3 Decomposition measures

Another group of computational measures can be summarised as decomposition methods. These include quadtree decomposition (Forsythe et al., 2003; Zheng, Chakraborty, Lin, & Rauschenberger, 2009) as well as space-based decomposition (Reinecke et al., 2013). The notion of these decomposition methods consists in dividing an image into multiple quadrants, based on the homogeneity or equal amount of information within the areas. Each quadrant is then further divided into smaller ones until a certain criterion of homogeneity is achieved. An image that is divided into a small number of relatively large blocks is thus more homogenous than one that is divided into a large number of small quadrants (Forsythe et al., 2003). Accordingly, a correlation between the number of quadrants and visual complexity was found, although this was rather weak (Reinecke et al., 2013). Since space-based decomposition is a rather specific measure tailored towards websites, this was not further considered within the studies of this dissertation. An example for quadtree decomposition can be found in Figure 16.

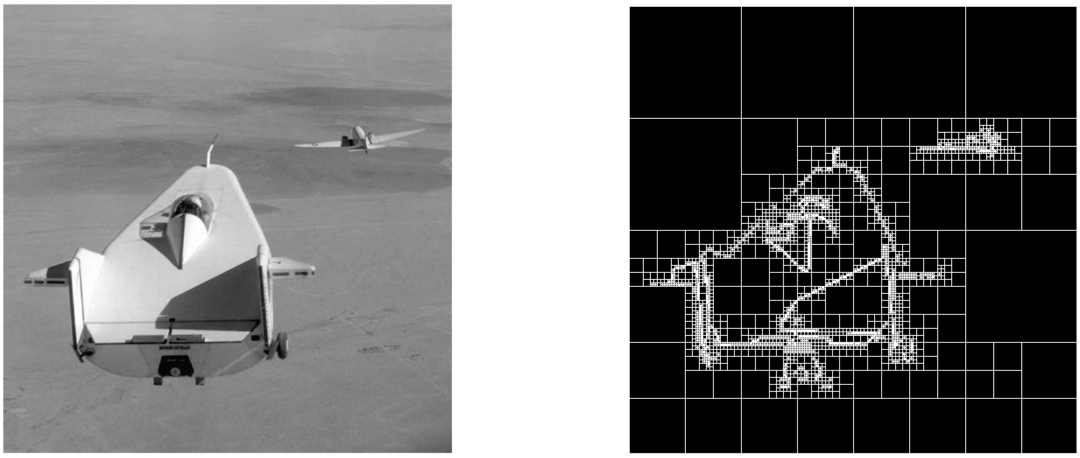


Figure 16. Example for quadtree decomposition from MathWorks (2020)

2.4.4 Structural measures

Next to the three groups of measures described, other computational parameters mainly represent aspects of the structural configuration of elements within an image. Not all of these have yet been considered within the context of visual complexity, since they were developed for different purposes such as aesthetics. However, based on findings about the dimensional structure of visual complexity, considering these within the context of visual complexity may allow for additional insights especially regarding structural facets of images. These computational measures of image structure can mainly be grouped into symmetry and balance parameters.
An accepted measure for the symmetry, which is appropriate especially for rather simple black and white images, is Bauerly and Liu's (2008) measure for symmetry. This is calculated according to a formula, which compares pixel values on both sides of one

or multiple axes of reflection. Among those used in the subsequent studies are the vertical and horizontal image axes as well as both diagonal axes (for quadratic pictures). In addition, the average of symmetries for all two respectively four axes was calculated. Within empirical investigations, some findings point towards good relations with subjective ratings of symmetry (Bauerly & Liu, 2008) while others showed rather small correlations (Hübner & Fillinger, 2016). Regarding the prediction of visual complexity, Gartus and Leder (2017) revealed significant relations with the average symmetry.

Another symmetry measure that is also suitable for more naturalistic stimuli was introduced by Elawady, Ducottet, Alata, Barat, and Colantoni (2017). As opposed to the previous measure, their methodology is also convenient for naturalistic images. The authors evaluated the performance of the method with regard to symmetry, while within research on visual complexity, this has yet only been used by Gartus and Leder (2018) to the best of my knowledge. An example image with the top five detected symmetries is depicted in Figure 17.

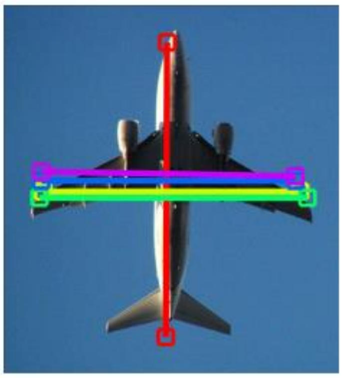


Figure 17. Example image with the top five detected symmetries (in the following order: red, yellow, green, blue, and magenta) from Elawady et al. (2017)

Another relevant aspect of organizational structure of an image structure is perceptual or visual balance. This describes, "how well the elements in a picture are arranged" (Hübner & Fillinger, 2016, p. 1). As suggested by Arnheim (1954), there is a center of perceptual "mass" within each image. This depends on the perceptual weight of its elements, which is again affected by multiple factors such as their size, colour, regularity or distance from the center of "mass". For example, black elements are usually assumed to have a larger perceptual weight than white pixels. Several measures for the quantification of visual balance exist. Among these are the Assessment of Preference for Balance (APB) (Wilson & Chatterjee, 2005), the Deviation of the Center of Mass (DCM) and Homogeneity (Hübner & Fillinger, 2016). The APB measure primarily relies on symmetry, so that measures are calculated both for the vertical, horizontal and diagonal image axes. Next to that, it takes into account the relations between inner and outer areas within an image. The DCM score however focusses on the distance of the center of "mass", which is determined as a position of both the x- and y-axis, from the geometrical center of a picture. Moreover, homogeneity may also relate to the visual balance within an image. This quantifies how scattered the elements in a picture are with less scattering suggesting lower homogeneity and thus lower visual balance according to Hübner and Fillinger (2016). Homogeneity goes back to Shannon's (1948) information entropy. Although not yet investigated within the context of visual complexity, it may be hypothesized that the described measures of visual balance are related to the structural dimension as argued within paragraph 2.2.3.

2.4.5 Segmentation

Further computational measures can help to identify the semantic content of an image more specifically, such as the number of elements for example. For this purpose, especially image segmentation methods can be useful (for a review see Pal & Pal, 1993). Image segmentation is "the process of partitioning an image into meaningful regions or objects" (Vala & Baxi, 2013, p. 387). This can for example be used in order to automatically count the number of objects within the image, which can be achieved by using Matlab's Image Processing Toolbox (The Mathworks, Inc, 2018), as suggested in MathWorks (2019) for example.
Another approach to the segmentation of images for determining the number of features is the mean shift segmentation as described for example by Desnoyer and Wettergreen (2010) and Cheng (1995). This method can be used to segment images according to similarity in colour or intensity by iteratively calculating the mean for moving windows (Kaftan, Bell, & Aach, 2008). Similarly, k-means clustering can also be used for image segmentation (Ray & Turi, 1999).

2.4.6 Spatial frequency

Spatial frequency is relatively popular within vision research, since it has been shown that neurons within the visual cortex are sensitive to certain frequencies of spatial periodical stimuli or gratings (Campbell, Cooper, & Enroth-Cugell, 1969; Maffei & Fiorentini, 1973; Pollen & Ronner, 1983), in particular “the number of grating bars (light or dark bars) per unit of visual angle” (Maffei & Fiorentini, 1973, p. 1255). However, spatial frequencies can also be interesting within the context of visual complexity (e.g. Cavalcante et al., 2014; Corchs, Ciocca, Bricolo, & Gasparini, 2016). Authors of these two references proposed complexity measures that rely on spatial frequency. Moreover, Forsythe et al. (2003) found that icons were judged as more simple if they contained more low spatial frequency information in relation to high spatial frequency information.

An interesting approach to spatial frequency was also applied by Bradley et al. (2007), who determined the frequency of the median fast Fourier transform (FFT) power for each row and column and then averaged it, so that one value represented the whole image. Moreover, Chikhman, Bondarko, Danilova, Goluzina, and Shelepin (2012) found that the product of the squared spatial-frequency median and image area correlated highly with complexity ratings of hieroglyphs. Since the image area was controlled within the subsequent experiments, the squared spatial-frequency median was used.

Also building on spatial frequency, Näsänen, Kukkonen, and Rovamo (1993) introduced a measure for image complexity consisting of the product of median spatial frequency of the Fourier spectrum and the image area comprising 95% of the total contrast energy of the stimulus.

2.4.7 Colour, contrast, brightness

Furthermore, automated measures for analysing the colourfulness of an image reveal additional information in relation to visual complexity. Although conclusions about the influence of colourfulness on the perception of visual complexity are still not consistent (see paragraph 2.2.3 for a discussion), the incorporation of measures for colourfulness may still reveal valuable insights. Two typical methods for computational assessment of colourfulness were established by Yendrikhovskij, Ridder, Fedorovskaya, and Blommaert (1997) as well as by Hasler and Suesstrunk (2003). The former authors used the average colour saturation value of pixels within an image as well as their standard deviation within the CIELUV colour space. They could show a very high correlation between their colourfulness measure and subjective ratings. Hasler and Suesstrunk (2003) however used a more perceptually based approach within the sRGB colour space by calculating colourfulness as the difference against grey.

Next to these measures of colourfulness, colour and intensity entropy was calculated as suggested by Zheng et al. (2009).

Moreover, measures of contrast and brightness were applied as according to Bradley et al. (2011). They defined the measure of brightness as the mean RGB (red, green, blue) pixel value average across all pixels, while their standard deviation was considered as a measure of contrast. Additionally, measures for contrast and hue quality as well as hue count were proposed by Ke, Tang, and Jing (2006), these can give a clue about the simplicity of a picture with regard to the colours.

Moreover, the entropy of the greyscale intensity histogram for a picture can serve as an indicator of visual complexity (Marin & Leder, 2013). This gives a measure of randomness within the image, since entropy increases with a larger variation of pixel intensities. If all pixels have the same intensity, entropy is zero.

2.4.8 Other measures

Rosenholtz et al. (2007) investigated visual clutter, which was defined as a “state, in which excess items, or their representation or organization, lead to a degradation of performance at some task” (p. 3). From this definition, a certain similarity to visual complexity is inherent, although visual clutter focusses more on task performance as in visual search or recognition. For a quantification of this concept, the authors proposed two measures: feature congestion and subband entropy. For the former, the three features colour, orientation and luminance contrast were incorporated, since it is assumed that these determine the degree to which an added new item would draw attention due to a higher level of clutter. The latter, subband entropy, however refers to the similarity in luminance, which is presumed to reflect the amount of visual information in a display.

Next to the single measures, various of these can of course be combined, as for example also done by Gartus and Leder (2017). For example, compressed file sizes for

the different formats such as JPEG or GIF can also be assessed for the edge images generated by Canny, Sobel, Perimeter and RMS method.
In conclusion, a big variety of computational measures have been proposed in previous research. While not all of the reported measures have yet been directly associated with visual complexity, these may still provide interesting information in this context as stated before. Within previous research, computational measures explained considerable proportions of the variance within mean complexity ratings could be explained by these (e.g. Gartus & Leder, 2017; Marin & Leder, 2013). This emphasizes their suitability for the prediction of visual complexity ratings, for example for unknown pictures. However, next to the objective image data, individual ratings can be affected by a number of other factors. Among these can be for example familiarity, interest, liking and aesthetical preference (see paragraph 2.2.5), which can however hardly be captured using computational measures. In order to consider these factors as well as interindividual differences within the cognitive processing of visual complexity, an integration of eye tracking methodology may reveal interesting insights. Theoretical foundations as well as resulting measures are discussed within the next paragraph.

2.5 Visual perception and eye tracking

In order to investigate the attentional processes and also consider interindividual differences within the perception of visual complexity, ocular parameters can be promising. In order to allow for a better understanding of eye movements and the underlying cognitive processes, I will firstly describe some relevant foundations of visual perception and attention before focussing on the ocular parameters which can reflect these processes.

2.5.1 Visual perception and attention

The Gestalt psychologists investigated the principles of perceptual grouping and object perception in general as one of the first ones (as described in 2.2.2) and contributed to the research foundations on visual perception. Subsequently, a large number of researchers focussed on visual perception and attention (see e.g. Bruce, Green, & Georgeson, 2006). Due to the large scope of this field of research, I will point out only theories and principles that appear especially relevant as background for this work. For the sake of brevity, I will not go into details with regard to the neurophysiological foundations for visual perception.

One relevant theory for perception in general is Broadbent's (1958) filter theory, which particularly stresses the selectivity of human perception processes. He differentiated between two stages of processing, a preattentive stage, which encodes simple physical properties and a second, serial attentive state, which encodes more abstract properties. This notion is still quite influential today (Driver, 2001). Broadbent (1958) assumed that the second stage has only limited capacity of attention available and thus information needs to pass a selective filter, which takes action at the early first stage of processing and only lets relevant information pass. The theory could however not explain important findings such as the cocktail-party effect and therefore misses important points, why it is not considered as valid today (Eysenck & Keane, 2000). Due to these findings, which were opposed to early selection, two consequent approaches were taken. While the theories of late selection (e.g. Deutsch & Deutsch, 1963) stated that even unattended stimuli are fully processed and selection only happens before entry into memory, Treisman (1964) instead proposed within her attenuation model that in contrast to Broadbent's (1958) model, unattended stimuli were “attenuated” but not filtered out completely. Within their later feature integration theory, Treisman and Gelade (1980) however also differentiated between two stages of processing. Within the first preattentive stage, an object is broken down into features such as orientation, location and colour. Within the second focussed attention stage, these features are recombined so that the whole object can be perceived instead of individual features (Goldstein, 2010; Quinlan, 2003). The theory also incorporates neurophysiological findings of the “what” and “where” pathways, which are related to findings that visual cortical areas are organized into one pathway for object vision and one for spatial vision (Ungerleider, 1994). Another approach for explaining the perception of objects is Biederman's (1987) recognition-by-components theory. Within this, the author proposes that we perceive objects by separating them into basic units of objects, which he called ‘geons’. These can be 3-dimensional shapes such as cones or cylinders, which can be assembled to an unlimited number of objects (see for example Figure 18). This theory can however hardly explain mechanisms for the perception of complex natural scenes.

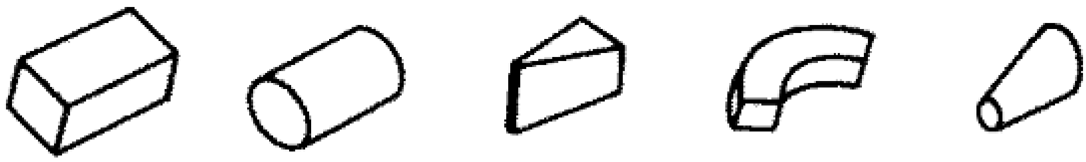


Figure 18. Examples of Geons proposed by Biederman (1987)

Similar to earlier approaches, Marr (1980) thought of the perception process as consisting of different stages. Within his influential theory, he proposed that visual perception acts as an information-processing system which transforms the visual input on the retina into an alternative description of the structures of the image. On the base of the retinal image, a primal sketch is constructed as a representation for elementary global features of an image such as edges, corners, curves and boundaries. Within the next stage, orientation as well as the rough depth and motion are incorporated, resulting in the 2½D sketch, which already gives a first spatial description of the world. Finally, the 3D model representation integrates information from both retinal images in order to construct a stable three-dimensional representation of the world which is independent of the observer’s position but centred on the object. With regard to visual complexity, especially the construction of the primal sketch may play an important role, with more complex images eventually making the construction harder (Riglis, 1998).

Since more than 25 years, an extensive debate emerged between two approaches to processing and selection of stimuli that still affects research today (see e.g. Theeuwes, 2010). This concerns the role of top-down processing, which depends on voluntary control, and bottom-up processing, which is related to features of the stimulus, for visual attention.

Among the early representatives of top-down processing are Helmholtz (1867) and Gregory (1970). While Helmholtz (1867) stated within his likelihood principle that perceptions of visual forms or patterns reflect the most likely object or form, Gregory (1970) argued that perception is based on a cognitive hypothesis. According to him, knowledge and information that is stored from previous experiences interacts with sensory events in order to make inferences about what we perceive. According to Gordon (2004), this theory is particularly influential because it can explain various phenomena such as the extraction of objects from background clutter or the perception of ambiguous objects, which relies on previous knowledge. The importance of top-down influences could be shown for example within Posner's (1980) cueing paradigm and within a row of experiments by Folk, Remington, and Johnston (1992). These revealed that attentional capture is affected by attentional control, which had been induced by task demands. These findings are further supported by neurophysiological correlates (Bressler, Tang, Sylvester, Shulman, & Corbetta, 2008; Giesbrecht, Woldorff, Song, & Mangun, 2003). All in all, results from visual cueing studies show that observers can volitionally direct their attention towards a particular location in space or a certain feature, which underlines the role of top-down control.

Representatives of bottom-up processing however stress that stimulus features also play an important role in the processing of stimuli. Gibson (1966) for example, as one of the first ones, followed the bottom-up approach, arguing that perception is direct and follows only one direction, from sensory input to higher-level processing. Treisman and Gelade's (1980) Feature Integration Theory similarly is an early example for bottom-up theories, since it proposes that processing at the early stage is only related to stimulus features such as colour, orientation and location. Moreover, saliency is traditionally considered as one of the most important aspects of bottom-up processing (Theeuwes, 2010). According to Itti and Koch (2001), saliency is independent of a specific task, computed pre-attentively and can lead to the pop-out of certain elements from a visual scene. From computational analysis of visual attributes of an image, saliency maps can be produced. These represent the degree of saliency of locations within an image. Several researchers found that they can predict human gaze (e.g. Itti & Koch, 2001; Schauerte & Stiefelhagen, 2012) while others argue that saliency does not necessarily predict fixation location during visual search and observed correlations cannot be unambiguously attributed to image attributes (Henderson, Brockmole, Castelhano, & Mack, 2007). They suggest that cognitive factors play an important role in gaze control. Obviously, bottom-up processes cannot explain all aspects of visual attention alone, since voluntary control also plays an important role (Duchowski, 2017; Pinto, van der Leij, Sligte, Lamme, & Scholte, 2013; Theeuwes, 2010).

In general, it is well recognized that both bottom-up and top-down processes affect processing and attention within different stages (Pinto et al., 2013; Theeuwes, 2010). However, there is still a debate on how both processes work together. While Theeuwes (2010) for example suggests that stimulus-driven properties play a role at the early stage and volitional top-down control later in time, Pinto et al. (2013) argue for two independently operating systems. Other authors more generally criticise the dichotomous differentiation between bottom-up and top-down attentional control, because both do not account for findings of selection biases, which cannot be explained by saliency or volitional aspects either (Awh, Belopolsky, & Theeuwes, 2012). Instead, they propose to integrate selection history as a third dimension next to goals and physical salience (Awh et al., 2012).

In conclusion, the last paragraph revealed that attentional aspects play an important role within visual perception. This of course also affects the processing of visual complexity. First of all, stimulus features can affect the attention and selection within the bottom-up processes. Furthermore, previous experience and knowledge can influence the top-down processing. In order to consider and utilize visual attention within the investigation of visual complexity, eye movements can allow for considerable additional insights. Within the following, I will thus take a closer look at the investigation of eye movements for the study of visual attention in order to lay the foundations for a better understanding of the subsequent discussion of parameters for visual complexity.

2.5.2 Visual attention and eye movements

Although the first eye tracking investigations date back until the 19th century (for an overview see for example van Gompel, 2007), Buswell (1935) was the first to apply a systematic approach for the exploration of eye movements and fixations when viewing complex pictures instead of text or simple patterns (van Gompel, 2007). Yarbus (1967) later posed different questions to the participants of his relatively influential experiment, while their eye movements were recorded when viewing the same image several times. These questions addressed for example the material circumstances, ages, clothing or activities of persons depicted within the image. He found that depending on the task, the distribution of fixations varied significantly (see Figure 19).

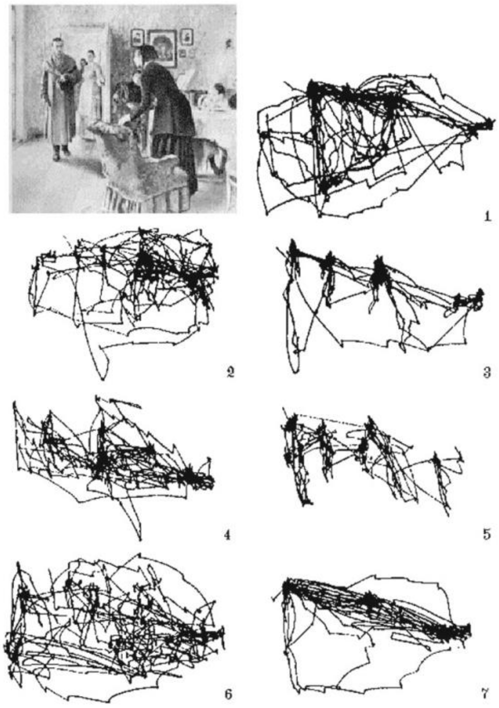


Figure 19. Yarbus' (1967) recordings of eye movement from the same subject with seven different tasks (taken from Duchowski, 2017, p. 9)

Accordingly, Yarbus (1967) stated that eye movements provide insights into cognitive processes since these reflect the observers' attention. These findings could be extended by the work of Noton and Stark (1971a, 1971b, 1971c). They showed that participants often fixate specific features of an object such as curves and angles, which provide the most information. According to the authors, this happens serially, so that features are assembled one after another while being matched with internal (memory) representations of the object. However, even without varying tasks, the order of eye movements or scanpaths of subjects viewing the same image can vary both between individuals but also between multiple observations of the same individual.
Both Yarbus' (1967) and Noton and Stark's (1971a) works built the foundations for the use of eye tracking in order to gain insights into processes of visual attention. In contrast to the Gestalt psychologists, their findings strongly point towards a serial construction of mental representations through sequentially fixating multiple regions of interest instead of perceiving a scene as a whole (Duchowski, 2017). However, it is important to note that (visual) attention may consist of more than just foveal gaze (Duchowski, 2017). An obvious example for this phenomenon is that in order to see weakly glowing stars within the night sky, it is useful to look slightly next to them (because of

the higher number of rods next to the fovea, which are designed for perception at low light levels). Attention can be directed at the star, although it is not within the fovea. Within a research context, this was for example shown by Posner, Snyder, and Davidson (1980), who found that attention can be independent from the foveal direction of gaze, for example when stimuli are presented at a certain angle next to a fixation point. The finding that objects can be perceived and categorized without direct spatial attention was also supported by several more recent research works (e.g. Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002; Reddy, Wilken, & Koch, 2004). Even though many people assume that attention is strictly related to the fovea, because this is mainly the case within everyday life, this is not necessarily the case. Instead, Posner et al. (1980) for example compared attention with a spotlight, which “enhances the efficiency of detection of events within its beam” (Posner et al., 1980, p. 172). Accordingly, the authors separated two aspects of visual attention: orienting and detecting. Orienting encompasses the pointing of attention into a direction which does not necessarily require eye movements. Detecting on the other hand, meaning “the contact between the attentional system and the input signal” (Posner et al., 1980, p. 173) however requires an attentional spotlight, which goes along with eye movements. The distinction between different types of attention was also supported by research of Reddy, Moradi, and Koch (2007), who found different patterns of neural activity within the cortex for effects of spatial attention and a task-based component of attention.

Within more recent research on scene perception, it could be shown that observers can understand the gist of a scene very quickly, even when a scene is shown as short as 40ms (Castelhano & Henderson, 2008) and before the eyes begin to move (Graef, 2005). But where do people look? It was discovered early on that particularly informative areas are often fixated (Antes, 1974; Mackworth & Morandi, 1967). Moreover, the saliency of different locations seems to play a role (Mannan, Ruddock, & Wooding, 1995, 1996; Parkhurst, Law, & Niebur, 2002; Parkhurst & Niebur, 2003), which supports the role of attentional bottom-up processes discussed before (see 2.5.1). As Da Silva, Courboulay, and Estraillier (2011) could show, saliency can also play a role for the perception of visual complexity. They found that saliency and attention maps can serve as estimators for the image complexity of natural scenes. In sum however, very little efforts have been taken until today in order to investigate visual complexity by means of using ocular parameters, which can help to describe and quantify viewing behaviour. This is surprising, since considering the previously described findings, eye tracking can be helpful in order to gain insights into attentional processes and beyond. Rare approaches in this direction were however undertaken by Bradley et al. (2011) and Madan et al. (2017). While the main focus of both studies was on emotion, they also found a greater number of fixations, shorter fixation durations and longer scanpaths for scenes compared to the less complex figure-ground images. They did however not investigate in more detail which aspects of visual complexity produced these results.

In conclusion, eye movements such as fixations can provide insights into attentional processes. Particularly early fixations are often spread around informative and salient areas. These may also help to draw conclusions about the visual complexity of a picture. For example, when larger areas of an image are salient, which leads to a larger number of fixations, the picture may be perceived as more visually complex. The next paragraph will focus on ocular parameters, which could relate to visual complexity. Thereby, it is important to keep in mind that perception is often associated with, but does not necessarily depend on visual attention.

2.5.3 Ocular parameters for visual complexity

Recorded eye tracking data allows for various inferences concerning visual processing and visual attention, as discussed within the previous paragraph. Different types of ocular parameters can thereby be considered, such as gaze, blinks or pupil parameters. The former particularly allow for insights into the localization of visual attention, which plays an important role for example within the perception of scenes. According to Duchowski (2017), three types of eye movements are important within the context of gaze: fixations, saccades and smooth pursuit movements. I will focus on the former two within the subsequent discussion of relevant parameters since smooth pursuit movements are of minor relevance when static stimuli are used. Additionally, sequences of multiple fixations and saccades are scanpaths (see also Rötting, 2001), which can also be of interest and will therefore be discussed. The subsequent paragraphs are thus structured into fixation measures, saccade measures, scanpath measures, pupillometry and blinks. All of the described parameters were used for the prediction of visual complexity ratings within study 3 and 4, while the number of fixations, scanpath length and spatial density were also investigated experimentally. Next to the measures, subparts from their calculation were also included for explorative analyses.

2.5.3.1 Fixation measures

Fixations can be described as “pauses over informative regions of interest” (Salvucci & Goldberg, 2000, p. 71) or “eye movements that stabilize the retina over a stationary object of interest” (Duchowski, 2017, p. 44). However, these also encompass miniature eye movements such as microsaccades, drift and tremor (Duchowski, 2017; Martinez-Conde & Macknik, 2015). On average, fixations often last between 200 and 300 milliseconds (ms), but can be as short as 30 ms and as long as several seconds (Holmqvist & Andersson, 2017). In general, longer fixations are often associated with a deeper and more effortful cognitive processing, which is supported by findings from scene perception, usability and visual search (for a review, see Holmqvist & Andersson, 2017).

Within the context of visual complexity, fixations can be of special interest since as pointed out in the definitions, these are usually related to interesting regions, which may appear in larger number within more complex stimuli. Consequently, Bradley et al. (2011) and Madan et al. (2017) found a positive relation between visual complexity and **number of fixations**. Within Bradley et al.'s (2011) work, this also went along with shorter average **fixation durations** for complex images. On the other hand, Moffitt (1980) concluded that fixation duration increases with a larger number of items per fixation (when the information value of each item is held constant) and also increases with a larger information value of each item (when the number of items per fixation is held constant). Nuthmann (2017) additionally found that image features such as luminance, clutter and edge density also had an impact on fixation durations within different experimental tasks. Both the number and duration of fixations can thus be of interest for the investigation of visual complexity, since these may serve as indicators for the quantity and information amount of elements in an image, as for example suggested by Bradley et al. (2011). Although often considered as noise that is present when attempting to hold the gaze steady during fixations (Duchowski, 2017), **drift** was also included as a parameter, although there was no specific hypothesis regarding possible relations with visual complexity.

2.5.3.2 Saccade measures

Saccades on the other hand are usually described as “rapid movements between fixations” (Salvucci & Goldberg, 2000, p. 71) or “rapid eye movements used in repositioning the fovea to a new location in the visual environment” (Duchowski, 2017, p. 40). In contrast to these, pursuit movements help to track slowly moving objects (Rötting, 2001, p. 76).

Concerning relevant parameters, the **number of saccades** similar to the number of fixations as well as the (mean) **amplitude** or **length of saccades** can be of interest within the investigation of visual complexity. For the former, Kotval and Goldberg (1998) for example revealed a larger number of saccades for an extensive search within a computer interface. Moreover, Phillips and Edelman (2008) found that saccade amplitude accounted for a large part of the variance of visual search performance. Furthermore, Bradley et al. (2011) found clear differences between figure-ground and scene images with regard to saccade amplitude.

Moreover, the **fixation to saccade ratio** compares the time spent with the processing of information (as assumed to reflect in fixation time) with the time used for searching (saccade time) (Kotval & Goldberg, 1998). High ratios can thus indicate a higher amount of information processing as opposed to search. However, Kotval and Goldberg (1998) could however not find differences for this measure among different interfaces. All in all, the previous findings suggest that saccade parameters can also serve as indicators of visual complexity.

Velocity was also included as a parameter. Referring to this, Hutton and Tegally (2005) for example found a decrease in velocity for an attentionally demanding secondary task, when subjects performed smooth pursuit eye movements. Moreover, Savage, Potter, and Tatler (2013) showed higher saccade peak velocities for a high cognitive load condition.

2.5.3.3 Scanpath measures

Scanpaths are sequences of consecutive fixations and saccades (Goldberg & Helfman, 2010). From these, a variety of measures that can be derived. These will be described in the following, including some findings from previous research.

One of the typical scanpath parameters is **scanpath length**, which is calculated by summing the distances between gazepoints (Kotval & Goldberg, 1998). The previous authors for example used the length of scanpaths as a measure for search behaviour in websites with longer scanpaths indicating less efficient search behaviour. Similarly, Renshaw, Finlay, Tyfa, and Ward (2003) used scanpath length as an indicator of good versus bad design of graphs with shorter scanpaths for the well designed graphs. Similarly, Simonin, Kieffer, and Carbonell (2005) found that display layouts with a better visual comfort were also those with shorter scanpaths. With regard to visual complexity, both Bradley et al. (2011) and Madan et al. (2017) found longer scanpaths for more complex pictures.

Another scanpath parameter is **convex hull area**. This represents the area circumscribed by the entire scanpath. Therefore, it can also be used as an indicator of visual search, with smaller areas indicating more efficient search behaviour (Goldberg & Kotval, 1999). Using a circumscribing convex hull instead of a circle reduces the influence of small deviations in gazepoint samples as visualized in Figure 20.

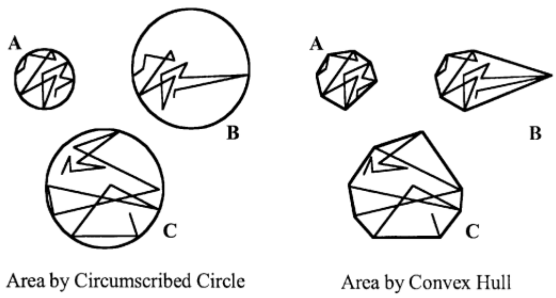
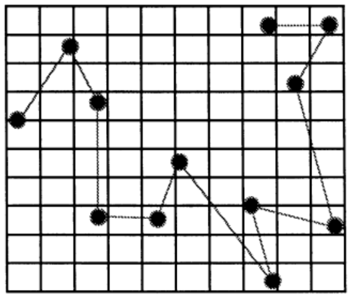


Figure 20. Visualization of the convex hull area, from Goldberg and Kotval (1999)

Moreover, **spatial density** describes the spatial distribution of scanpath nodes within a grid, for example of 10 by 10 squares, as a percentage of the number of cells with nodes compared to the total number (see Figure 21). This was again used as an indicator of visual search within previous studies. When nodes are distributed evenly in the whole visual field, this produces larger values, indicating more extensive and less efficient search behaviour according to Goldberg and Kotval (1999). On the other hand, smaller spatial density values can point to a more efficient and direct search. Differences in spatial density between different interfaces have been found by Kotval and Goldberg (1998).



Spatial Density = 12 /100

Figure 21. Spatial density visualization, from Goldberg and Kotval (1999)

Based on the defined grid that the picture area was split into, an additional measure was calculated. This is the standard deviation of the number of scanpath nodes in all cells of the spatial density matrix (**SD Nr. of Nodes**) and thus gives an estimate of the variability of the fixation distribution. This is assumed to indicate the regularity of the gaze behaviour. Within an additional measure, the same calculation takes into account only those cells of the matrix with at least one scanpath node (**SD Nr. of Nodes > 0**). This approach may increase the general variance of the measure particularly for short presentation durations, since there are many of the cells within the matrix without scanpath nodes. Moreover, another new measure called **Organisation of Fixations** was calculated. Based on the distribution of fixations within the grid cells, a value of 1 for this measure indicates that all fixations are focussed in one cell while smaller values indicate that these are widely distributed across many cells.

Similar to spatial density, **transition density** also provides information of the area of search but additionally adds a temporal component by integrating a representation of the transitions to and from a defined number of areas (Kotval & Goldberg, 1998). Thereby, a high density may again point towards a rather inefficient search, while a low density can indicate a more directed and efficient search behaviour (see also Figure 22). Like convex hull area and spatial density, transition density may be hypothesized to relate to visual complexity. Different levels of visual complexity may have similar effects as the manipulations of the design of user interfaces used by Goldberg and

Kotval (1999) with regard to the search behaviour and thus affect these parameters accordingly.

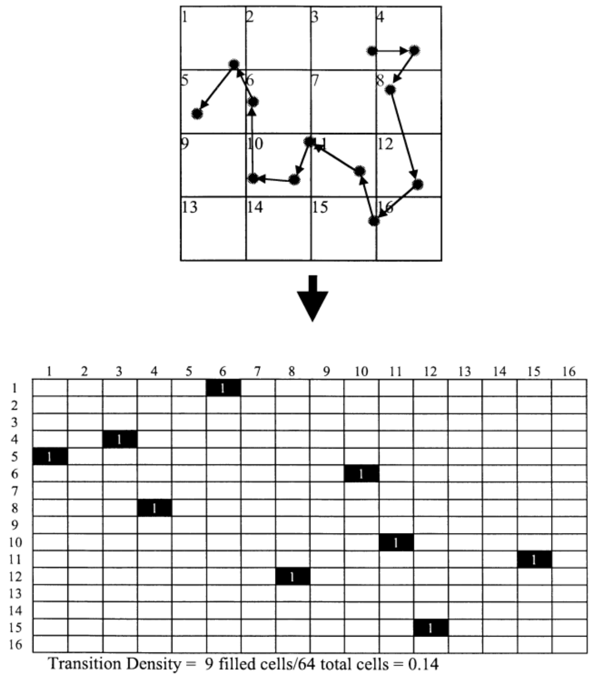


Figure 22. Transition density matrix, from Goldberg and Kotval (1999)

Building on Goldberg and Kotval's (1999) spatial and transition density, two further measures were introduced by Krejtz, Szmidt, Duchowski, and Krejtz (2014) and Krejtz et al. (2015) for the quantification of gaze patterns. Both are based on Shannon's (1948) entropy and should thus serve as estimates of the uncertainty or predictability of a gaze pattern. Therefore, the image is first split into a number of areas of interest (AOIs) or grid areas, similar to the spatial density measure (see Figure 21 for visualization). The primary measure, **stationary entropy,** focusses on the probability distribution of fixations within the different AOIs, from which the average level of uncertainty within the spatial distribution of their sequence is calculated. A higher stationary entropy points towards a homogeneous distribution of visual attention across the different image areas, while a small value suggests that fixations are relatively concentrated on few AOIs. The **transition entropy** however provides a measure for the predictability of visual scanning pattern by calculating the entropy of the transitions between AOIs. A minimum transition entropy value of zero suggests that transitions from a source AOI always go to the same destination AOI, while a large transition entropy means that transitions go from the source AOI to any destination AOI with equal likelihood, producing a larger randomness and thus entropy.

Both measures have been reported to be related to sleepiness and lane departure within a driving context (Shiferaw et al., 2018) as well as secondary cognitive load (Allsop, Gray, Bülthoff, & Chuang, 2017) and anxiety (Allsop & Gray, 2014). However, since they incorporate the randomness versus predictability of gaze, it could be hypothesized that these might also be sensitive to changes in visual complexity.

Further measures focus on the spatial distribution of fixations. Among these is first of all the **Nearest Neighbour Index** (NNI), as described by Clark and Evans (1954). With regard to the distribution of fixations, this describes if they are ordered (with values larger than one), random (with values equal to one) or clustered (with values smaller than one) (Duchowski, 2017). This has often been used within the context of mental workload (e.g. Di Nocera et al., 2015; Di Nocera, Terenzi, & Camilli, 2006), but might also reflect different search strategies, for example of how clinicians observe electrocardiograms as in Davies, Vigo, Harper, and Jay (2016). As for the other measures, it might be hypothesized that the NNI might also be related to visual complexity, since the distribution of fixations with regard to clustering or ordering may also be affected by the visual complexity of the stimulus.

Another measure for the quantification of the time course of eye movements is the coefficient **K**, which indicates **focal or ambient attention** (Duchowski & Krejtz, 2017; Krejtz, Duchowski, Krejtz, Szarkowska, & Kopacz, 2016). It takes into account both fixation durations as well as saccade amplitudes for the calculation. Positive values of the coefficient, resulting from longer fixations followed by short saccades, indicate focal processing while negative values of the coefficient due to short fixations followed by long saccades indicate ambient processing. Previous research with this has for example revealed effects within a cartographic task (Krejtz, Coltekin, Duchowski, & Niedzielska, 2017).

2.5.3.4 Pupillometry

Next to gaze parameters, the assessment of the pupil size allows for the calculation of further measures which may relate to visual complexity. Next to adaptations to natural lighting conditions, it has been found that the pupil can also change its size due to cognitive processes. In particular, pupillary responses have been (successfully) used as estimates for the intensity of mental activity and information processing (Laeng, Sirois, & Gredebäck, 2012; Sirois & Brisson, 2014). Following the early work of Hess and Polt (1960), which showed sensitivity of pupil size for arousal and interest, one focus within the subsequent research was cognitive effort and mental workload (e.g. Iqbal, Zheng, & Bailey, 2004; Palinko, Kun, Shyrokov, & Heeman, 2010; Pomplun & Sunkara, 2003) as well as attention (Burge et al., 2013; Geva, Zivan, Warsha, & Olchik, 2013). Due to its sensitivity for light, it has to be considered that luminance levels should either be controlled or light-independent measures of pupillary activity could be considered in order to avoid possible confounding. With regard to the latter aspect, a number of measures have been proposed that address this issue. Among these are the Index of Cognitive Activity (ICA), the Index of Pupillary Activity (IPA) and the low frequency/high frequency (LF/HF) ratio.

The **Index of Cognitive Activity (ICA)** was developed by Marshall (2002) as a measure of the cognitive effort of a user during the interaction with a visual display. It is based on the assumption that effort-related changes in pupil dilation appear faster than light-related changes, which however usually have a larger amplitude. The measure has shown to positively relate to cognitive load within multiple previous research works (Demberg, Sayeed, Mahr, & Müller, 2013; Dlugosch, Conti, & Bengler, 2013; Schwalm, 2009). The **Index of Pupillary Activity (IPA**) (Duchowski et al., 2018) uses a similar approach as the ICA and was shown to be sensitive to task difficulty by the authors. Next to the final parameters, partial parameters involved within the calculation such as the rise and drop based on the filtering procedure as well as possible alternative calculations such as using a discrete instead of a continuous filter approach, the pupil area instead of the diameter or the raw samples instead of filtered data were included exploratively. Finally, the **LF/HF ratio** of power spectral densities of pupillary signal was proposed as an additional luminance-independent measure by Peysakhovich, Causse, Scannella, and Dehais (2015). This considers more closely the interaction between effects of luminance and cognition in order to exclude the influence of the former as far as possible. The authors found that while luminance affected both low and high frequencies similarly, the influence of cognitive aspects (load on memory in particular) had a larger influence on low frequencies within the pupillary signal. They could thus show a significant effect of memory load for their measure independent of luminance. While all three pupillary measures have not yet been investigated within the context of visual complexity to the best of my knowledge, it might be hypothesized that visually complex stimuli may, similar to previous findings, put higher cognitive demands on subjects. Therefore, these measures might likewise be sensitive to effects of visual complexity.

2.5.3.5 Blinks

Blink-based parameters are typically not among the most popular measures within human-machine interaction. However, some previous research works have revealed effects that might be interesting also with regard to visual complexity. Concerning cognitive load for example, Savage et al. (2013) showed that blink frequency increased within a high cognitive load condition. Similar results of an increased **blink rate** for high mental workload conditions were found by Recarte, Pérez, Conchillo, and Nunes (2008). The authors additionally investigated the level of visual demand by introducing an independent visual search task. This however revealed opposite effects compared to mental workload, with a blink inhibition for the condition of higher visual demand. Faure, Lobjois, and Benguigui (2016) similarly showed that more demanding driving environments decreased blink frequency, while increasing as a cognitive secondary task was introduced. According to the authors, eye blinks are not affected by light conditions as opposed to pupil diameter. This may make eye blink frequency and other blink-based measures interesting for the further use as indicators of visual complexity or mental workload. Next to blink frequency, the percentage of eye closure (**PERCLOS**) (Wierwille, Ellsworth, Wreggit, Fairbanks, & Kirn, 1994) is another typical blink-based parameter. While originally a measure for fatigue, Halverson, Estepp, Christensen, and Monnin (2012) found that it could also be used for the classification of workload levels. Eventually, this might also reveal as sensitive for differences in visual complexity.

All in all, various ocular parameters have been developed within previous research. While some of these have yet been studied in the context of visual complexity or visual demand, many have been originally established within different contexts such as search behaviour or mental workload. Nevertheless, these might provide insights into the attentional and cognitive processes involved within the perception of visual complexity. Investigating these may first of all allow for both further insights into perception processes as a consequence of the experimental analysis of the effects of influencing variables, while their integration within the prediction of visual complexity ratings may contribute to more accurate models. Since ocular parameters may allow to quantify certain cognitive and interindividual aspects of information processing, this information may account for additional variance beyond computational parameters.

2.6 Research agenda and goals

The goal of this dissertation is to bring forward research on visual complexity within human-machine interaction in various regards. Subsequently, the current state of research on visual complexity is summarized and gaps within the existing body of research are pointed out. Consequently, the structure of studies for addressing these research gaps is presented.

Multiple definitions for visual complexity exist, with one of the most popular ones describing it as “the level of detail or intricacy contained within an image” (Forsythe, 2009, p. 158; Snodgrass & Vanderwart, 1980, p. 183). Within existing information processing models, the construct is commonly specified in the very beginning, before further cognitive processes happen (e.g. Deng & Poole, 2010; Leder et al., 2004). A large variety of different influencing variables have been proposed for visual complexity, however few approaches have yet been undertaken to systematically investigate their influence. However, Ichikawa (1985) and Chipman (1977) underlined the relevance of both a quantitative and a structural dimension using experimental approaches. Concerning the relations with other constructs, most studies have focussed on connections with aesthetical preference (e.g. Berlyne, 1974; Geissler et al., 2006; Tuch et al., 2012). Within human-machine interaction, visual complexity is starting to gain attention for example concerning the design of websites (e.g. Tuch et al., 2012) or driving (e.g. Edquist et al., 2012). Computational measures have been successfully used as indicators of visual complexity (e.g. Gartus & Leder, 2017), while relations with ocular parameters have only very rarely been investigated (Bradley et al., 2011; Madan et al., 2017).

Accordingly, gaps within the actual state of research concern the dimensionality of the construct. This has not yet been conclusively agreed on, since many existing research works, focussing on the dimensionality of influencing variables, contradict each other. This impedes a general understanding and common definition of the construct. Moreover, much of the research on influencing variables of visual complexity is based on correlational studies while few experimental approaches haven been taken. These would allow for a better control of possible confounding factors as well as conclusions about causal effects. In particular, this holds true for applied research in the context of human-machine interaction.

Furthermore, it is surprising that eye tracking has yet only rarely been used for the investigation of visual complexity. This would provide both better insights into attentional and cognitive processes but can also be used for the prediction of visual complexity ratings in combination with computational measures. Within previous research, predictions have been based on computational measures and therefore allow only for the prediction of mean complexity ratings for stimuli but not for the consideration of interindividual differences between subjects. Finally, hardly any research exists that would permit qualified conclusions regarding the relation between visual complexity and mental workload. This would however underline the role of the construct visual complexity particularly within human factors.

These research gaps are addressed within four studies, which were conducted within the scope of this dissertation project. Their structure and particular focus is visualized in Figure 23.

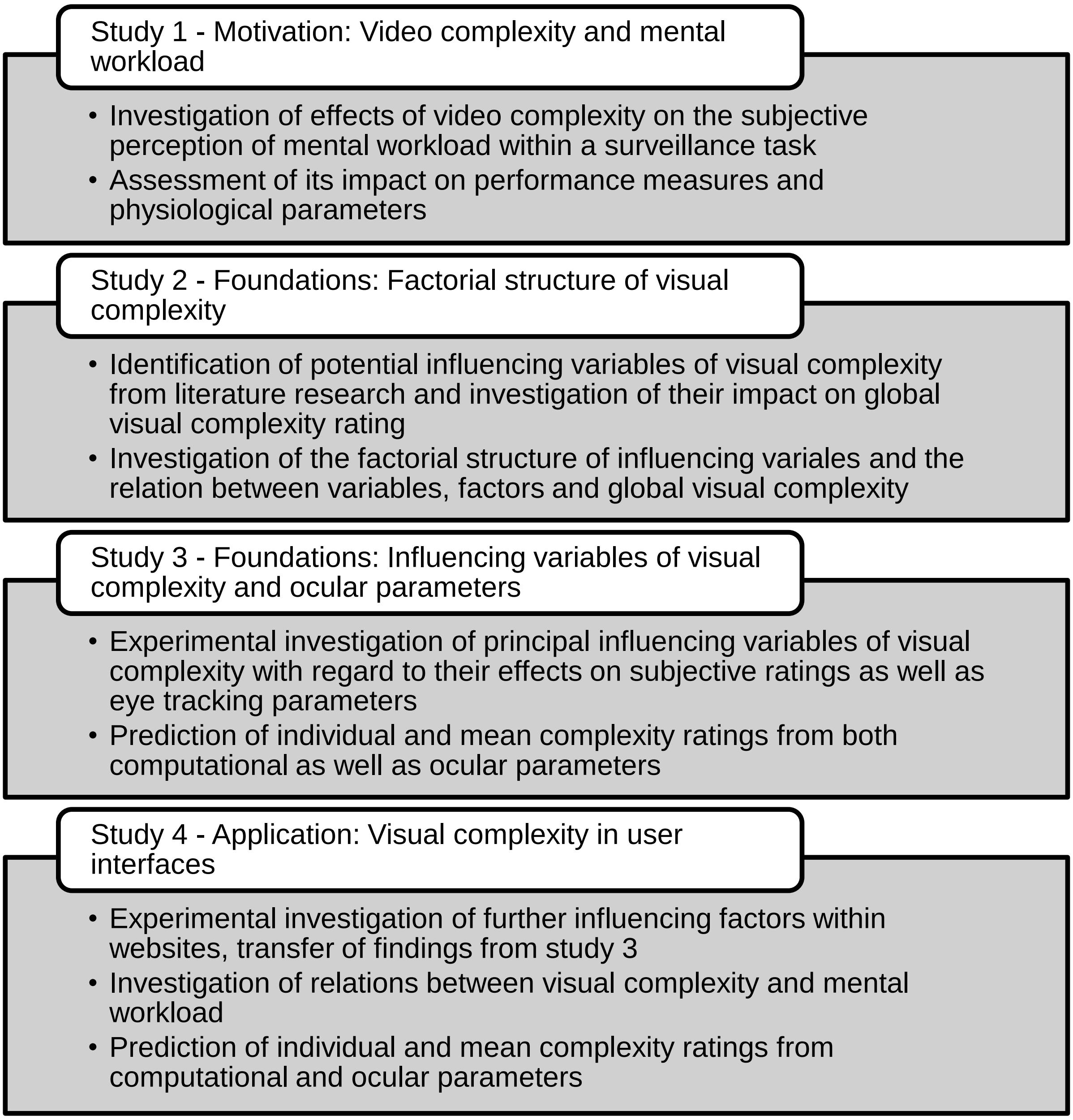


Figure 23. Structure of the four studies conducted of this dissertation project and their research focus

The obtained results will then be integrated into a research model of visual complexity in human-machine interaction, which will also incorporate findings from previous research literature. This will include relevant influencing variables of visual complexity as well as effects on other constructs such as mental workload and performance as well as associations with ocular and computational parameters. This is described and presented within paragraph 7.2.

3. Study 1 - Motivation: Video complexity and mental workload

The goal of the first study on complexity and autocycling frequencies of videos in control rooms is to identify effects of video complexity on the mental workload within an applied task for control room operators. Thereby, the relevance of complexity within the context of human-machine interaction is investigated. Within the following, the background of the study is addressed before the research questions are deducted. Subsequently, the method of the study is described and results are reported and later discussed.

3.1 Background

Within the research project “Video in control rooms: Mental workload analysis”, conducted on behalf of the German Federal Institute for Occupational Safety and Health, a central project goal consisted in the investigation of how different video display options within control rooms affect the mental workload of operators. Within the presented part of the project, effects of autocycling frequency were investigated. Autocycling describes the consecutive automated change between video signals from different cameras on one screen. With regard to video autocycling, little empirical evidence exists regarding the effects on operators’ mental workload. No appropriate publications could be identified within an extensive literature research concerning the effects of using autocycling or different autocycling frequencies on operators’ mental workload. In particular, experimental investigations are missing, which would be particularly helpful for providing reliable recommendations concerning the workplace design in control rooms using video transmission. While Pikaar et al. (2015) suggest that a maximum of 12-16 camera images can be handled as simultaneous livestreams by operators at low task complexity, this can hardly be transferred to the use of autocycling in control rooms. Although Deutsches Institut für Normung e.V. (2008) within the DIN EN ISO 11064 generally suggests to avoid using autocycling in control rooms, it can in some cases be necessary for the monitoring of a larger number of cameras.

Next to autocycling frequency, other aspects may strongly contribute to the mental workload level of control room operators and eventually interact with video autocycling.

Bruijn, Jansen, Lenior, Schreibers, and Pikaar (2016) for example already drew attention to the importance of image complexity within CCTV control rooms. A number of complexity factors were summarized by the authors, including crowding, 'behaviour' and movement of the target and the number of distractors. However, as for autocycling frequency, no experimental investigations of the effect of image complexity could be identified, which makes it hard to draw profound conclusions about its effects particularly on mental workload.

In conclusion, little literature on the effects of video complexity and video autocycling on operator mental workload exists. However, as argued in paragraph 2.3.1, it can be hypothesized that complexity affects mental workload due to the higher demand for cognitive resources required for information processing and at the same time limited capacities, for example in the visual short-term memory (Alvarez & Cavanagh, 2004; Wickens, 2008). Previous findings from applied contexts such as driving (Edquist et al., 2012; Verwey, 2000), aircraft (Svensson et al., 1997) or air traffic control (Djokic et al., 2010) support this theoretical assumption.

Next to video complexity, autocycling frequency can be hypothesized to affect mental workload of control room operators within a monitoring task. Although this cannot be directly deducted from rarely existing previous literature, the demand for cognitive resources is similarly likely to increase with faster autocycling frequencies, since the operator is required to adapt to the context of different cameras backgrounds more often and within a shorter timeframe.

These two factors might not only have an impact on subjective ratings, but also affect performance measures such as response latencies and hit or error rates as well as physiological measures that are typically used as indicators of mental workload. These can contribute to the assessment of a comprehensive image of the workload. A number of frequently used physiological indicators of mental workload are based on the electrocardiogram (ECG) (Manzey, 1998) as well as the registration of eye movements (Marquart, Cabrall, & Winter, 2015). The ECG assesses cardiovascular activity and allows the continuous monitoring of mental workload. Additionally, it is rather unobtrusive, for example compared to electroencephalogram (EEG) and can therefore be assessed more easily within field studies (Roscoe, 1992). Consequently, it has been used for example for the assessment of pilot workload (Roscoe, 1992; Wilson, 2002). With regard to the derived ECG measures, it can generally be differentiated between heart rate and heart rate variability (HRV). Both have shown to be sensitive for differences in mental workload, however HRV is often ascribed a larger relevance (for a review, see Manzey, 1998). Two HRV measures often used as workload indicators are the Root Mean Square of Successive Differences (RMSSD) for subsequent R-R intervals as well as the power spectral density in the low frequency component (LF) from 0.04 to 0.15 Hz (Cinaz, La Marca, Arnrich, & Tröster, 2010; Fallahi, Motamedzade, Heidarimoghadam, Soltanian, & Miyake, 2016; Tjolleng et al., 2017). Both RMSSD as well as LF power values typically decrease for higher levels of mental workload (Cinaz et al., 2010; Heine et al., 2017).

Moreover, ocular parameters can serve as correlates of mental workload and thus provide further insights into the effects of the experimental manipulations. With regard to these, a number of parameters have previously been investigated in relation to different workload levels. Among these are basic measures such as the number of fixations or fixation duration, which have also been reported as related to visual complexity (see paragraph 2.5.3), but also a number of others (see for example Marquart et al., 2015). I will specifically focus on the Percentage of Eyelid Closure (PERCLOS), which is defined as the percentage of time, in which the eyelid covers 80 % or more of the pupil (Marquart et al., 2015). This has typically been used as a measure of alertness or drowsiness (Dinges & Grace, 1998). However, since Halverson et al. (2012) discovered its usefulness for the prediction of mental workload, it has become more popular also within this area (e.g. Schneider & Deml, 2016).

Based on the theoretical background, the following research questions are investigated within this study:

3.2 Research Questions

1. Does a higher level of complexity in videos lead to an increase in the subjective perception of operators' mental workload within a surveillance task?
2. Does autocycling frequency positively affect the subjective perception of operators' mental workload within a surveillance task?
3. Do increased video complexity and autocycling frequency affect performance measures and thus lead to higher response latencies and error rates?

4. Do higher video complexity and autocycling frequency of videos affect the physiological measures RMSSD, LF Power and PERCLOS?

3.3 Method

Given the little previous work, the methodological approach for this study is rather explorative, while at the same time requiring an experimental design in order to provide reliable results. The details regarding the study implementation are described within the following.

3.3.1 Participants

The sample consisted of 34 persons, who were employed full-time within a control room. 15 of the participants were working within a traffic control room (ship lock or public transportation), 15 in a security control room (private security companies, police or swimming lifeguards) and four in system monitoring (from industry, energy or water supply). Participants were recruited with help of the control rooms. The study was conducted during working hours and was not further compensated. Among all participants were 6 women (17.6 %) and 28 men (82.4%) with an average age of 43.6 years (*SD* = 9.3). Age of all participants ranged from 28 to 60 years, with 6.4 years (*SD* = 5.8) of professional experience.

3.3.2 Experimental design and procedure

A controlled laboratory study was conducted within different control rooms across Germany. Therefore, a mobile control room experimental setup consisting of a PC and three 24 inch displays was constructed (see Figure 24). The middle display was used for the presentation of videos while the display on the left showed a map, where in line with the cover story the currently active camera was marked in red. The right display was occupied with a secondary mental rotation task (Schneider & Deml, 2016) based on Shepard and Metzler's (1988) two-dimensional objects, which was conducted in order to assess spare mental capacity that is not occupied by the primary task (Mulder, 1979). This is however not reported here for the sake of brevity, but can be found in Ries and Deml (2019).

For the implementation of the experimental routine, a software was programmed in order to coordinate the video playback and synchronize the assessment of multiple performance and physiological measures.

Figure 24. Experimental setup of study 1

The design of the experiment consisted of a 2 (video complexity) x 3 (autocycling frequency) repeated measures design, resulting in six experimental blocks, which were presented in a random order (see Figure 25).
Video complexity was manipulated by the amount of crowding in the video referring to Bruijn et al. (2016), who identified this factor as relevant for the image complexity in CCTV control rooms. Accordingly, videos with many persons or objects were selected for the condition of high complexity while videos with few persons were selected for the condition of low complexity. The video material will be described in more detail in 3.3.3. With regard to the manipulation of autocycling frequency, the video image either changed every three seconds (fast autocycling), six seconds (medium autocycling) or nine seconds (slow autocycling). Using this experimental design, effects of the two

independent variables video complexity and autocycling frequency on multiple dependent variables were investigated in order to draw a comprehensive picture of the operators' mental workload. Among these were subjective ratings of mental workload, which are described in paragraph 3.3.3.2, as well as performance and physiological measures. While both the percentage of correct reactions to displayed warning symbols as well as the required reaction times serve as measures of performance, the physiological measures are described in more detail within paragraph 3.3.3.3.

Independent variables:

		Autocycling frequency		
		Slow (9s)	Medium (6s)	Fast (3s)
Video	Low	LC-slow	LC-medium	LC-fast
Complexity	High	HC-slow	HC-medium	HC-fast

Dependent variables:

- Subjective: ratings of mental workload
- Performance: Percentage of correct reactions and Reaction Times
- Physiological measures: ECG and Eye tracking

Figure 25. Experimental design of study 1

In the beginning of the experiment, participants were instructed with regard to the experimental task. Based on a cover story, they should imagine working in a police control room, where they were responsible for the video surveillance of different areas within a city. Their primary task was the monitoring of events in the videos, which were presented in the middle screen. In irregular intervals, small black and white warning symbols appeared at random positions on the video stream, which participants had to react to by pressing the space bar. During the six experimental blocks, eye tracking, performance and physiological data were continuously assessed. After each block, a computerized questionnaire was administered in order to capture the participants' perceived level of mental workload (see paragraph 3.3.3.2).

3.3.3 Material

3.3.3.1 Videos

Manipulation of the image complexity of videos was implemented by selecting according videos with a high or low degree of crowding, based on the number of persons visible in the video. The aspect of crowding is assumed to affect image complexity in videos according to Bruijn et al. (2016). Consequently, videos with many persons were selected for the high complexity conditions and videos with few persons for the low complexity condition. Videos were collected from various online platforms, most of which were originally created for computer vision research, such as the CUHK Crowd Dataset (Shao, Loy, & Wang, 2017) or the i-lids Dataset (AVSS, 2007). Examples for videos of low and high complexity are depicted in Figure 26 and Figure 27. The selected videos were then cut into pieces of three, six or nine seconds for the manipulation of autocycling frequency, which required the original videos to be of sufficient length for the manipulation. The selected and edited videos were then presented sequentially within the experiment.

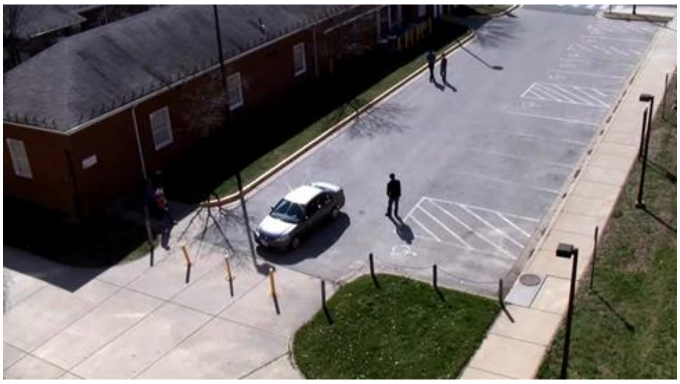


Figure 26. Low complexity video material

Figure 27. High complexity video material

3.3.3.2 Questionnaires

There were several questionnaires administered after each of the six experimental blocks. The main focus within this work is on the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988), which was used in order to assess the mental workload of the participants during their task. This was presented without the scale of physical demand, which appeared of minor relevance within the study. Following the NASA - Raw Task Load Index (RTLX) approach (Byers, Bittner, & Hill, 1989), the mean value of the remaining five scales mental demand, temporal demand, performance, effort and frustration served as the total value for mental workload. The reported values are thus within the range between 1 and 20. Among the other questionnaires that were used was for example the System Usability Scale (SUS; Brooke, 1996). Their results will not be reported here for the sake of brevity, but can be found in Ries and Deml (2019).

3.3.3.3 Devices

ECG

For the recording of the electrocardiogram (ECG), the biosignal recorder Varioport-B of the company Becker Meditec was used. The ECG module can record the signal with a resolution of 0.002 mV within a range of ± 5.4 mV. For this study, a sampling rate of 512 Hz was selected. Thereby, three single-use electrodes were applied to the chest of the participants at the positions suggested by the manufacturer (see Figure 28).

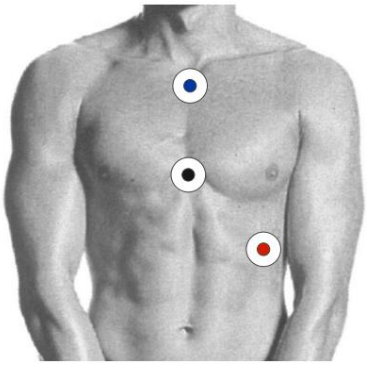


Figure 28. ECG electrode positions, taken from Becker (2016)

From the raw data, QRS-complexes were first detected and from these the Root Mean Square of Successive Differences (RMSSD) as well as power spectral density in the low-frequency (LF) range according to the Welch method as a measure of heart rate variability were calculated with the help of own software routines.

Eye-Tracking

During the experiment, participants were wearing a Dikablis Professional eye-tracker (Ergoneers GmbH) with a sample rate of 60 Hz. This allows the assessment of the pupil with an accuracy of 0.05° of visual angle and of gaze with an accuracy of 0.1°-0.3°. The fieldcam records the visual field of the participant with a resolution of 1920x1080 pixels and an angle of 40°-90°, while the two eye cams record with a resolution of 648x488 pixels. For this study, both the analysis software D-Lab (Ergoneers GmbH, 2017) as well as own software routines were used in order to calculate the Percentage of Eyelid Closure (PERCLOS).

3.3.4 Statistical Analysis

For the analysis of the collected physiological data, primarily the median absolute deviation (MAD) was used, which is a robust method for dealing with outliers (Leys, Ley, Klein, Bernard, & Licata, 2013). I used a moderately conservative criterion of 2.5 to exclude outliers from the further analyses. The subsequent inferential statistical analysis was conducted with Linear Mixed-Effect Regressions (LMER) or Generalized Linear Mixed-Effect Regressions (GLMER) in case of the RMSSD with a logarithmic link function due to the shape of the distribution of residuals. Therefore, the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) with function lmer as well as the *MASS* package (Venables & Ripley, 2002) with function glmmPQL-function were used in software R (R. Core Team, 2018). In comparison with the more traditional approach of using repeated measures ANOVAs, mixed models such as LMER or GLMER allow for the estimation of variance components for random factors such as participants instead of aggregating data, which allows researchers to consider all factors and thereby achieve a better understanding of the underlying data (Baayen, Davidson, & Bates, 2008; Judd, Westfall, & Kenny, 2012). Further advantages of mixed models include the handling of missing and unbalanced data as well as the performance with small numbers of observations (Baayen et al., 2008). Effects of the fixed factors were analysed using *car*'s (Fox & Weisberg, 2019) Anova function as well as *emmeans* (Lenth, 2019) for post-hoc tests. Result plots were created with the help of the R-package *ggplot2* (Wickham, 2016). All error bars within the plots depict the 95% confidence interval. In the following, the χ^2-statistics and Tukey post-hoc tests from these are reported for reasons of brevity.

3.4 Results

Within the repeated measures 3 (autocycling frequency) x 2 (complexity) experimental design, effects of both factors on various dependent variables were investigated. The analyses of the multiple behavioural and physiological measures allow for a comprehensive view on the mental workload of the operators. The dependent variables are reported within the following paragraph. First of all, I will focus on rating data before addressing performance and physiological measures.

3.4.1 Subjective Ratings

Mental workload – NASA-RTLX

With regard to the subjective ratings of mental workload, which were assessed using the NASA-RTLX, the mean value of the five scales mental demand, temporal demand, performance, effort and frustration was used as the total value for mental workload. The statistical analysis of these first of all revealed a main effect of autocycling frequency on ratings of mental workload, $\chi^2(2) = 38.13$, $p < .0001$. Tukey post hoc tests showed that mental workload for video sequences of three seconds was rated as significantly higher compared to six, $\beta = 0.68$, $SE = 0.25$, $p < .05$, and nine seconds, $\beta = 1.52$, $SE = 0.25$, $p < .0001$. Moreover, mental workload was rated significantly higher for blocks with video sequences of six seconds compared to those with sequences of nine seconds, $\beta = 0.84$, $SE = 0.25$, $p < .01$.

Moreover, mental workload was rated significantly higher for experimental blocks with highly complex video material compared to those with little complex material, $\chi^2(1) = 8.47$, $p < .01$. The interaction between autocycling frequency and complexity had no significant effect on mental workload ratings, $\chi^2(2) = 1.10$, $p = .58$.

The rating data are visualized in Figure 29 and details regarding the regression model are reported in appendix 9.1. Within all following graphs, error bars depict the 95% confidence interval.

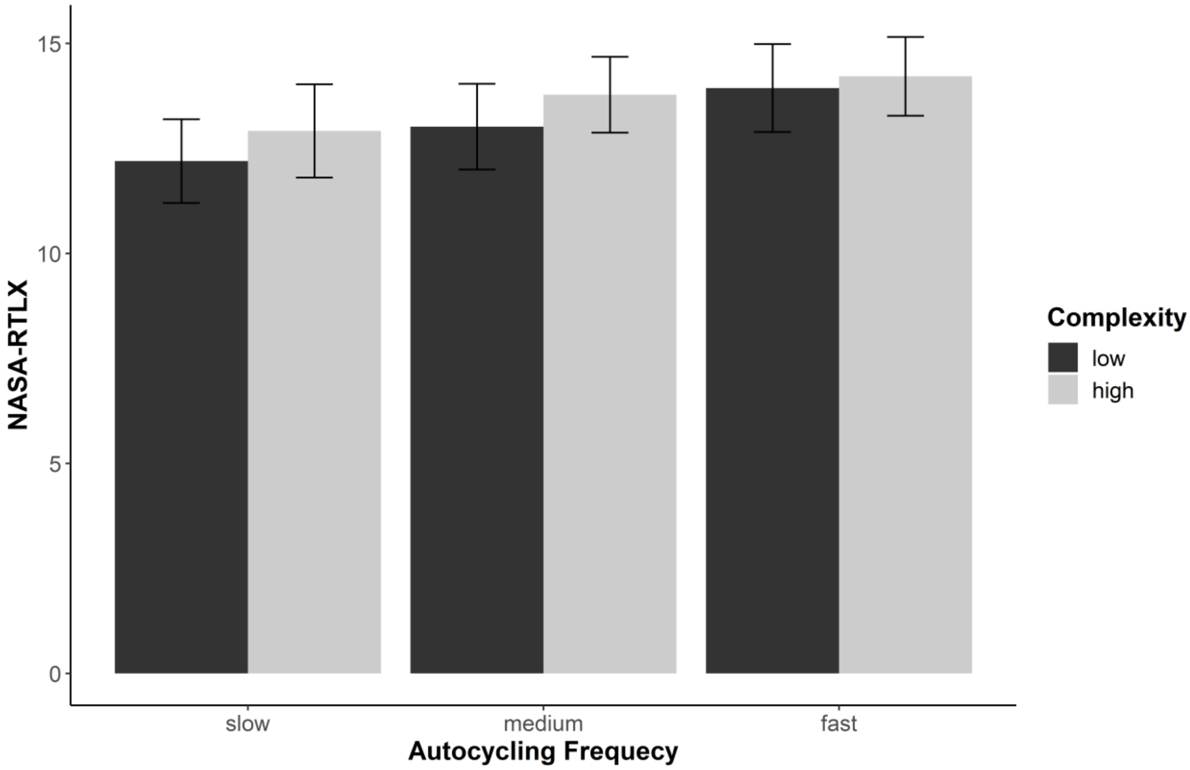


Figure 29. NASA-RTLX ratings in study 1.

3.4.2 Performance Measures

Percentage of correct reactions

Video complexity significantly affected the percentage of correct reactions to the displayed alarm symbols with more accurate reactions for experimental blocks with less complex material, $\chi^2(1) = 20.51$, $p < .0001$. No significant effects were found for autocycling frequency, $\chi^2(2) = 1.07$, $p = .58$, nor the interaction of both factors, $\chi^2(2) = 0.25$, $p = .88$. Data is visualized in Figure 30 and details concerning the regression model are reported in appendix 9.2.

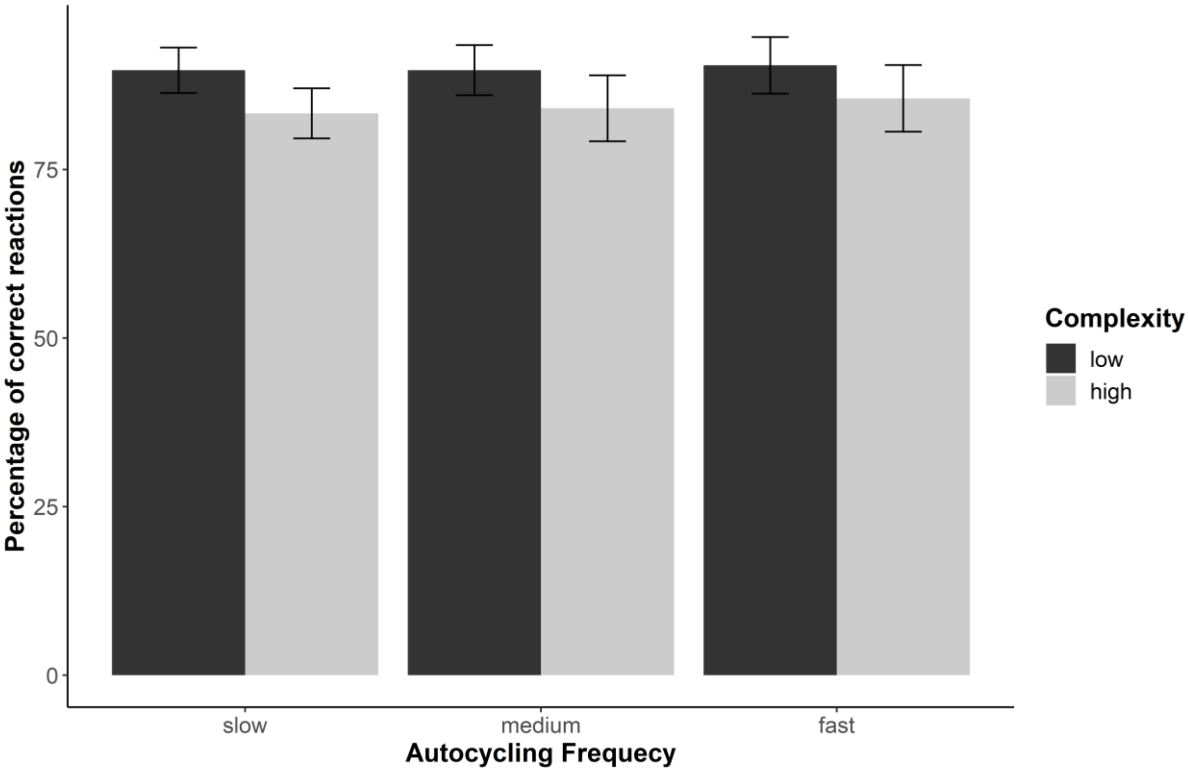


Figure 30. Percentage of correct reactions in study 1.

Reaction time

Regarding the performance measure reaction time, again a significant influence of video complexity could be found with faster reactions for experimental blocks of low complexity, $\chi^2(1) = 52.76$, $p < .0001$. While autocycling frequency had no significant effect on reaction time, $\chi^2(2) = 2.49$, $p = .29$, there was a significant interaction between both factors, $\chi^2(2) = 10.85$, $p < .01$. Due to the hybrid type of interaction, this does however not restrict the interpretability of the complexity main effect (Bortz & Schuster, 2010; Leigh & Kinnear, 1980). Data are visualized in Figure 31 and details concerning the regression model are reported in appendix 9.3.

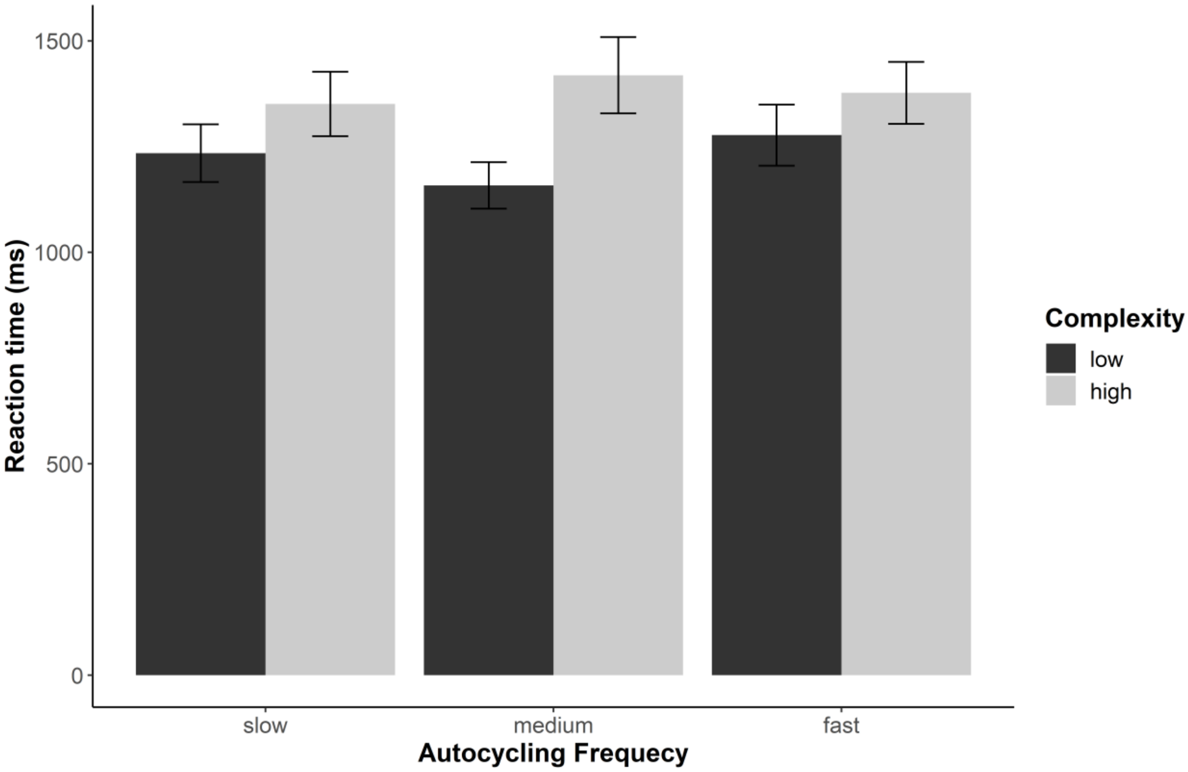


Figure 31. Reaction time in study 1.

3.4.3 Physiological Measures

In order to complete the image on the mental workload state of operators, a selection of workload-related physiological measures was analysed. Among these are both ECG and eye tracking parameters. These are reported within the following. A more comprehensive overview of all assessed measures is reported in Ries and Deml (2019).

ECG - RMSSD

First of all, the ECG measure RMSSD (the Root Mean Square of Successive Differences of subsequent R-R intervals) is analysed. Thereby, a significant effect of autocycling frequency was found, $\chi^2(2) = 7.70$, $p < .05$. Tukey post hoc tests revealed significantly larger RMSSD values for blocks with video sequences of nine seconds compared to three seconds, $\beta = 0.063$, $SE = 0.024$, $p < .05$, while there were no significant differences between blocks of nine and six seconds, $\beta = 0.049$, $SE = 0.024$, $p = .10$, and between blocks of six and three seconds, $\beta = 0.014$, $SE = 0.025$, $p = .84$. Moreover, neither video complexity, $\chi^2(1) = 0.97$, $p = .33$ nor the interaction of autocycling frequency and video complexity, $\chi^2(2) = 0.03$, $p = .99$, had a significant influence on

RMSSD values. Data are visualized in Figure 32 and details about the regression model are reported in appendix 9.4.

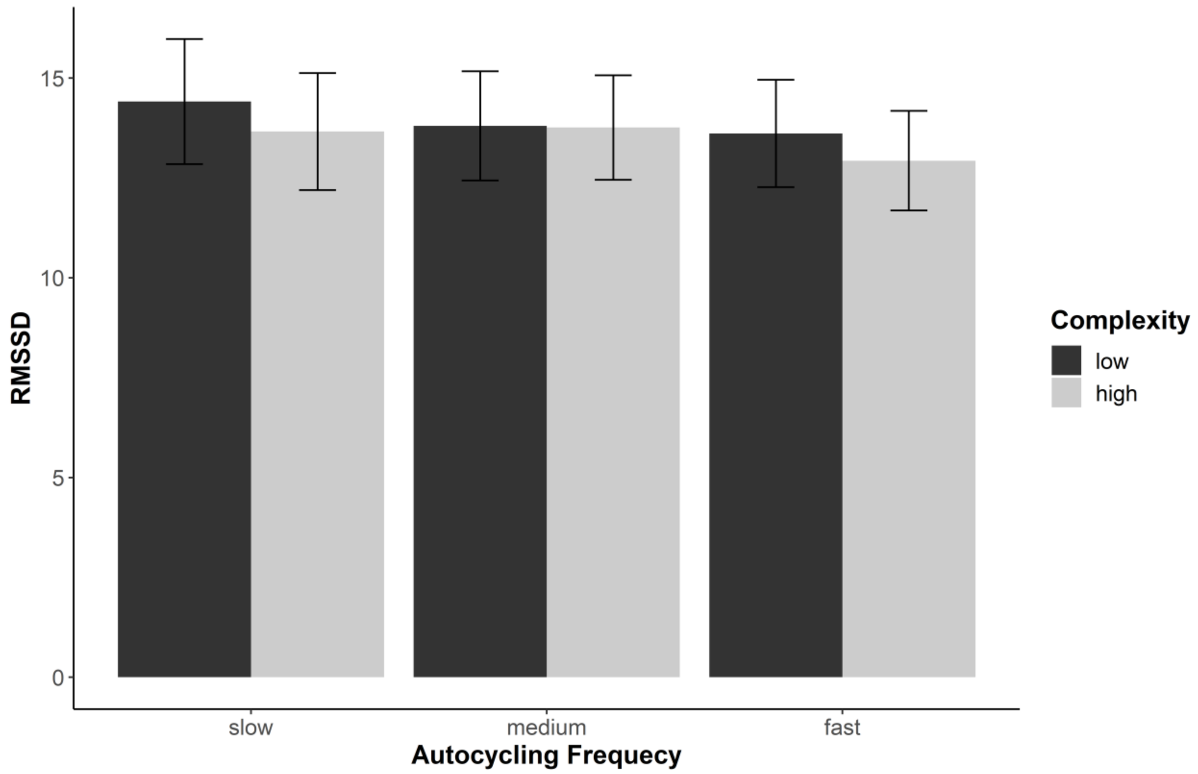


Figure 32. RMSSD in study 1.

ECG – Low frequency power

For the power of the power spectral density in the ECG low frequency component (LF) from 0.04 to 0.15 Hz, a significant effect of video complexity was found with higher power for less complex videos, $\chi^2(1) = 4.82$, $p < .05$. Neither autocycling frequency, $\chi^2(2) = 3.03$, $p = .22$, nor the interaction of both factors, $\chi^2(2) = 1.75$, $p = .42$, had a significant effect on low frequency power. Data are visualized in Figure 33 and the regression model is reported in appendix 9.5.

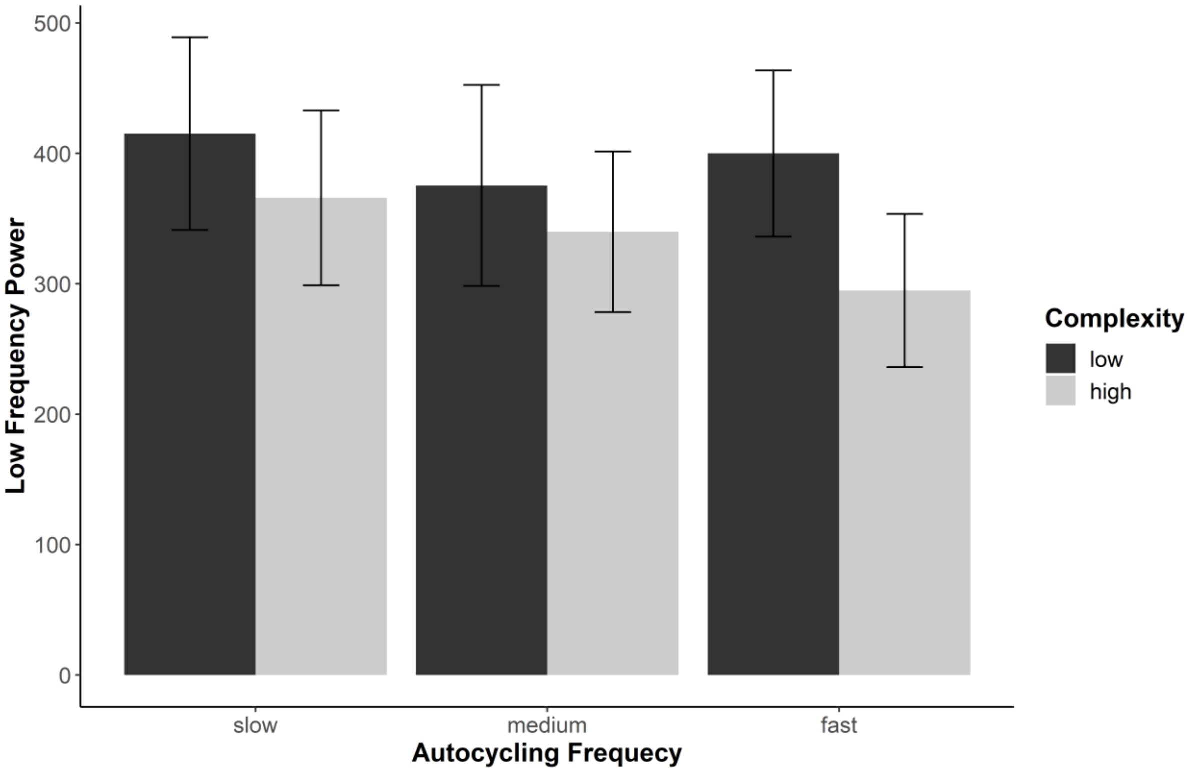


Figure 33. Low Frequency Power in study 1.

Eye-tracking - Percentage of Eyelid Closure (PERCLOS)

Autocycling frequency, $\chi^2(2) = 24.35$, $p < .0001$, had a significant effect on the percentage of eyelid closure with higher values for experimental blocks with video sequences of nine seconds compared to sequences of three seconds, β = 0.81, *SE* = 0.17, $p < .0001$, and six seconds, β = 0.50, *SE* = 0.17, $p < .01$, while there was no significant difference between blocks with sequences of six and three seconds, β = 0.31, *SE* = 0.16, $p = .15$. There was also a main effect of video complexity on PERCLOS values with lower PERCLOS values for more complex videos, $\chi^2(2) = 21.18$, $p < .0001$. The interaction between both factors had no significant effect on PERCLOS values, $\chi^2(2) = 0.52$, $p = .77$. Data are visualized in Figure 34 and details concerning the regression model are reported in appendix 9.6.

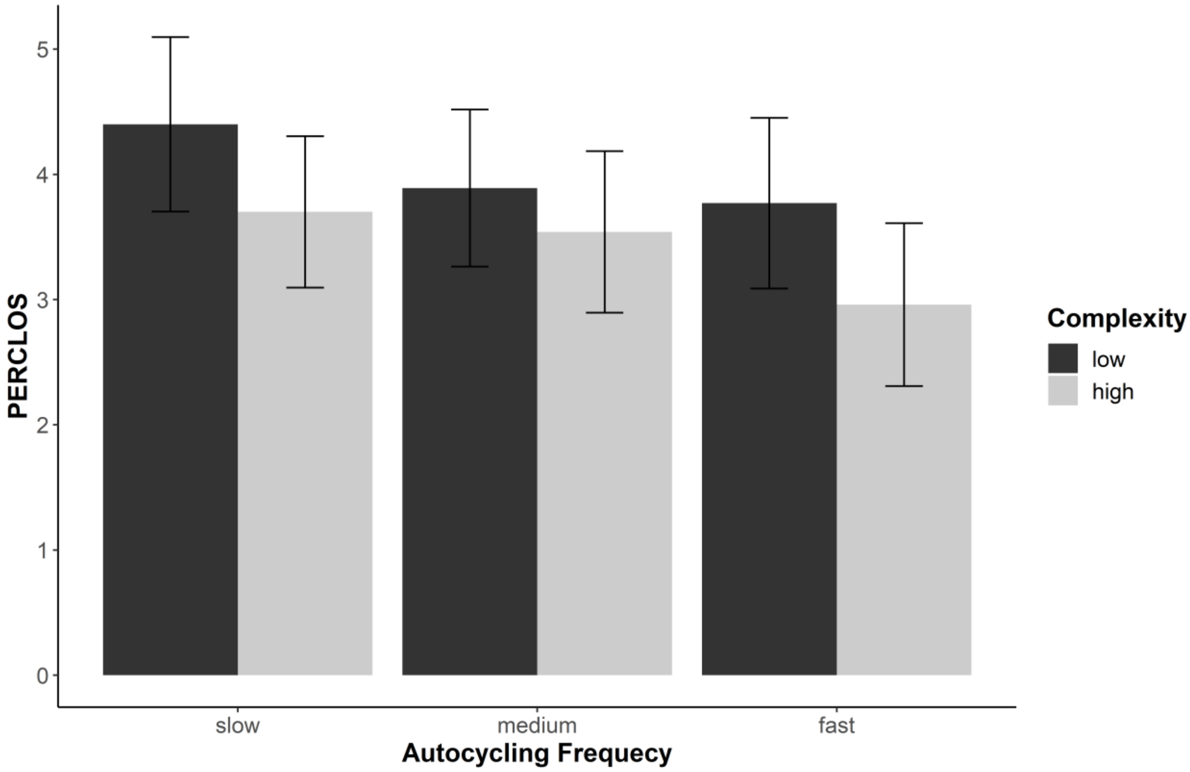


Figure 34. Percentage of Eyelid Closure (PERCLOS) in study 1.

3.5 Discussion

The overall results of this first study stress the relevance of video complexity within the workplace of a control room setting. Both autocycling frequency and complexity of videos significantly affected mental workload ratings with faster autocycling and larger video complexity causing increased mental workload ratings in a CCTV surveillance task.

Moreover, complexity particularly affected the two performance measures, with a lower number of correct reactions and longer response times for more complex videos. Since both performance indicators were negatively affected by the higher level of complexity, this underlines its impact on mental workload. The performance decrements may however also partly be explained by an obstructed detection of alarm symbols due to the higher number of elements in complex videos.

Effects of both autocycling frequency and video complexity on physiological measures further underlined their impact on the participants' workload level beyond subjective and performance measures. First of all, lower RMSSD values were found for faster autocycling frequencies. According to the research literature (Cinaz et al., 2010; Heine et al., 2017), this supports the increase in mental workload associated with faster autocycling. Moreover, significantly lower power in the LF range was found for complex videos, which is also typically associated with increased mental workload (Heine et al., 2017; Mehler, Reimer, & Wang, 2011). Finally, both autocycling frequency and video complexity affected the eye tracking parameter PERCLOS with lower PERCLOS values for both faster autocycling and complex videos. Since PERCLOS is typically positively related to sleepiness and performance decrements (Marquart et al., 2015) and could be used for the accurate classification of workload levels (Halverson et al., 2012), the results appear to plausibly relate to the subjective ratings. In addition, since PERCLOS measures the percentage of time in which the eyelid is closed, it can also be strongly related to eye blinks. These were found to increase with the level of visual demand (Recarte et al., 2008). Within this context, the negative relation between PERCLOS and both autocycling frequency and video complexity further supports the workload effects identified by means of subjective, performance and further physiological measures.

In conclusion, all three types of employed measures underline the role of image complexity within the CCTV control room setting. Within this study, it has shown to affect the workload state of operators next to other features such as the autocycling frequency of videos. Accordingly, the complexity of video material should be considered within the workplace design in CCTV control rooms. Unlike autocycling frequency, it cannot be directly adapted in order to optimize the workload level of operators in most cases. However, it might be important for example to select an appropriate autocycling frequency, which is adjusted to the complexity level in order to avoid overload.
But complexity may not only play a role within work tasks in the context of CCTV or video surveillance. It can also play an important role in many other domains of human machine interaction, since the interaction with graphical depictions is essential for many work-related activities and beyond. In this regard, the design of graphical user interfaces is one domain where particularly visual complexity may play a key role and where the finding of an impact on mental workload from this study can eventually be transferred to. This remains to be investigated within the subsequent studies.

Limitations

Regarding the limitations of this study, when experimenting within rather applied settings using realistic video material, effects of course largely depend on the validity of stimulus material. Within this study, many possible factors may have affected both ratings, performance and also physiological measures. Among these are for example video quality, lighting and contrast conditions or viewing angles. Due to the limited availability of suitable realistic video material within public databases, a certain variability may have remained within the material. Thus, it cannot be excluded that further factors may potentially have produced a confounding of the results. However, best care was taken to avoid systematic differences between complex and simple videos within the above-mentioned confounding factors. For all autocycling conditions of one complexity level, videos from the same sources were used so that confounding could largely be excluded in this respect. Regarding the original stimulus material, a pre-rating of the videos with an independent sample however could have further ensured its validity with regard to both differences in complexity levels but also the influence of possible confounding factors.

Moreover, the subjects' task of course plays an important role for the results. Within this study, subjects monitored CCTV video material, while an integrated reaction task required the response to small alarm symbols presented at random locations superimposed to the videos. It remains possible that a different experimental task with different attentional requirements might have produced diverging results. Thereby, it is possible that the background of the task, which was conveyed by the cover story of monitoring CCTV cameras within a police control room, may have played a role. It can be presumed to have an impact if the video material should be monitored for example in order to recognize violent or criminal acts or in order to estimate the number of passengers for public transportation. Next to the security demands, the frequency of required reactions may also influence the task perception and with that various dependent variables.

Finally, this study strictly speaking does not directly address visual complexity, which mostly refers to static images, but video image complexity. For the implementation of manipulations, I focussed on Bruijn et al.'s (2016) concept of image complexity for CCTV-systems. By selecting videos with a different amount of visible persons, the aspect of crowding was addressed, which is one factor of the image complexity of videos according to Bruijn et al. (2016). Crowding as related to the number of persons within the image might also be a relevant aspect for the visual complexity of static images, referring to the quantitative dimension (see paragraph 2.2.3). Thus, it might be presumed that both constructs overlap, for example with regard to the influence of quantitative and structural aspects, but additional aspects contribute to the complexity of videos such as movement, speed and behaviour (Bruijn et al., 2016). Within these regards, it differs from static visual complexity. However, the impact of complexity, as shown within this study by means of the effect on mental workload, will most likely also carry relevance within other contexts. In order to ensure if a transfer of these findings is valid, further research is needed.

Outlook

Within future studies, the concept of video or image complexity as primarily brought up by Bruijn et al. (2016) can further be investigated, also in comparison to the concept of visual complexity of images. While Bruijn et al. (2016) already suggested a number of factors, which determine the complexity of videos, few of these have yet been experimentally investigated. For the visual complexity of images, a number of different variables and factors (see 2.2.3) have been identified, although these are not generally agreed on either. The relation between both constructs and potential similarities between influencing variables could be addressed in subsequent studies. Relevant influencing variables for both could for example be quantitative aspects such as the number of elements as well as structural aspects such as symmetry. However, the temporal aspect of movement within dynamic videos is obviously a unique feature of video complexity. Approaches for the quantification or measurement of video complexity yet remain to be investigated. One way could for example be the use of file compression methods. With regard to visual complexity, file sizes of image compression methods such as JPEG sometimes serve as complexity measures (e.g. Tuch et al., 2011). Since video files basically consist of a large number of single frames, this approach could eventually be transferred to this material as well. Compressed video file sizes of video formats such as MP4 could then act as indicators of video complexity. To the best of my knowledge, this approach has not been pursued yet and thus remains to be tested. Moreover, referring to an aspect mentioned before, the relevance of different task and work contexts for the perception of complexity remains another interesting aspect to be investigated within subsequent studies. Are videos perceived as more complex if a high level of attention is required due to the security demands of the task? This could be operationalized for example by using the same video material within different task instructions and cover stories, such as the context of airport security with high security demands in comparison to the monitoring of passenger numbers at a local train station with low security demands. Findings could also provide for a better understanding of the influence of aspects such as experience and motivation on complexity perception.

4. Study 2: Foundations: Factorial structure of visual complexity

While the first study underlined the relevance of complexity within human-machine interaction, the next study will closely investigate the construct visual complexity and its influencing variables. The existing definitions as well as the factorial structure are still not commonly agreed on within the research community. Hence, this study provides a better understanding of visual complexity while integrating existing findings.

4.1 Background

Within previous research works, a number of variables have been identified that were assumed to contribute to visual complexity. An extensive overview of these findings is provided in paragraph 2.2.3. Within the former research literature however, results are often strongly diverging with a number of different influencing variables and factors obtained by different authors. With regard to the perception of scene images for example, Oliva et al. (2004) revealed quantity of objects, clutter, openness, symmetry, organization and variety of colours as relevant variables for visual complexity of scenes within a hierarchical grouping. Riglis (1998) however identified symmetry, similarity, smoothness of curves and angles present, minimum description length and familiarity with visual stimuli as the primary aspects based on a literature research. Nadal et al. (2010) on the other hand selected the seven features unintelligibility of the elements ("the difficulty to identify the elements in the image"), disorganization ("the difficulty to organize the elements into a coherent scene", both from Nadal et al., 2010, p. 178), amount of elements, variety of elements, asymmetry, variety of colours and three-dimensional appearance as the most relevant ones based on a literature research. As it can be seen, the identified variables partly overlap between these references, however there remains considerable nonconformity with regard to many others. A similar picture emerges when literature focussing on the visual complexity of user interfaces and websites is considered (see also paragraph 2.3.2). For example, Miniukovich and Angeli (2014) suggested the eight variables symmetry, ease of grouping, prototypicality, grid, edge congestion, figure-ground contrast, colour variability and visual clutter, which they classified according to the three main dimensions information organization, information discriminability and information amount (see Figure 13). In a more recent work however, Miniukovich et al. (2018) selected nine aspects instead, which were classified according to the four main facets quantity of information, variety of visual form, spatial organization and perceivability of detail. The nine aspects considered were number of distinct units of information, number of groups of units of information, variety of colours, variety of sizes, vertical symmetry, content alignment point, congestion, figure-ground contrast and amount of white space (see Table 1). On the other hand, Deng and Poole (2010) for example proposed within a more general framework the dimensions visual diversity, which encompasses the different types of elements within a webpage as well as visual richness, which describes the amount of information such as text, graphics and links as key aspects. Michailidou et al. (2008) however named density and diversity as most relevant aspects.

As it can be seen, the information from existing literature provides a good starting point for a closer investigation and profound definition of the construct visual complexity. However, a number of contradictions between the previous research works become apparent. These discrepancies arise both between different groups of stimuli such as scene images and screenshots from webpages or user interfaces but also appear between different research papers within one specific domain.

This may be related to the fact that much of the existing research such as Riglis (1998), Miniukovich and Angeli (2014) or Miniukovich et al. (2018) strongly rely on literature research for the identification and classification of influencing variables of visual complexity, without further challenging the relations between the identified influencing variables and factors and a global visual complexity score. Here, the incorporation of additional methodological approaches for a closer examination can contribute to more reliable findings and thus to a better understanding of the construct visual complexity. Two older studies have already followed this strategy. Using an experimental approach, Ichikawa (1985) identified both a faster and a slower cognitive process as being relevant within the perception of visual complexity. While the former serves the identification of quantity, the latter influences the detection of structure. Similarly, Chipman's (1977) findings that determinants of pattern complexity can be grouped into quantitative and qualitative variables are also based on experimental investigations. These two examples may thus point to a two-dimensional structure of the construct, with a quantitative and a qualitative factor. Both however only used rather simple black and white stimuli. Nadal et al. (2010) furthermore more closely investigated the structure of seven primarily selected variables using both artistic and non-artistic stimuli and identified the three dimensions “elements”, “disorganization” and “asymmetry”. While the amount of elements, variety of elements and variety of colours were associated with the first dimension “elements”, unintelligibility of elements and disorganization were associated with the second dimension “disorganization”. Finally, the dimension “asymmetry” consisted only of the variable asymmetry. Within the context of the previous experimental works, the three-dimensional structure discovered by Nadal et al. (2010) is at least in some parts surprising. Based on the earlier experimental findings by Ichikawa (1985) and Chipman (1977), it would be assumed that the variables (a)-symmetry and (dis)-organization were related as parts of a structural dimension.

In order to achieve more clarity with regard to the structure of influencing variables and their contributions to a global visual complexity score, it will be analysed more closely within this study. Thereby, multiple contexts are considered within a factor analytic approach. Based on the extensive review of visual complexity literature in foundational (see paragraph 2.2.3) as well as applied domains (see paragraph 2.3.2), a number of aspects of visual complexity were selected for further investigation. It was ensured that only basic variables with possible relevance for different kinds of stimuli (from photographs to screenshots of user interfaces) were selected. Among these, with respect to their incidence, the variables reported in Table 2 were selected. Unintelligibility of elements was however excluded within the process of study 2a) and not used in study 2b), which is described in detail within paragraph 4.4.

Table 2. *Identified potential influencing variables for further investigation*

Number of elements
Variety of elements
Density of elements
Variety of colours
Colour contrast
Organization
Symmetry
Visual balance
Unintelligibility of elements

In conclusion, this study has two primary goals: first of all, the relation between the identified potential influencing variables and a global rating of visual complexity will be investigated in order to draw conclusions regarding their influence on the perception of visual complexity. Secondly, the dimensional structure of all potential influencing variables and their impact on the global visual complexity rating will be addressed in order to provide insights about the dimensionality of the construct visual complexity.
Results of this study can help to achieve a better understanding of the construct and bring more clarity into the existing literature. Previously, many different relevant aspects have been proposed within different research works while yet few integrative attempts have been made in order to consolidate the gathered findings. Moreover, it remains unclear to this date if systematic differences exist between different types of stimuli. The existing literature provides an unclear image with a number of overlapping aspects for example between shape patterns, images and user interfaces, while other proposed aspects greatly vary between the different stimulus materials. This could suggest that visual complexity is a largely domain-specific construct instead of a broadly valid and general one. In order to address this point, two partial studies were conducted using different stimuli, with photographs as basic stimuli and screenshots of websites as application-related stimuli. Comparing findings from both allows for con-

clusions about the generalizability of the construct visual complexity as well as potential differences between regarding both the role of influencing variables and factors for the perception of visual complexity.

4.2 Research Questions

The aforementioned aspects will be investigated within this study according to the following research questions:

1. Do the identified influencing variables number of elements, variety of elements, (unintelligibility of elements), variety of colours, colour contrast, organization, symmetry and visual balance relate to a global rating of visual complexity, both for photographs and for webpage screenshots?
2. What is the factorial structure of the influencing variables, both for photographs and for webpage screenshots? Are all factors related to a global rating of visual complexity?
3. Are there systematic differences between photographs and webpage screenshots regarding the relations of influencing variables and factors with a global visual complexity rating?

4.3 Method

In order to investigate the research questions, two online studies were conducted. While the study design and procedure of both were very similar, study a) used photographs as stimuli while study b) addressed the perception of webpage screenshots[1]. In the following, I will describe the participants, materials used as well as study design, procedure and the statistical analysis of both partial studies in detail. Since features such as the participant sample as well as the stimuli differed between both studies, I will report these separately within paragraphs 4.3.1 and 4.3.2, while describing other

[1] Stimulus selection and data collection for study 2a) was done in close collaboration with master student Yang Xie while stimulus selection and data collection for study 2b) was accomplished in close collaboration with bachelor student Jessica Waibel

parts such as the study design, procedure or the statistical analyses jointly for both partial studies.

4.3.1 Participants

Within study a), 96 persons participated. From these, two were excluded from the further analysis due to a very fast completion of the online survey which reflected in high relative speed index (RSI) values (> 2.5) (Leiner, 2013) and suspicious rating patterns. Of the remaining 94 subjects, 38 were females (40.4%) and 56 (59.6%) males. The average age was 25.7 years (*SD* = 4.4). The majority (80 or 85.1%) of the participants were students.

In study b), 60 persons participated. Of those, 40 (66.7%) were females and 20 (33.3%) males. The average age was 29.5 years (*SD* = 12.7) and the majority of all subjects (34 or 56.7%) were students

4.3.2 Study design and procedure

For both studies, a repeated-measures design was used. In study a), there were the two groups of high and low visual complexity, which were operationalized by using either figure-ground or scene images. In study b) however, a 3 (complexity) x 3 (website type) design was used with low, medium and high visual complexity and screenshots of online shops, news pages and company websites.

Both studies were conducted online using the platform “SoSci Survey”. They began with information about the procedure as well as an informed consent. After agreeing, demographic details were enquired from the participants. Subsequently, one exercise trial with an extra stimulus image was included in each study in order to allow subjects to get accustomed to the study procedure. Afterwards, images were presented in random order with all questionnaire items as described in paragraph 4.3.3.2 listed directly underneath. All ratings had to be entered on seven-point Likert scales. This way, the experiment was performed in a self-paced manner, where subjects could evaluate the stimuli while making their judgements.

4.3.3 Material

4.3.3.1 Stimuli

In study a), photographs were used as stimuli in order to investigate the relations of potential influencing variables with each other as well as with the global complexity rating for these. In order to use strongly controlled material, 18 images from the International Affective Picture System (IAPS) (Lang, Bradley, & Cuthbert, 2008) were selected[2]. Since affective influences were not of interest for the research questions within this study, only pictures from the emotionally neutral category were selected. For the low complexity category, nine figure-ground images were used. All of these depict one object on a monotonous background. For the high complexity category, nine scene images without a clear figure-ground composition were used. This approach follows Bradley et al. (2007) and Bradley et al. (2011), where example images can be found. Since pictures from the IAPS database are not intended for publication, original images were not included.

In Study b), 36 website screenshots were used as stimuli. These were created from real websites from the three categories news, company websites and online-shops. These categories were decided upon according to Roth et al. (2010), who analysed the 100 most visited websites of Germany, Austria, Switzerland and the USA and extracted these as the most popular categories next to social networks. The latter were excluded, since the typical design of social network starting or home pages offers little room for experimental manipulation, often offering only the option to login, while varying strongly between different sites (Roth et al., 2010). During the selection, care was taken to include only relatively unknown websites in order to prevent possible effects of familiarity. Screenshots in a resolution of 1280 x 720 pixels were created for initially 72 websites of the three categories using the browser plugin “FireShot” (GetFireShot, 2021). From these, 36 were picked and assigned to one of the three groups low, medium and high visual complexity according to a pre-rating of four separate subjects. Two examples for stimuli from the category online-shop are depicted in Figure 35 and

[2] IAPS pictures used in this study were: Figure-Ground - Neutral: 6150, 7010, 7056, 7110, 7150, 7175, 7190, 7211, 7950; Scene - Neutral: 5120, 5455, 5731, 7234, 7496, 7510, 7560, 7590, 7595

Figure 36. The assignment table with all stimuli and categories can be found in appendix 9.9**.**

Figure 35. Company website – low visual complexity

Figure 36. Company website – high visual complexity

4.3.3.2 Questionnaires

Both studies used the same questionnaires in order to assess ratings from the participants. Within the first part, stimuli should be assessed by the subjects with regard to a number of potentially relevant related constructs as well as the possible influencing variables identified within the previous literature research. For the ratings, 7-point Likert scales ranging from very low (German: “sehr gering”) until very high (“sehr hoch”) were used, where only the extrema were labelled. With the help of these, the items visual complexity (“Visuelle Komplexität”), liking (“Gefallen”), interest (“Interesse”), familiarity with contents (“Vertrautheit mit Inhalten”) should be judged before addressing the seven influencing variables. Among these were the number of elements (“Anzahl an Elementen”), variety of elements (“Vielfalt an Elementen”), density of elements (“Dichte der Elemente”), colour variety (“Farbvielfalt”), colour contrast (“Farbkontrast”), organization (“Ordnung”), symmetry (“Symmetrie”) and visual balance (“Visuelle Balance”). The item unintelligibility of elements (“Unterscheidbarkeit der Elemente”) was used within study a) but later removed from the further analysis (as described in paragraph 4.4) and excluded within study b). In study b), subsequently the short version of the visual aesthetics of websites inventory (VisAWI) (Moshagen & Thielsch, 2013) consisting of four items was administered in order to assess the aesthetical appraisal of these. Moreover, the prototypicality of the websites was assessed using the item “The website looks like a typical website” (“Die Webseite sieht wie eine typische Webseite aus”). Within the following, not all acquired ratings are reported for the sake of brevity.

4.3.4 Statistical Analysis

For the statistical analysis of both studies, regressions were used for the examination of the relations between potential influencing variables and a global visual complexity rating. Due to the ordinal data structure resulting from the use of Likert scales, cumulative link mixed models were used for the analyses with the help of the R-package *ordinal* (Christensen, 2018) and the function clmm. These mixed models allowed for the integration of random effects for both subjects and stimuli, which were included with random intercepts. The advantages of using mixed models are described in more detail within paragraph 3.3.4. Analyses of variance were then conducted in order to examine relations of the influencing variables with global complexity ratings using the Anova function of R-package *car* (Fox & Weisberg, 2019).

For the analysis of the factorial structure of the influencing variables, factor analyses were used. The number of factors was determined using Horn's (1965) parallel analysis by means of the R package *psych* (Revelle, 2018) using the function fa.parallel. Subsequently, factor analyses were computed using Varimax rotation with the fa function in the same package. The adequacy of the resulting models was then examined by the root mean square of residuals (RMSR), the root mean square error of approximation (RMSEA) and the Tucker-Lewis Index (TLI), which all showed appropriate values.

The regression analyses of factor scores on the global visual complexity rating was again implemented with cumulative link mixed models using the clmm function of R package *ordinal* (Christensen, 2018).

4.4 Results

Within the following, results from both partial studies are reported one after another. For each study, descriptive data are depicted first of all, before regression and factor analyses results are reported.

Study a) Photographs

Ratings of visual complexity are visualized in Figure 37 in order to depict the range and variety of ratings for the stimuli used.

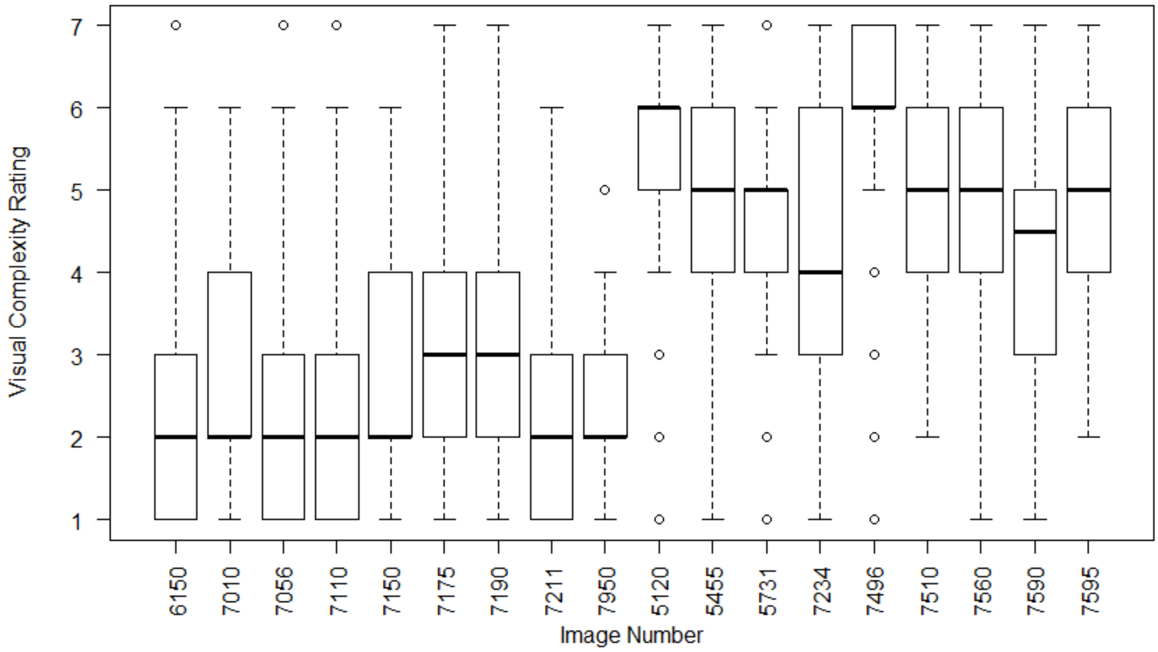


Figure 37. Boxplot of visual complexity ratings for all images used. (The nine images on the left, 6150-7950, are figure-ground images, while the nine images on the right, 5120-7595, are scene images)

Within the following, the relation between potential influencing variables, which had been identified within the previous literature research, and global complexity ratings are reported. Therefore, cumulative link mixed-model regressions were used. The detailed results of the regression model are presented in the appendix 9.7 in order to improve readability due to the large size of the table caused by the ordinal scaling of the predicting influencing variables. The ordinal regression model of all potential influencing variables on visual complexity ratings gave a marginal R^2 of .45 and a conditional R^2 of .60. Results from the subsequent analysis of variance, showing the statistical significance of all predictors are depicted in Table 3.

Table 3.

Analysis of variance for influencing variables

Predictors	**χ^2**	***p***
Number of elements	$\chi^2(6) = 44.57$	$\boldsymbol{p < .0001}$
Variety of elements	$\chi^2(6) = 43.78$	$\boldsymbol{p < .0001}$
Density of elements	$\chi^2(6) = 35.81$	$\boldsymbol{p < .0001}$
Variety of colours	$\chi^2(6) = 22.89$	$\boldsymbol{p < .0001}$
Colour contrast	$\chi^2(6) = 13.06$	$\boldsymbol{p < .05}$
Organization	$\chi^2(6) = 9.78$	$p = .13$
Symmetry	$\chi^2(6) = 17.89$	$\boldsymbol{p < .01}$
Visual Balance	$\chi^2(6) = 5.43$	$p = .49$

Furthermore, the dimensionality of the various influencing variables was investigated within a factor analysis. Therefore, first of all, the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (MSA) was examined in order to assess of the suitability of the data. The overall MSA was 0.81 (‘meritorious’ according to Kaiser, 1974), while the MSAs for all single scales were above 0.7, suggesting that the data allow for the use of exploratory factor analysis (Hutcheson & Sofroniou, 1999). Additionally, the significant Bartlett’s test ($p < .001$) suggests that the correlation matrix is not an identity matrix and there are relations between the variables. According to the subsequent Horn’s parallel analysis, a solution with three factors was selected. This was also in line with a visual analysis of the scree plot. The item unintelligibility of elements (“Unterscheidbarkeit der Elemente”) was eliminated because it did not contribute to a simple factor structure and failed to meet a minimum criterion of a primary factor loading of .4 or above.

Finally, a factor analysis with varimax rotation was conducted. Factor loadings are visualized in Figure 38. Additionally, a table with all factor loadings can be found in the appendix 9.8.

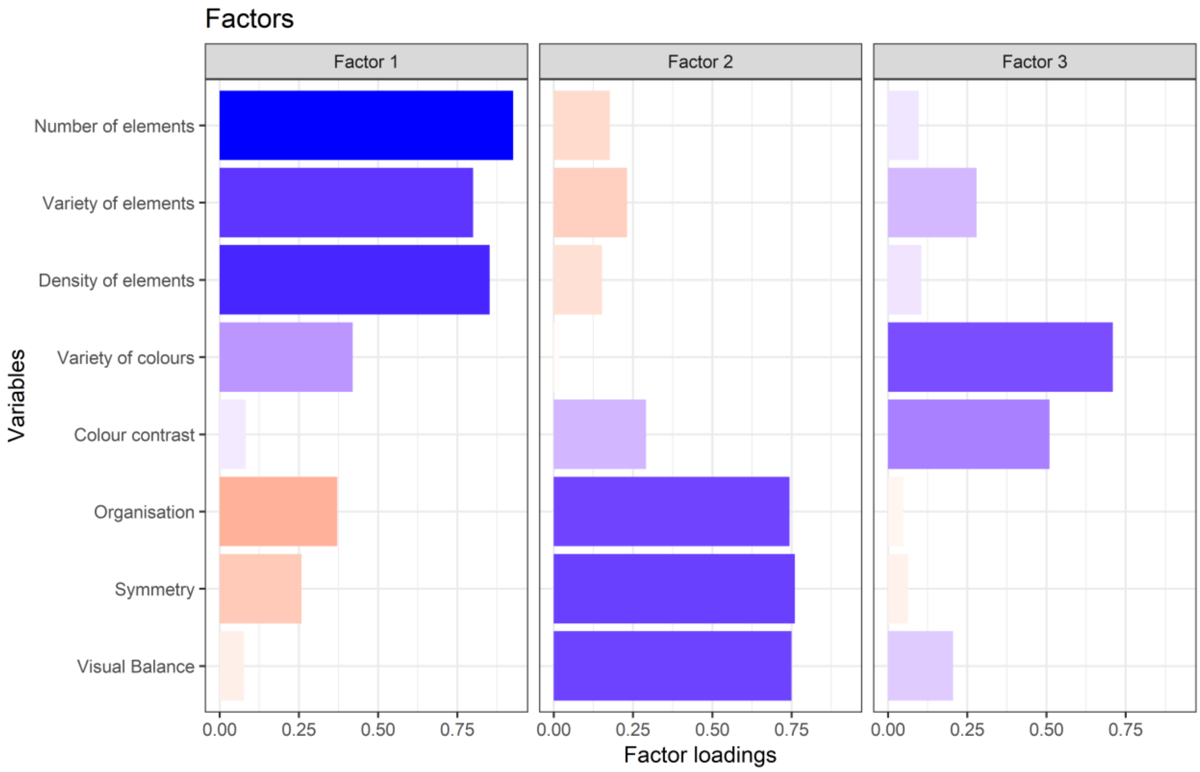


Figure 38. Factor loadings for three factor solution with Varimax rotation in study 2a)

For this factor structure, the adequacy of the model was evaluated. With a root mean square of residuals (RMSR) of 0.01, a root mean square error of approximation (RMSEA) of 0.045 and a Tucker-Lewis Index (TLI) of 0.99, the model seems adequate. Factor 1 was labelled “quantity” due to the high loadings of the items number of elements, variety of elements and density of elements. This factor explained a variance of 33%. The second factor was labelled “structure” because of high loadings of items organization, symmetry and visual balance. This factor accounted for 24% of the total variance. The third and last factor was labelled “colour” because primarily the two items variety of colours and colour contrast loaded on it. This factor explained 11% of the total variance.

The previously reported factor solution emerged from the relation of hypothesized influencing variables. In order to investigate the relation between the three factors and the global visual complexity ratings, finally an ordinal regression of factor scores on complexity ratings was computed. This showed significant positive relations between complexity ratings and factors quantity and colour as well as a negative relation with the factor structure as depicted in Table 4. The ordinal regression model of factor scores on visual complexity ratings gave a marginal R^2 of .43 and a conditional R^2 of .58.

Table 4.

Regression of factor scores on global visual complexity ratings

Predictors	Estimate	*SE*	*z*	*p*
Factor 1 (Quantity)	1.70	0.09	18.82	***p* < .0001**
Factor 2 (Structure)	-0.63	0.07	-8.68	***p* < .0001**
Factor 3 (Colour)	0.53	0.08	6.64	***p* < .0001**

Study b) Website Screenshots

Ratings of visual complexity for the website screenshots used as stimuli are depicted in Figure 39 in order to visualize their range and variety.

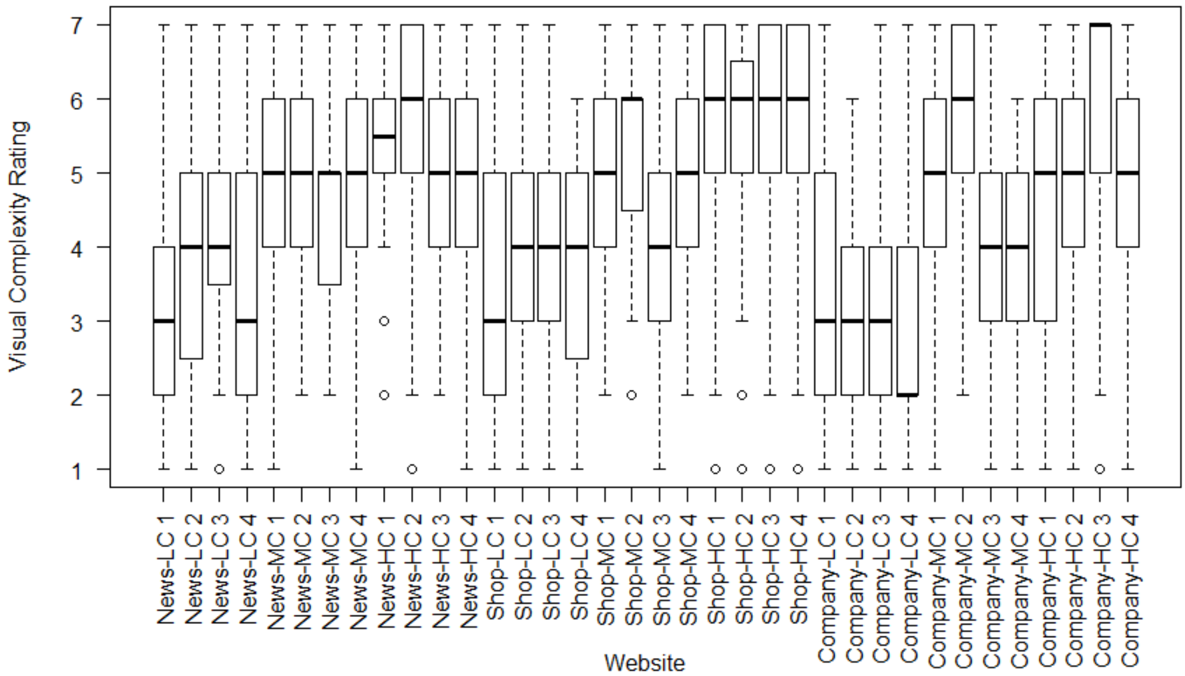


Figure 39. Boxplot of visual complexity ratings for all website screenshots used. The label describes the category (news, online-shops or company sites) as well as the

complexity level (high complexity – HC, medium complexity - MC or low complexity - LC) and number of the stimulus. The assignment table can be found in appendix 9.9.

As in study 2a), cumulative link mixed model regressions were calculated in order to investigate the relation between ratings of potential influencing variables and the global complexity ratings. The detailed results table is presented in the appendix 9.10 in order to improve readability. The ordinal regression model of all potential influencing variables on visual complexity ratings gave a marginal R^2 of .49 and a conditional R^2 of .60. Results from the subsequent analysis of variance, showing the statistical significance of all predictors are depicted in Table 5.

Table 5.

Analysis of variance for influencing variables

Predictors	χ^2	p
Number of elements	$\chi^2(6) = 147.48$	$\boldsymbol{p < .0001}$
Variety of elements	$\chi^2(6) = 17.42$	$\boldsymbol{p < .01}$
Density of elements	$\chi^2(6) = 50.48$	$\boldsymbol{p < .0001}$
Variety of colours	$\chi^2(6) = 11.61$	$p = .07$
Colour contrast	$\chi^2(6) = 5.69$	$p = .46$
Organization	$\chi^2(6) = 22.50$	$\boldsymbol{p < .001}$
Symmetry	$\chi^2(6) = 19.06$	$\boldsymbol{p < .01}$
Visual Balance	$\chi^2(6) = 12.19$	$p = .06$

As a next step, the dimensionality of the preselection of all potential eight influencing variables was investigated within a factor analysis. Within this partial study, the overall MSA was 0.76, while the MSAs for all single scales were 0.66 or above, suggesting that the sample is appropriate for performing factor analysis (Hutcheson & Sofroniou, 1999). Moreover, a significant ($p < .001$) Bartlett's test suggests that there are relations between the variables and therefore the correlation matrix is not an identity matrix. Like in the previous partial study, a solution with three factors was selected according to

Horn's parallel analysis and also in line with a visual analysis of the scree plot. Subsequently, the factor analysis with varimax rotation was conducted. Factor loadings are visualized in Figure 40. Additionally, a table with all factor loadings can be found in the appendix 9.11.

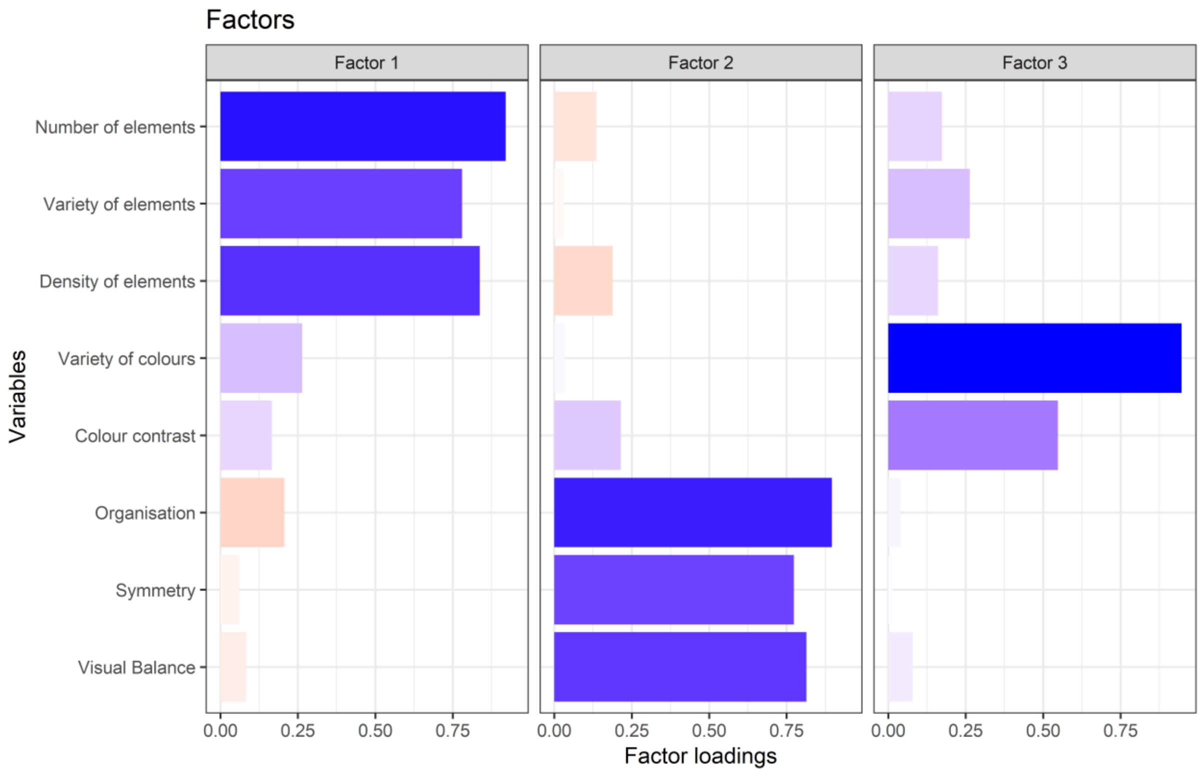


Figure 40. Factor loadings for three factor solution with Varimax rotation in study 2b)

With a root mean square of residuals (RMSR) of 0.01, a root mean square error of approximation (RMSEA) of 0.058 and a Tucker-Lewis Index (TLI) of 0.98, the model seems adequate.

The resulting factor structure was similar to study 2a). Factor 1 ("quantity") with high loadings of the items number of elements, variety of elements and density of elements, explained a variance of 29%. Factor 2 ("structure") with high loadings of items organization, symmetry and visual balance, accounted for 27% of the total variance within this study. Finally, factor 3 ("colour") with primary loadings of the items variety of colours and colour contrast, explained 17% of the total variance.

At last, the relation between factor scores and the global visual complexity was investigated using an ordinal regression. As in study 2a), this showed a significant positive effect of factor 1 ("quantity) and factor 3 ("colour") as well as a negative effect of factor

2 (“structure”) on complexity ratings as depicted in Table 6. The ordinal regression model of factor scores on visual complexity ratings gave a marginal R^2 of .45 and a conditional R^2 of .57.

Table 6.
Regression of factor scores on global visual complexity ratings

Predictors	Estimate	*SE*	*z*	*p*
Factor 1 (Quantity)	1.83	0.07	25.94	$\boldsymbol{p < .0001}$
Factor 2 (Structure)	-0.38	0.06	-6.68	$\boldsymbol{p < .0001}$
Factor 3 (Colour)	0.39	0.05	7.56	$\boldsymbol{p < .0001}$

4.5 Discussion

This study was conducted in order to achieve a better understanding of the construct visual complexity as well as the impact of different influencing variables. Within two partial studies, photographs and website screenshots were used as different domains of stimuli as to draw conclusions about the generalizability of the construct. For the study, a number of potential influencing variables had been identified within a previous literature research. Subjects then rated the presented stimuli with regard to these as well as for global visual complexity. The acquired data was analysed first of all with regard to the relation between potential influencing variables and the global visual complexity rating. Moreover, the factorial structure of influencing variables was examined by means of factor analyses. Finally, the relation between factor scores and global visual complexity ratings was investigated within regressions in order to check for the significance of the dimensions with regard to visual complexity.

First of all, results from the regression with potential influencing variables for photographs revealed that the number of elements, variety of elements, density of elements, variety of colours, colour contrast and symmetry were significantly related to the global visual complexity rating. On the other hand, both organisation and visual balance did not show a significant relation. For website screenshots, number of elements, variety of elements and density of elements again showed a significant relation. However, colour contrast was apparently not significantly related to the global rating while variety of colours was marginally significant with $p = .07$. Opposed to study 2a), organisation showed a clear relation with visual complexity for website screenshots while symmetry was significantly related as in the first partial study. Finally, visual balance gained marginal significance for website screenshots with $p = .06$.

Comparing both partial studies, regressions with potential influencing variables revealed some similarities but also differences. The two colour-related factors variety of colours and colour contrast appear to have a larger impact on photographs than on website screenshots, while both organisation and visual balance (even though the latter was only marginally significant) appeared to have a larger effect for website screenshots. In hindsight, this may not appear too surprising. It might be argued that structural aspects such as the organisation and visual balance are of larger relevance for the visual complexity of user interfaces, which is underlined by the evidently negative relation between visual complexity and organisation as reported in appendix 9.10. Within less complex websites, items might for example be arranged along a grid following a clear structure. This clear structure may however be less present within photographs of natural scenes. Here, colour as a natural feature may have a larger impact. Since the variety of colours and colour contrast are not primarily design-features within IAPS-images, these may point towards a higher degree of visual clutter and thus serve as a proxy for visual complexity. This is in line for example with the evidently positive relationship between the variety of colours and visual complexity as reported in appendix 9.7.

Within the factor analyses for both partial studies, very similar three-factor solutions could be identified for both photographs as well as for website screenshots. Thereby, a first “quantity”-factor received high loadings from the items number of elements, variety of elements and density of elements. It is labelled quantity since the number of elements loaded highest on this factor within both partial studies. The two other items variety and density of elements are obviously strongly related to this. This retrospectively seem very plausible, since it is likely that the density and variety of elements increase with the number of elements, although there may be instances where this is not the case. The second factor was labelled “structure” due to the high loadings of symmetry, organisation and visual balance, which contribute to similar parts. As a post-hoc explanation for this finding, it can be quoted that symmetry and visual balance are strongly related (Hübner & Fillinger, 2016). Moreover, symmetry facilitates perceptual grouping (Wagemans, Elder et al., 2012), which may be hypothesized to apply to visual balance as well. Finally, the rating of organisation (or “Anordnung” in German) may similarly encompass several aspects that fall beyond the principles of perceptual grouping such as proximity or common fate (Wagemans, Elder et al., 2012). However, no further instructions or explanations regarding this item were given to the subjects, therefore this can only be assumed. Finally, the third factor was labelled “colour” due to the high loadings of the items variety of colours and colour contrast. All of the three factors are of relevance for the perception of visual complexity as suggested by the significant relations between factor scores and global visual complexity ratings.

The identified factorial structure is mainly in line with previous findings from experimental research literature. Similar to Chipman (1977) and Ichikawa (1985), it emphasizes the meaning of both a quantitative as well as a structural factor of visual complexity, while additionally bringing up a third, colour-related factor. This might not have played a major role within previous investigations, since these in some cases used black and white stimuli, which also applies for both Chipman (1977) and Ichikawa (1985). Other researchers classified colour within a quantitative factor (Nadal et al., 2010) or as related to variety (Miniukovich et al., 2018). Yet, the findings from this study clearly suggest colour as an independent factor.

The factor analysis results clearly contradict those revealed by Nadal et al. (2010). They similarly identified three factors, however with a very different structure of factor loadings, which makes the interpretation rather difficult. While in their work, unintelligibility of elements and disorganization loaded highly on a second factor, asymmetry loaded heavily on a third factor. This is also opposed to experimental findings by Chipman (1977) and Ichikawa (1985), that suggest a two-factorial structure with a quantitative and a structural factor. The structural factor is typically found to encompass both aspects of symmetry and organisation (Chipman & Mendelson, 1979). Eventually, the surprising factor structure identified by Nadal et al. (2010) can in part be related to the negative formulation of items (e.g. asymmetry, disorganization, unintelligibility of elements), which may have confused participants (Colosi, 2005). Depending on the scale labelling (for example from “completely disagree” to “completely agree”), this might have led to a potential double negation, which may be an aspect that contributed to findings of a different factorial structure. In conclusion, the results of this study showed the same factorial structure of visual complexity within the two different stimulus domains photographs and website screenshots, based on the findings from both partial studies. This suggests that the construct is universal instead of domain-specific and contributes to a better understanding and definition of the construct visual complexity. Finally, within regressions of factor scores on global visual complexity ratings, the impact of the three factors on the perception of visual complexity was investigated. As for the factor analysis, results were very similar between both partial studies, showing significant relations with all three factors. While the “quantity”-factor showed a positive relation like the “colour”-factor, a negative relation was found for the “structure”-factor. Taken together with the loadings of the factor analysis, this suggests that a higher number, variety and density of elements is related to a higher level of visual complexity. This also applies for a higher variety of colours and colour contrast. Moreover, a larger degree of symmetry, organisation and visual balance goes along with a lower level of visual complexity. This again seems plausible and is in line with findings from previous research (e.g. Chipman & Mendelson, 1979; Harper, Jay, Michailidou, & Quan, 2013). Importantly, the factor scores accounted for a considerable proportion of the total variance of global complexity ratings (58% for photographs and 49% for website screenshots). This is especially remarkable when compared to the variance explained by all influencing factors (61% and 52%, respectively).

Limitations

However, there are of course some limitations. First of all, photographs and website screenshots were investigated within this study. While these revealed robust results with regard to the factorial structure and relations between factor scores and global visual complexity ratings, this of course does not mean that similar results will necessarily show for all other stimulus domains. It remains to be investigated if the identified factorial structure also holds true in contexts such as driving or for artworks.

Moreover, all findings within this study are based on questionnaire data and thus provide a good understanding of subjective perceptions. However, these have not been complemented with more “objective” measures such as performance parameters or ocular and computational measures yet. These however may complete the picture by allowing further insights into cognitive processes associated with the perception of visual complexity.

What remains unclear is the understandability of items. While these were selected carefully based on existing research literature, no further explanation was given as part of the instruction of this study. This was intentionally avoided, so that subjects were not influenced in a certain way. Instead, they should make their unbiased judgements for the different items based on their understanding of these. This approach necessarily produces the possibility that subjects interpreted items such as visual balance or unintelligibility of items differently. While the latter was excluded from the further analyses because it did not show a primary factor loading of .4 or above, all other items however seemed appropriate for factor analysis as suggested by decent MSAs. Since these rely on the shared variance of items, this would likely not be the case if items were not understandable, which might rather have gone along with either no variance or random ratings.

As a last aspect, the presentation durations of stimuli were not controlled within the online surveys. The reason for this was to allow subjects to closely inspect the pictures while making their judgements. This however also means that subjects might have spent more time observing for example the more complex images. This could then have affected the subjective rating. In this regard, Cardaci, Di Gesù, Petrou, and Tabacchi (2009) for example showed a relation between visual complexity and perceived presentation duration, while Palumbo, Ogden, Makin, and Bertamini (2014) could find no such relation. Further studies may however be helpful in order to address the possible relation between visual complexity and perceived presentation duration, which might possibly influence subjective ratings.

Outlook

First of all, consequent works can help to examine the validity of the identified factor structure as well as the relevance of influencing variables within other domains. For example, it might be investigated if the findings on visual complexity from this study can be transferred to the context of driving, where visual complexity can be highly relevant (e.g. Horberry et al., 2006). Despite its effects on lane keeping performance for example (Horberry et al., 2006), few investigations have so far more closely addressed the role of visual complexity within this context.

Moreover, the findings from this study can be further substantiated. Particularly, in order to systematically examine the impact of the identified influencing variables on visual complexity beyond correlational analyses, experimental research designs could be used. Within these, single dimensions could be systematically manipulated in order to assess their effects on visual complexity ratings. Moreover, additional measures might be integrated in order to allow for further insights and a better understanding of the cognitive processes associated with the perception of visual complexity. In particular, ocular parameters from eye tracking will likely provide further insights into attentional mechanisms. Moreover, computational measures could be used for the investigation of relations and thereby, for example in combination with ocular parameters, for the prediction of visual complexity ratings. This approach will be implemented within the next two studies.

5. Study 3: Foundations: Influencing variables of visual complexity and ocular parameters

While the previous study allowed for a detailed overview of the construct visual complexity as well as its relations with influencing variables, these have rarely been investigated systematically by means of experimental methodology. Within strongly controlled conditions, this study investigates the effects of manipulations of relevant influencing variables both on subjective ratings as well as various ocular parameters. Moreover, it explores if eye tracking can contribute to the prediction of complexity ratings beyond computational measures.

5.1 Background

According to Standish (2008), both a quantitative as well as a qualitative or structural aspect are essential for complexity in general. Findings of researchers such as Ichikawa (1985) and Chipman (1977) suggest a similar image of quantitative and structural aspects specifically with regard to visual complexity (for an extensive overview of research literature see also paragraph 2.2.3). The importance of both factors for the perception and processing of visual complexity was also supported within the previous study. However, their influence on global visual complexity ratings for pictures has rarely been directly addressed in experimental investigations before. These however allow for causal reasoning, going beyond the analysis of correlational associations. Even though Ichikawa's (1985) findings are based on an experimental approach, these do not allow for conclusions regarding global complexity judgements, but rather focussed on the underlying cognitive processes. In this regard, an experimental manipulation of selected influencing factors and the assessment of ratings contribute to a consolidation of findings. Based on Chipman (1977) and Ichikawa (1985) as well as the findings from study 2, the number of elements as well as symmetry were selected as the most important aspects for both the quantitative as well as the structural factor of visual complexity, respectively. These accounted for the largest part of variance in global visual complexity ratings within the previous study. The third factor “colour” was not considered within this study in order to focus on the two other factors and keep the experimental design simple.

Beyond the general advantages of an experimental study design, this approach furthermore allows for the analysis of various ocular parameters. On the one hand, these can significantly contribute to deeper insights into cognitive and especially attentional processes during the perception of visual complexity. On the other hand, they can complement rating data and thus provide a more comprehensive assessment of visual complexity and eventually contribute to the prediction of visual complexity ratings beyond computational measures. However, ocular parameters have only rarely been used in research on visual complexity to this date (see paragraph 2.5.3). Within the few studies incorporating eye tracking methodology, Bradley et al. (2011) and Madan et al. (2017) showed a larger number of fixations and longer scanpaths for more complex images. Scanpath length is typically calculated as the sum of all distances between fixations (Goldberg & Kotval, 1999). However, in research in the field of user interface design, the study of eye movements has been common for some time and has yielded interesting findings regarding the use of multiple parameters for interface evaluations (Goldberg & Kotval, 1999; Holmqvist & Andersson, 2017; Kotval & Goldberg, 1998). Some of these parameters may also be informative for the investigation of visual complexity in addition to subjective ratings. One of these parameters is spatial density, which can be used as an indicator of the extent of visual search in interfaces (Goldberg & Kotval, 1999). It serves as a measure for the spatial distribution of gaze and is calculated by dividing the interface or stimulus into a grid (for example 10 x 10) and assessing the number of cells containing at least one scanpath node divided by the total number of grid cells (Goldberg & Kotval, 1999). This is also visualized in Figure 41. While larger spatial distribution indicates extensive search, smaller values can point to a more directed search in interfaces. Generalizing the assessment of the spatial distribution of gaze beyond interfaces, the spatial density can potentially also be highly valuable in visual complexity research. Next to the number of fixations and the scanpath length, which were shown to relate to visual complexity (Bradley et al., 2011; Madan et al., 2017) and can, in the case of scanpath length, reflect search behaviour within user interfaces (Goldberg & Kotval, 1999), spatial density can complement this picture by integrating the spatial dimension of gaze distribution.

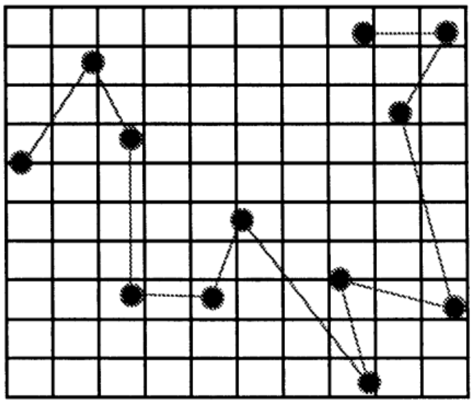


Figure 41. Spatial density visualization, taken from Goldberg and Kotval (1999). This would be a spatial density of 12/100.

While potentially reflecting effects of experimental manipulations, ocular parameters can also add to the prediction of visual complexity ratings beyond computational measures. The latter have been used for the prediction of mean complexity ratings within previous research (e.g. Gartus & Leder, 2017; Tuch et al., 2012, see also paragraph 2.4). Beyond these, ocular parameters can extend the possibilities for prediction in two ways. First of all, both volitional (top-down) as well as automatic (bottom-up) aspects of attention can influence gaze behaviour (see discussion in paragraph 2.5.1 and 2.5.2), which may also be reflected in ocular parameters. Secondly, parameters are calculated on trial-level, thus for each picture and subject individually. As opposed to computational measures, which are picture-specific, this allows the consideration of intraindividual differences. This variance between persons, which may relate to different experience, expectations, motivation or personality, might also affect gaze patterns as well as pupillometry. Thus, a higher level of detail can be achieved within the prediction. Therefore, the integration of ocular parameters appears beneficial in many respects.

In conclusion, within this study, I primarily used an experimental approach for the investigation of the most relevant influencing variables of visual complexity. The number of elements and symmetry were therefore selected, representing both the quantitative and the structural factor of visual complexity. In order to exclude possibilities of confounding as far as possible, abstract shape patterns were used as stimuli. These offered the possibility of controlling the experimental factors without side effects of other aspects such as previous experience or individual preference. With this approach, the effects on both visual complexity ratings as well as ocular parameters are investigated. Moreover, computational as well as ocular parameters are used for the prediction of complexity ratings. In the following, research questions according to these goals are specified.

5.2 Research questions

1. What are the effects of manipulations of quantitative (specifically the number of elements) as well as structural (specifically symmetry) aspects in stimuli on visual complexity ratings?
2. What are the effects of manipulations of quantitative (specifically the number of elements) as well as structural (specifically symmetry) aspects in stimuli on the ocular parameters number of fixations, scanpath length and spatial density?
3. How well can mean and single visual complexity ratings be predicted from computational and ocular parameters?

5.3 Method

For the investigation of these research questions, a laboratory eye tracking experiment was conducted, which is described in detail in the following.

5.3.1 Participants

33 persons participated within the experiment, among them 20 male (60.6%) and 13 (39.4%) females with a mean age of 24.3 years (*SD* = 3.7). 27 of the participants (81.8%) were students, while the other 6 (18.2%) were employed. The study lasted approximately 30 minutes and participation was refunded with 5€. All participants had normal or corrected-to-normal vision.

5.3.2 Experimental design and procedure

In order to study the effects of quantity and structure as principle visual complexity factors, a repeated-measures 4 (number of elements) x 3 (symmetry) x 3 (type of element) experimental design was used. Within this, the first two factors were of substantial interest. The factor number of elements had four levels with either one, five, nine or 13 elements in a picture. Symmetry was adapted using the method suggested by Bauerly and Liu (2008) where symmetry is calculated as the similarity of pixels on opposite sides of an axis of reflection. Bauerly and Liu's (2008) original formula was adapted so that symmetries of white pixels were not considered in the calculation because of the white background. The mean value of horizontal, vertical and both diagonal axes served as overall symmetry measure. The stimulus pictures were adjusted in order to match one of the three levels *no symmetry* ($s = 0$), *medium symmetry* ($0.35 < s < 0.65$) or *perfect symmetry* ($s = 1$). Finally, as type of element, either dots, squares or crosses were used. This factor was included to increase variation within stimuli although there was no theory-based hypothesis concerning this factor. This also holds true for the two different arrangements of elements that were used for each factor level. An overview over the experimental design is given in Figure 42.

Independent variables:		
Number of elements	Symmetry	Type of element
• 1 element • 5 elements • 9 elements • 13 elements	• no symmetry • medium symmetry • perfect symmetry	• dot • square • cross

Dependent variables:
• Subjective ratings of visual complexity • Ocular parameters: Number of fixations, scanpath length, spatial density

Figure 42. Experimental design for study 3

Participants were first of all instructed that they could abort the experiment at any time without consequences as well as that data was stored anonymously and used for research purposes before they gave written informed consent. The experiment then started with instructions for the participants, which explained that their task would be to carefully watch images with shape patterns before giving ratings for each image directly after its presentation. After the subsequent calibration of the eye tracker, the experimental procedure was started using *OpenSesame 3.2.6* (Mathôt, Schreij, & Theeuwes, 2012). First of all, participants could get accustomed with the experimental procedure with three practice trials containing stimuli similar to the experimental images. The following main part of the experiment comprised the randomized presentation of 72 stimulus pictures, which are described in detail within paragraph 5.3.3. Each picture was presented for 6000ms, preceded by a fixation cross and followed by a short questionnaire, which will be described in paragraph 5.3.3. All ratings were self-paced. After 36 trials, there was a break that subjects could use to move or relax.

5.3.3 Material

Stimuli

For the experiment, a stimulus set consisting of 72 black and white images with shape patterns was created according to the experimental design[3]. Thus, each combination of the factor levels number of elements, symmetry and type of element is represented by two images with different arrangements of elements. The images were exported as GIF-files with a resolution of 1024 x 1024 pixels. Examples of pictures are shown in Figure 43.

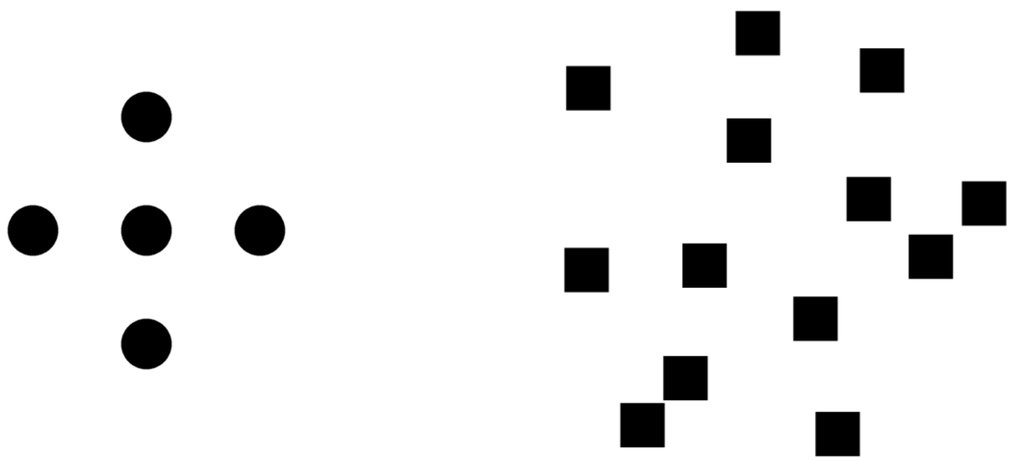


Figure 43. Two examples for stimulus images in study 3; left side: nine elements, perfect symmetry, dots; right side: 13 elements, no symmetry, square.

For all stimuli, the range of computational measures described in paragraph 2.4 was calculated except colour-related measures, since only black and white stimuli were used. This resulted in a 150 measures and combinations of measures in total.

Questionnaire

Within the questionnaire following each experimental trial, participants were asked to rate the visual complexity as well as the liking of the picture in 7-point scales ranging from 1 = “very low” to 7 = “very high”.

[3] Stimulus creation was accomplished in close collaboration with master student Sibylle de Vandière

Apparatus

All stimuli were displayed on a 24-inch LCD monitor (LG 24MB56HQ; display dimension = 52.69 cm × 29.64 cm; resolution = 1920 × 1080 pixels; refresh rate = 60 Hz). The eye-to-screen distance amounted to approximately 98 cm. Picture size was 1024 × 1024 pixels, thus corresponding to a visual angle of 16.3° × 16.3°.

For the assessment of eye movements, a video-based SR Research Eyelink 1000 Plus eye tracker was used. Monocular eye position data of the dominant eye were sampled at 2000 Hz. It was used with a desktop mount for screen-based eye tracking, while the participants' heads were stabilized in a chin-rest in order to achieve a higher accuracy. After the detection of blinks, saccades and fixations were detected within the software Data Viewer (SR Research, 2019), using the standard configuration that classifies an eye movement as a saccade when it exceeds 30°/s velocity or 8000°/s^2 acceleration. The time intervals between saccades were defined as fixations. From the assessed data, all ocular parameters described in paragraph 2.5.3 were calculated for each trial, resulting in 46 measures in total.

5.3.4 Statistical Analysis

Data was analysed in the software R (R Core Team, 2017). For the statistical analysis of ocular parameters, linear mixed effects models were fitted within the R-package *lme4* (Bates et al., 2015). For the statistical analysis of visual complexity ratings from a seven point Likert scale, cumulative link mixed models implemented in the R-package *ordinal* (Christensen, 2018) were used. In both cases, number of elements and symmetry were entered as fixed effects while as random intercepts for subjects, stimuli and object types were used. Factor level "1 element" was treated as a control condition and excluded from statistical analyses because symmetry manipulations within this condition with one element were not assumed to have actual influence on their perception. All ratings are however depicted within the figures for the sake of completeness. Main and interaction effects of the factors were analysed within likelihood ratio test using the chi square distribution by means of the Anova function from package *car* (Fox & Weisberg, 2019) as well as post-hoc tests in package *emmeans* (Lenth, 2019). Arguments for the use of mixed effects models are discussed in detail by Judd et al. (2012) and within paragraph 3.3.4. Result plots were created with the help of the R-package *ggplot2* (Wickham, 2016). All error bars within the plots depict the 95% confidence interval. The procedure for the prediction of mean and single complexity ratings is described in detail within the subsequent results to improve readability.

5.4 Results

Within the following, the results of this study are reported. First of all, the effects of the described experimental design on visual complexity ratings as well as three ocular parameters are investigated. This allows for reliable causal conclusions regarding the impact of selected influencing variables on the perception of visual complexity.
Subsequently, prediction models for mean and single visual complexity ratings are examined. These rely on both computational and ocular parameters.

5.4.1 Visual complexity ratings

First of all, the number of elements in a picture had a significant effect on the rating of visual complexity, $\chi^2(2) = 120.55$, $p < .0001$. Furthermore, the factor symmetry also significantly affected the rating of visual complexity, $\chi^2(2) = 96.89$, $p < .0001$. Finally, the interaction between number of elements and symmetry also affected the rating of visual complexity, $\chi^2(4) = 10.14$, $p = <.05$. Due to the ordinal type of interaction, this does not restrict the interpretability of the two main effects.
The marginal R^2 was .34 and the conditional R^2 was .56 for the underlying regression model, which is reported in more detail within appendix 9.12. Rating scores are visualized in Figure 44.

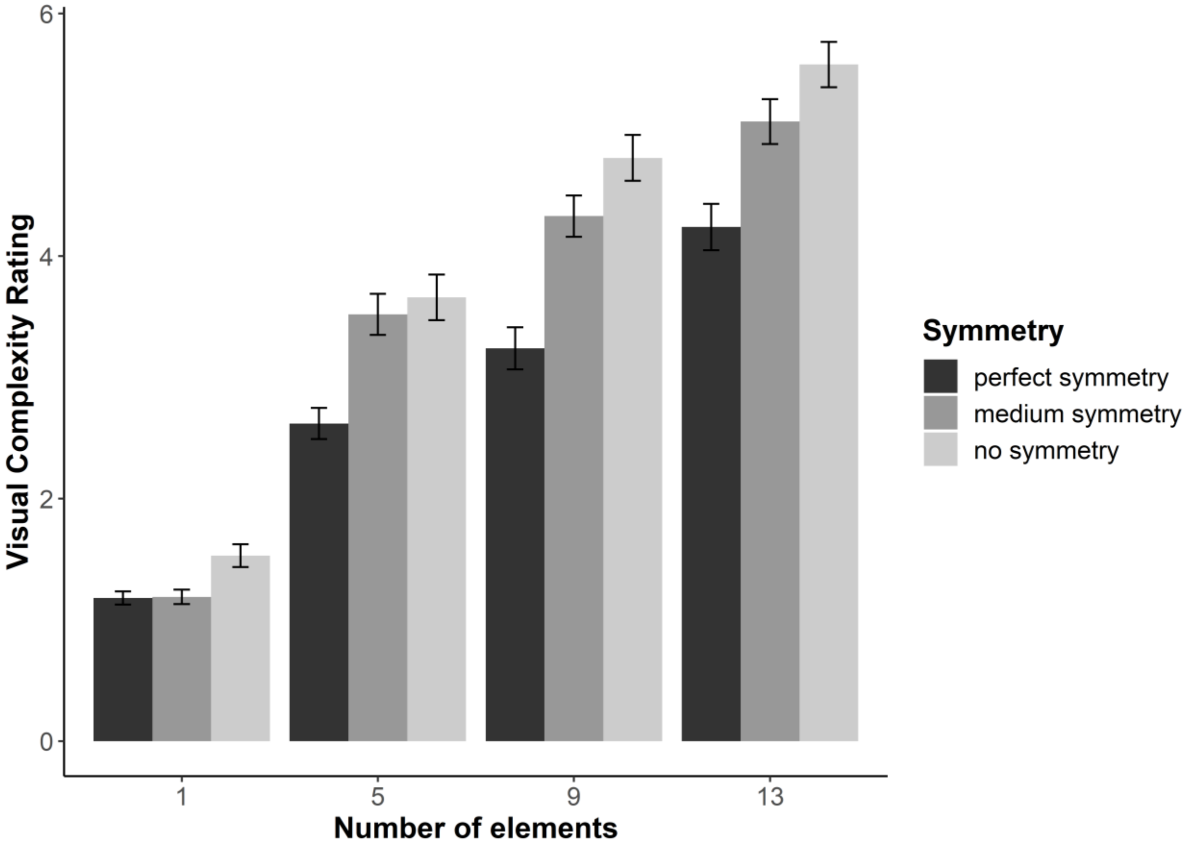


Figure 44. Subjective ratings of visual complexity in study 3

5.4.2 Ocular parameters

Number of Fixations

First of all, the number of elements in a picture had a significant effect on the number of fixations, $\chi^2(2) = 126.80$, $p < .0001$. Tukey post hoc tests showed that the number of fixations was higher for pictures with 13 elements compared to pictures with nine, $\beta = 1.24$, $SE = 0.23$, $p < .0001$, and five elements, $\beta = 2.53$, $SE = 0.23$, $p < .0001$. Similarly, the number of fixations was higher for pictures with nine elements compared to pictures with five elements, $\beta = 1.29$, $SE = 0.23$, $p < .0001$.

Moreover, the factor symmetry had a significant effect on the number of fixations, $\chi^2(2) = 41.39$, $p < .0001$. Tukey post hoc tests revealed that the number of fixations was higher for asymmetrical compared to symmetrical pictures, $\beta = 1.37$, $SE = 0.23$, $p < .0001$, while the difference between asymmetrical pictures and those with medium symmetry was not significant, $\beta = 0.29$, $SE = 0.23$, $p = .42$. Furthermore, the number of fixations was higher for pictures with medium symmetry than for symmetrical pictures, $\beta = 1.08$, $SE = 0.23$, $p < .001$.

The interaction between the number of elements and symmetry had no significant effect on the number of fixations, $\chi^2(4) = 7.43$, $p = .12$. Values for number of fixations are visualized in Figure 45. While the marginal R^2 was .05, the conditional R^2 was .56 for the underlying regression model. More information about this can be found within the appendix 9.13.

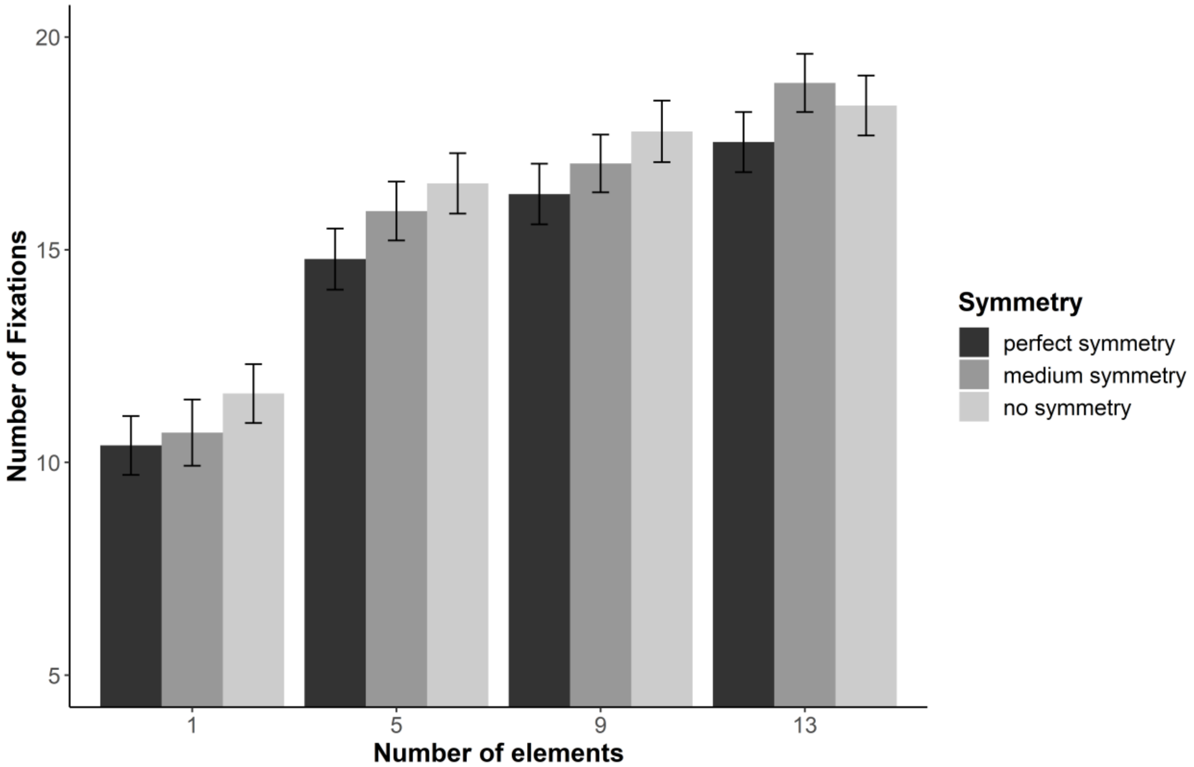


Figure 45. Number of Fixations in study 3

Scanpath Length

For the preprocessing of scanpath length data, trials with a scanpath length of zero pixels were primarily excluded from further analyses. Subsequently, the median absolute deviation (MAD), which is a robust approach for dealing with outliers (Leys et al., 2013), was used with a moderately conservative criterion of 2.5 to exclude remaining outliers. Within the linear mixed effects models, random intercepts were included for subjects and trials as for the other measures. However, type of object did not account for any variance and was therefore not included as a random intercept in order to avoid a singular fit of the model.

Results revealed that the number of elements, $\chi^2(2) = 80.19$, $p < .0001$, symmetry, $\chi^2(2) = 59.86$, $p < .0001$, and the interaction of the two factors, $\chi^2(4) = 88.66$, $p < .0001$, had a significant effect on the scanpath length in pixels. Scanpath lengths are depicted in Figure 46. For the underlying regression model, the marginal R^2 was .19, while the conditional R^2 was .55. More information about this can be found within the appendix 9.14.

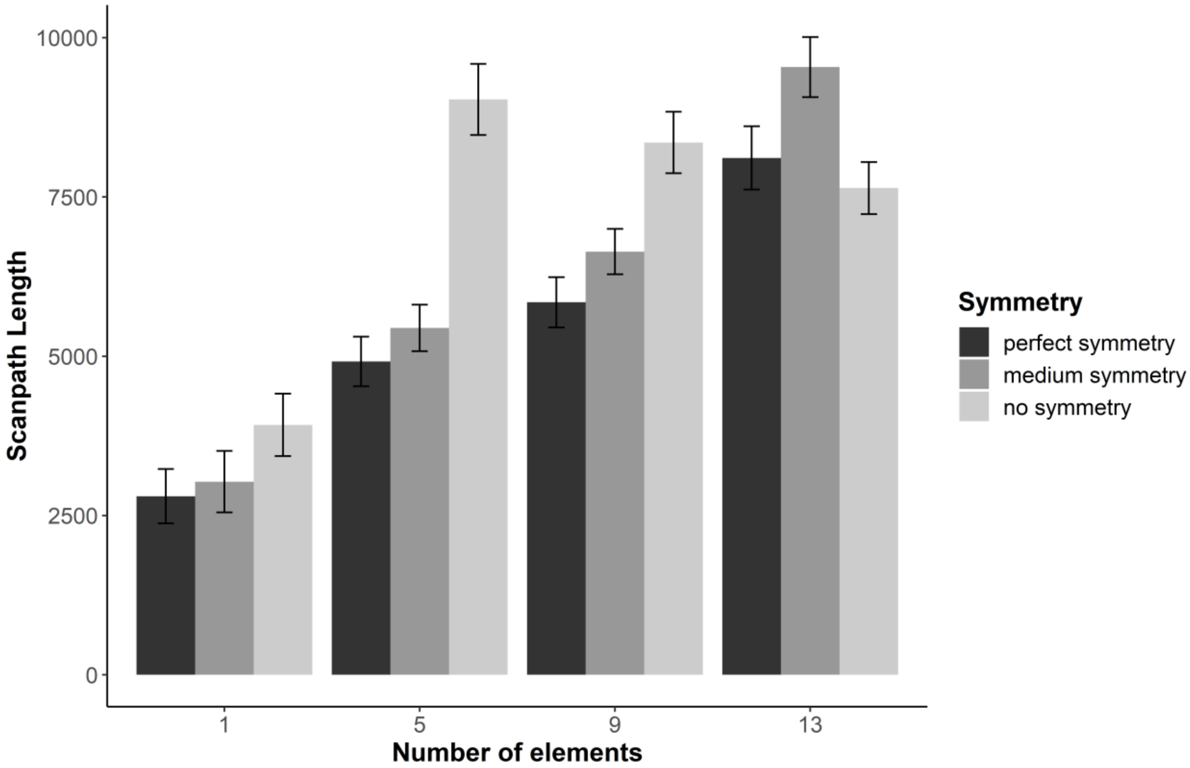


Figure 46. Scanpath length (in pixel) in study 3

Spatial Density

For the measure of spatial density, the number of elements, $\chi^2(2) = 280.35$, $p < .0001$, and symmetry, $\chi^2(2) = 23.22$, $p < .0001$, as well as the interaction of the two factors, $\chi^2(4) = 16.42$, $p < .001$, had a significant effect on the spatial density. The scores are illustrated in Figure 47. The marginal R^2 was .20 and the conditional R^2 was .53 for the underlying regression model. More information about this can be found within the appendix 9.15.

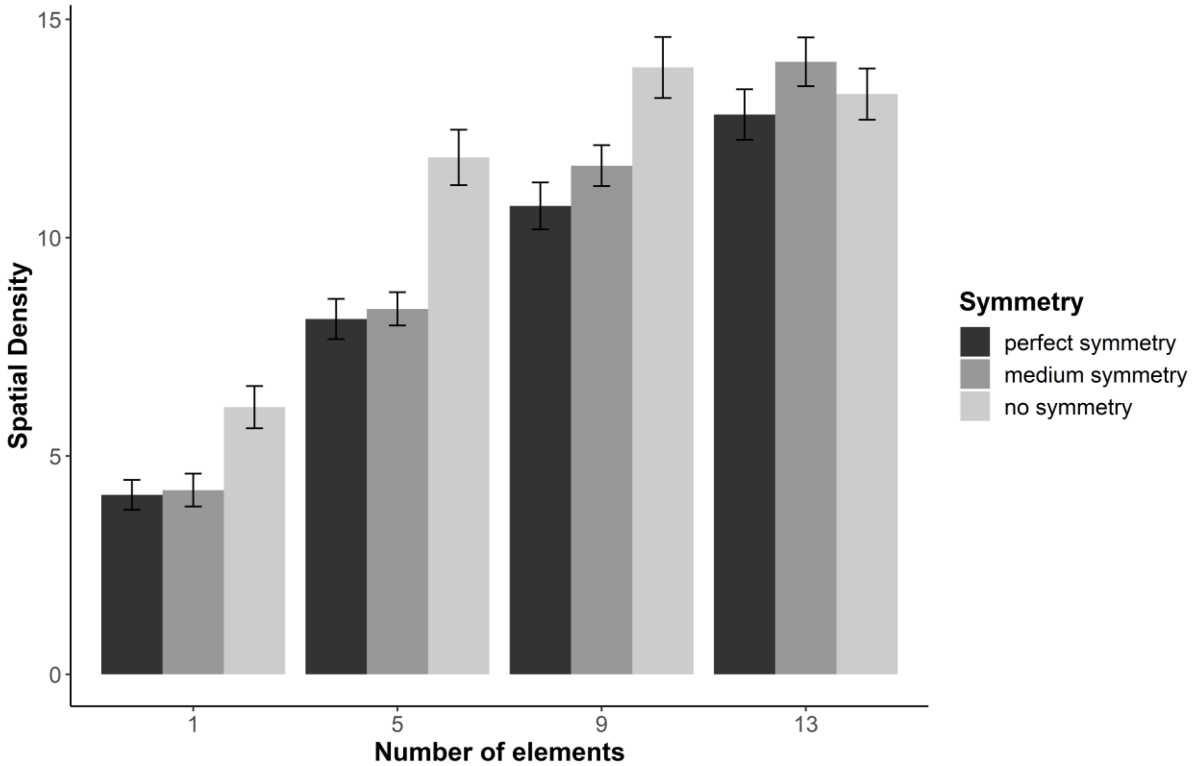


Figure 47. Spatial density (in percent) in study 3

5.4.3 Prediction of visual complexity ratings

Finally, both computational and ocular parameters were used for the prediction of visual complexity ratings. In the first step, the mean visual complexity ratings for each picture were addressed by means of a least absolute shrinkage and selection operator (LASSO) regression (Tibshirani, 1996). This is a form of a regularized regression that can be used to assess the combined effect of many potentially correlated variables. Within the regularization process, coefficients of the regression variables are penalized. Since a number of these are shrinked to zero, this allows for variable selection. The tuning parameter λ thereby controls the strength of the penalty. It is well suited for the application in cases with a relatively small number of observations and a large number of predictors and allows for good model interpretability while reducing overfitting.

Within the analysis, both computational (150) and ocular (46) measures served as potential predictors, resulting in a total number of 196 variables. First of all, mean visual complexity ratings as well as the mean of all ocular parameters were calculated for

each stimulus picture across all subjects, resulting in 72 values each. Since the computational parameters were analysed for each image, these are available on image-level per se and not on the trial-level. The computational measures are described in detail within paragraph 2.4, while the ocular measures are depicted in paragraph 2.5.3. Subsequently, one predictor with near zero variation was excluded, while the remaining predictors were centered and scaled. While the latter is necessary for LASSO regressions, both help to prevent scaling problems and improve interpretability. After that, the dataset was randomly split into an 80% training and 20% test set. This was done in order to compare prediction errors between both with regard to overfitting.
Consequently, a cross-validated LASSO regression model was fit with the help of the R-package *glmnet* (Friedman, Hastie, & Tibshirani, 2010). The penalty parameter λ was chosen based on the criterion of the mean-squared error (MSE) according to the "tolerance" model for a more parsimonious fit based on 100 penalty values (see Figure 48).

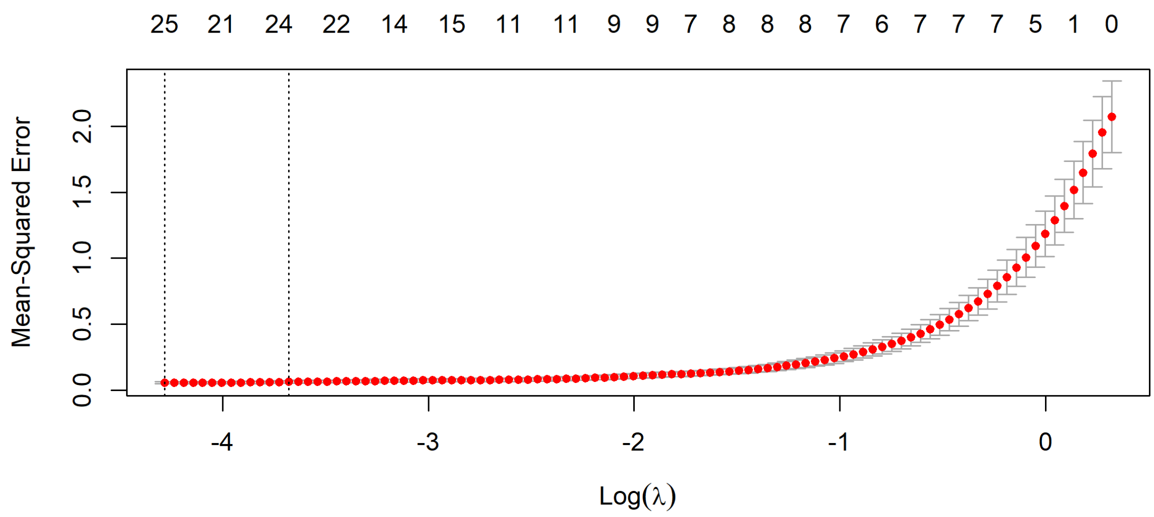


Figure 48. Mean-Squared Error (MSE) in relation to λ from lasso regression in study 3. The two dotted lines represent the optimal (left) and tolerance (right) fit lambda

24 predictors were selected, which are listed within Table 7. Of these, the upper five variables denote ocular parameters as described in more detail within paragraph 2.5.3. The others are computational parameters, among these combinations of compression (e.g. GIF, JPEG; TIFF) and edge (e.g. Perimeter, Canny) measures, including both

mean and standard deviations of the pixel distribution within an image. Other selected variables stem from decomposition measures such as the number of quads of different sizes, or comprise measures for symmetry, visual balance and homogeneity. Since the examination of statistical significance of LASSO predictors is based on strong assumptions (Lockhart, Taylor, Tibshirani, & Tibshirani, 2014), which are not testable when the number of predictors exceeds the number of observations as argued by Wasserman (2013), only the coefficient estimates are reported here.

Table 7.

Selected variables from lasso regression with coefficients for prediction of mean complexity ratings in study 3

Variable	**Coefficient estimate**
(Intercept)	4.40
Number of Blinks	-0.076
Number of Saccades	0.12
Average Drift	-0.027
Coefficient K	-0.0050
ICA drop	0.0098
TIFF filesize	0.22
Ratio TIFF filesize to image pixels	0.000035
Mean Edge Canny GIF	0.029
Mean Edge Perimeter GIF	0.000042
Mean Edge Perimeter TIFF	0.0000000000000000010
Mean Edge Perimeter PNG	0.0000000000000000020
SD Edge Canny GIF	0.061
SD Edge Canny JPEG	0.011
SD Edge Canny TIFF	0.00018
Mean x SD Edge Perimeter GIF	0.17
Mean x SD Edge Perimeter TIFF	0.0040
Mean x SD Edge Perimeter PNG	0.0000000000000038
Symmetry diagonal top left bottom right (\)	-0.20
Quads 1x1	0.0051
Quads 32x32	0.098
Quads 128x128	0.060
APB Horizontal Inner Outer	0.085
APB Vertical Inner Outer	-0.031
Homogeneity	0.62

Note. Due to the large differences in size, I offended the suggestions of the APA to report two decimal places but instead decided to report two valid places for each coefficient.

The correlations between the 24 predictors as well as the criterion visual complexity (VC) rating in the top line are visualized in Figure 49.

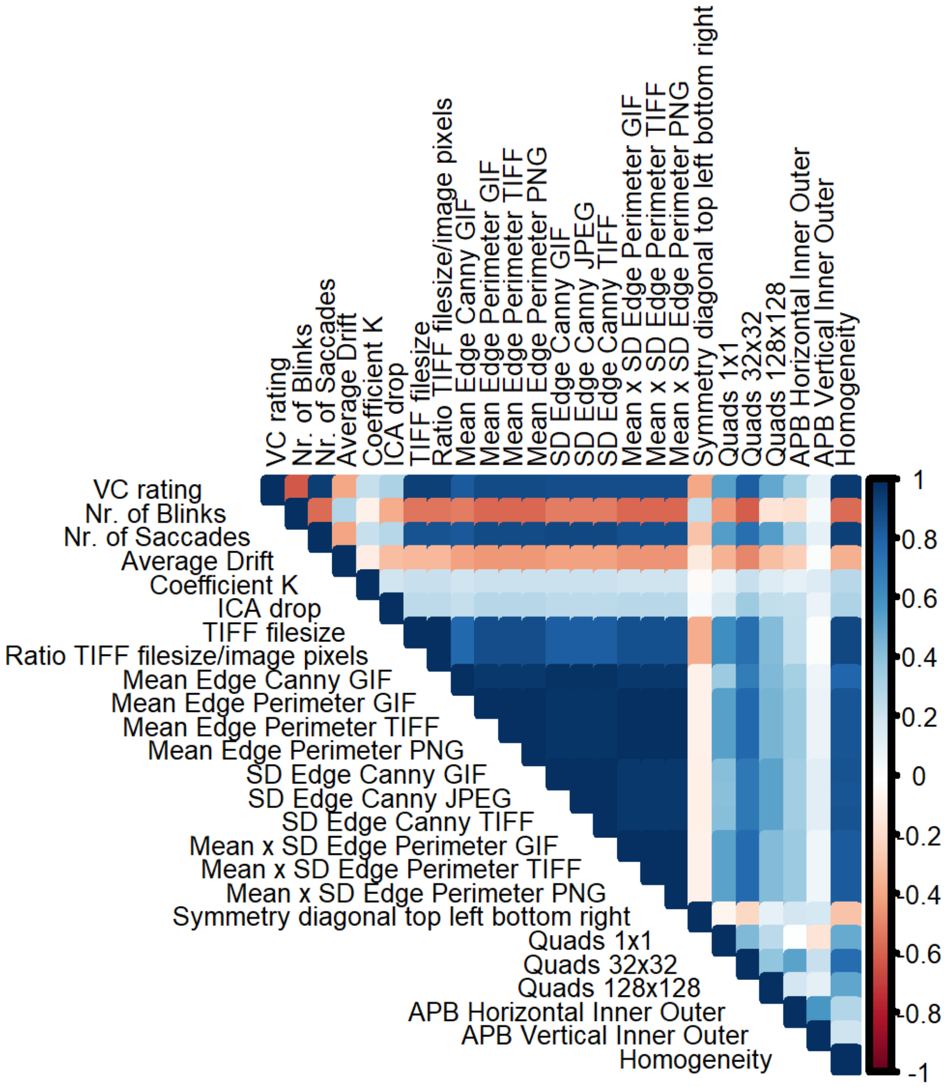


Figure 49. Correlations of selected predictors for mean complexity rating in study 3

The regression model fit on the training data gave an R^2 of .99 and a mean-squared error (*MSE*) of 0.062. The model was then used for the prediction of test data. A plot of the actual mean visual complexity ratings versus the according predictions from the regression model is depicted in Figure 50. The R^2 within the test data was .97 and the *MSE* 0.084.

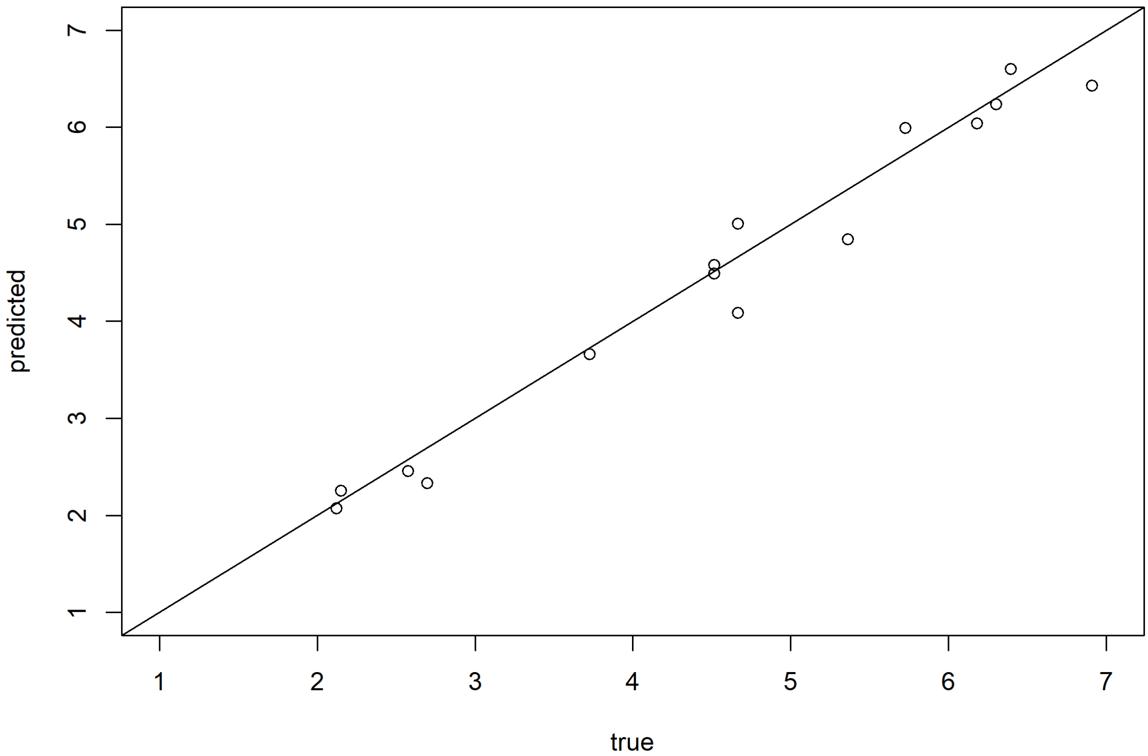


Figure 50. True versus predicted mean visual complexity ratings of the test data in study 3.

In a next step, prediction models for individual visual complexity ratings were investigated. The combination of prerequisites within the data could hardly be met with well established “standard” methods. The latter included ordinal scaling of the criterion visual complexity rating due to the seven-point Likert scale that was used, a clustered data structure both by subject and by stimulus image as well as a large number of potential predictors. Therefore, three different methodologies were evaluated within an explorative approach.

As a first step, the variables selected previously within the LASSO regression model of mean visual complexity ratings were used within a mixed-effects ordinal regression with a random intercept for subjects. It could be hypothesized that due to their impact on mean visual complexity ratings, they should also account for considerable variance within single ratings.

First of all, from the 24 predictors selected within the LASSO regression model, the variables Ratio TIFF filesize to image pixels, Mean Edge Perimeter TIFF, Mean Edge Perimeter PNG, SD Edge Canny JPEG, SD Edge Canny TIFF, Mean x SD Edge Perimeter TIFF, Mean x SD Edge Perimeter PNG were removed because of collinearity. The remaining 17 variables were than included within the ordinal mixed-effects regression model, for which again a set of training data consisting of 80% of the total observations were used. The model was implemented with the help of the function olmm from R-package *vcrpart* (Bürgin & Ritschard, 2017), with a random intercept for subjects. Within the training data, this achieved a correct classification rate of .47. Detailed information concerning the regression model can be found within appendix 9.16.

The regression model was then used for the prediction of single visual complexity ratings within the previously unknown test data. When the subject vector is taken into account, a correction classification rate of .45 could be achieved. If this is ignored and the population-averaged response probabilities are considered instead, the correct classification rate is .39. Confusion matrices for training and test data, the latter both with subject vector considered and ignored, are visualized within Figure 51.

Next to the correct classification rate, which indicates the percentage at which the predicted category is the same as the actual, additional evaluation metrics that also take into account the distance between categories may be of particular interest given the ordinal scaling of the visual complexity rating. In this regard, Gaudette and Japkowicz (2009) suggested to use the mean squared error (MSE) or the mean adjusted error (MAE), even though these are actually intended for continuous data. Baccianella, Esuli, and Sebastiani (2009) however showed that both these measures may perform poorly with imbalanced categories and suggest to use an adapted, macroaveraged version of the MAE, which is calculated for each category and then averaged, to that each category is given equal weight. All evaluation measures for both training and test data are reported in Table 8.

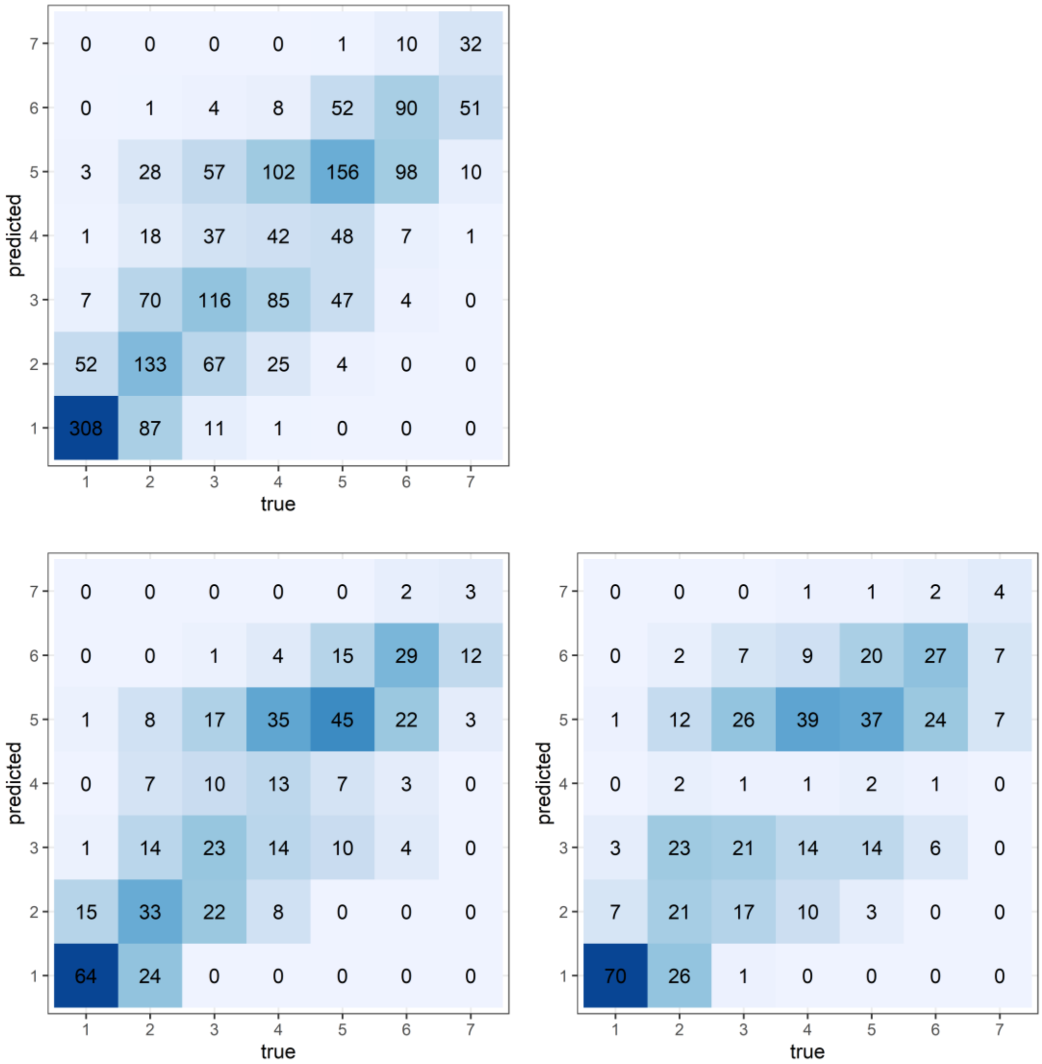


Figure 51. Confusion matrices of ordinal mixed regression for single visual complexity ratings within training data (top) and test data with subject vector considered (bottom left) and ignored (bottom right) in study 3

Within the next step, a random forest approach was used for the classification of single visual complexity ratings. Random forests (Breiman, 2001) are a popular tree-based technique, which is based on fitting large collections of de-correlated regression or classification trees to bootstrap-sampled versions of data and subsequently averaging the results (Hastie, Tibshirani, & Friedman, 2017). At each step, a random selection of predictor variables is considered. Classical random forests however do not offer the possibility to take into account a clustered data structure or ordinal scaling of data.

Random forests including all available predictor variables were modelled with the help of the R-package *ranger* (Wright & Ziegler, 2017). In order to improve prediction performance of the final random forest model, the parameters *number of variables randomly sampled as candidates at each split* (mtry), *minimum node size* and *sample size* were tuned within a grid search. The number of trees was set to 500. This procedure provided the lowest out-of-bag error of .58 and, hence, a correct classification rate of .42 for an mtry of 14, a minimum node size of 4 and a sample size of .50. The variable importance values for the top 30 variables in the final model are visualized within Figure 52. Variable importance describes the accumulated improvement of the split-criterion for the split variables (Hastie et al., 2017).

The final random forest allowed for a correct classification rate of .38 within the test data. All evaluation measures for both training and test data are reported in Table 8. Confusion matrices are visualized in Figure 53.

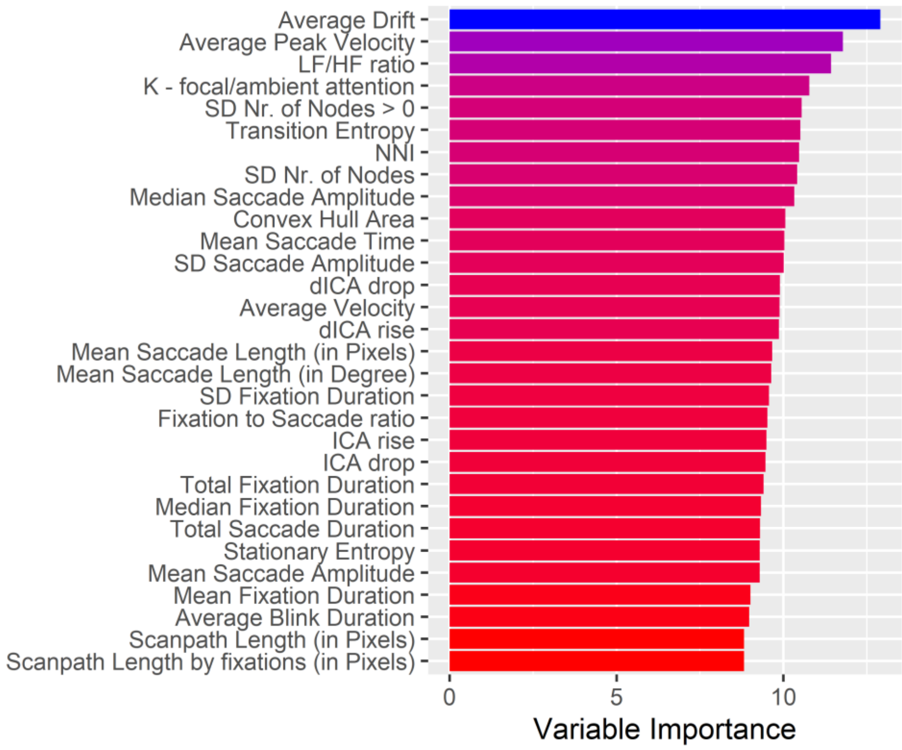


Figure 52. Variable importance values in the final random forest model for the prediction of single visual complexity ratings in study 3

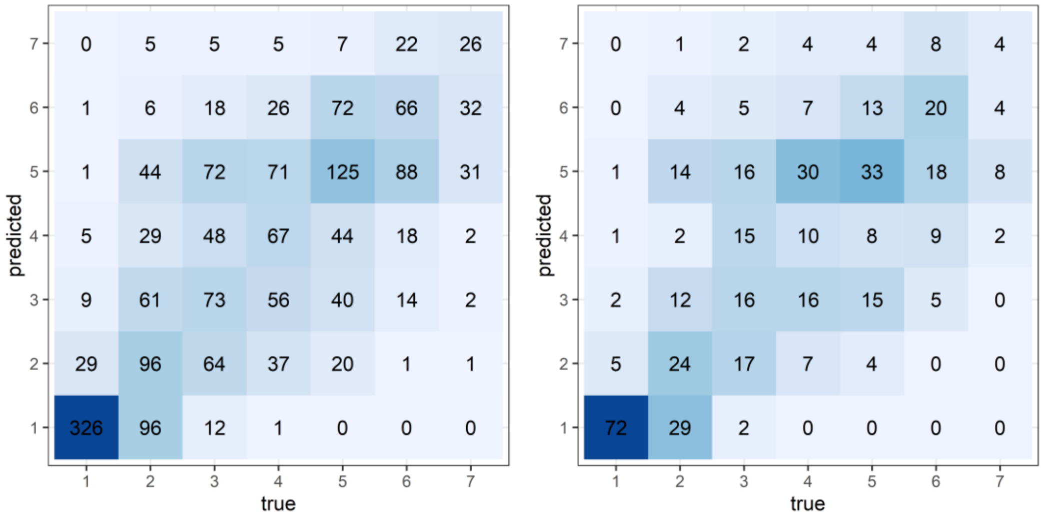


Figure 53. Confusion matrices of random forest for single visual complexity ratings within training (left) and test data (right) in study 3

Finally, an approach of variable selection by means of the L1-penalized Lasso estimation for generalized linear mixed models was used as implemented within the R-package *glmmLasso* (Groll, 2018; Groll & Tutz, 2014). While providing the functionality of Lasso variable selection, this approach can also take into account the clustered data structure with random effects for subjects and stimuli and offers cumulative link models for ordinal responses, which are suitable for the seven-point Likert scale ratings of visual complexity.

After determining the optimal penalty parameter λ based on the Bayesian information criterion (BIC) (Schwarz, 1978), the final model was built based on the optimal parameters. Thereby, three different approaches were taken. The first model encompassed two random effects for both subjects and stimuli and achieved a correct classification rate of .47. Within this, the variable Homogeneity was selected. Details regarding the regression model are reported in appendix 9.17. Within the test data, this model showed a correct classification rate of .46. A detailed overview of all evaluation measures for both training and test data is reported in Table 8. Confusion matrices are visualized in Figure 54.

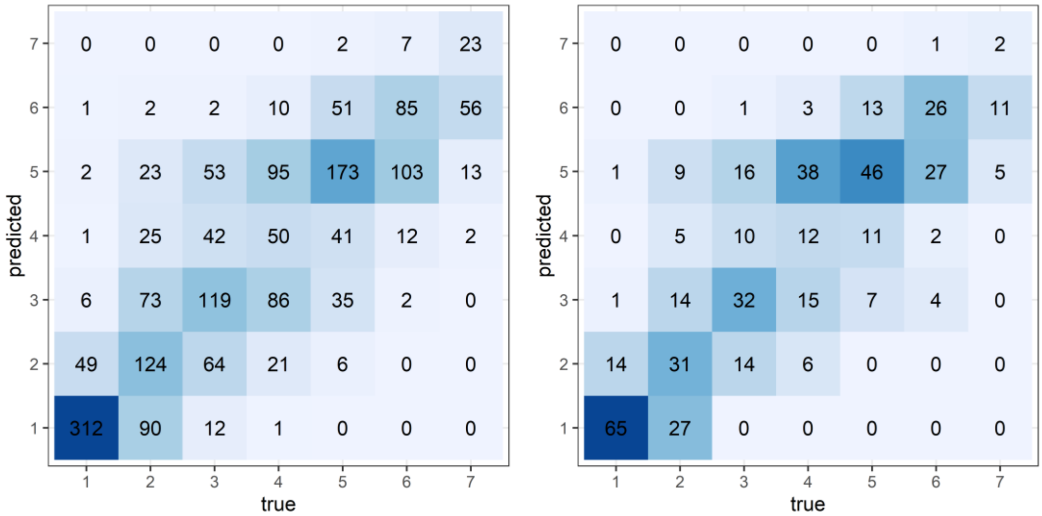


Figure 54. Confusion matrices of glmmLasso with random effects for subjects and stimuli for single visual complexity ratings within training (left) and test data (right) in study 3

The second glmmLasso model encompassed one random effect for subjects, in order to investigate the prediction performance when stimulus information are not considered or available. This achieved a correct classification rate of .44. Within this, the variables TIFF and Homogeneity were selected. Detailed information regarding the model and coefficients are reported within appendix 9.18. Within the test data, this model gave a correct classification rate of .42. A detailed overview of all evaluation measures for both training and test data is reported in Table 8. Confusion matrices are visualized in Figure 55.

Finally, the third glmmLasso model encompassed no random effects in order to investigate the prediction performance without considering any additional information apart from the automated and ocular measures. This achieved a correct classification rate of .42 with variable Homogeneity selected. Detailed information concerning the final model can be found in appendix 9.19. Within the test data, this model gave a correct classification rate of .37. A detailed overview of all evaluation measures for both training and test data is reported in Table 8. Confusion matrices are visualized in Figure 56.

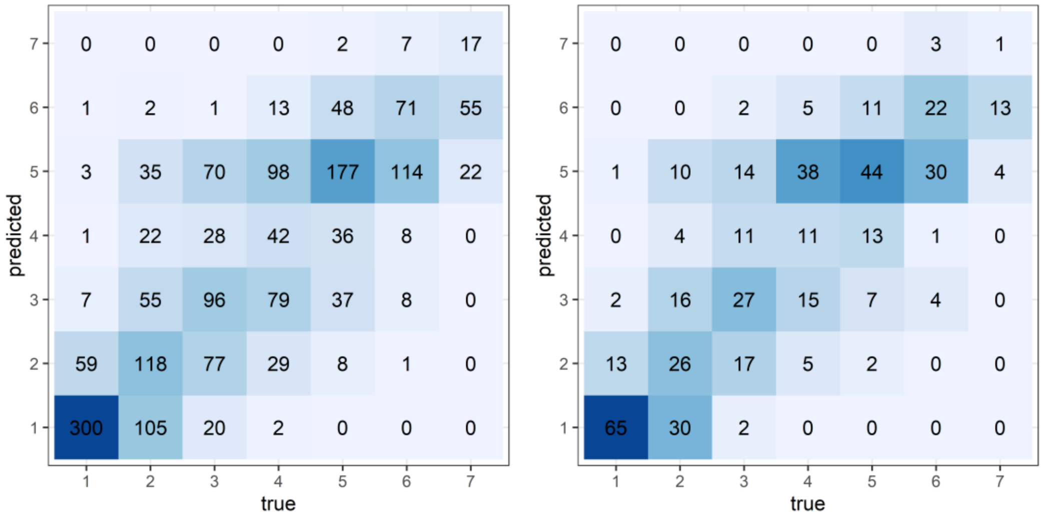


Figure 55. Confusion matrices of glmmLasso with a random effect for subjects for single visual complexity ratings within training (left) and test data (right) in study 3

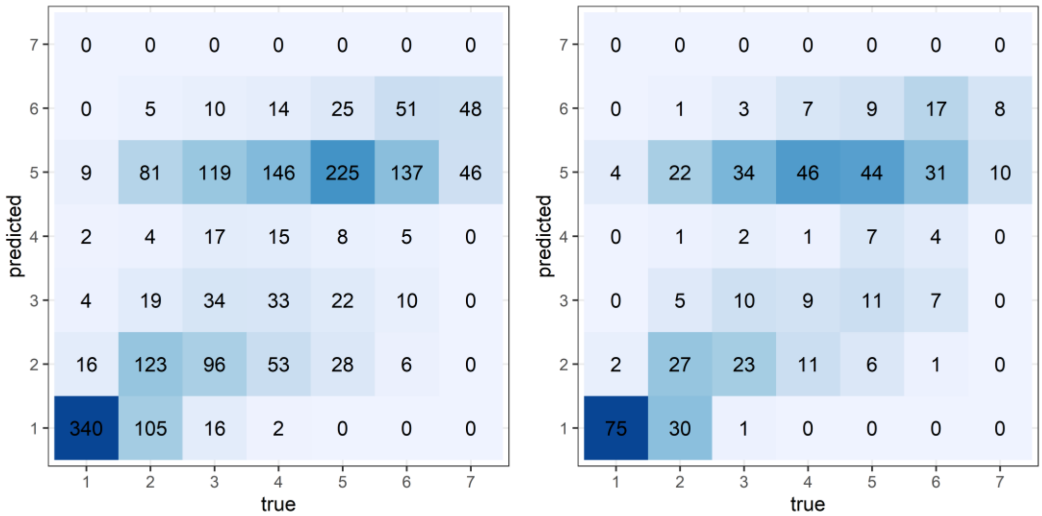


Figure 56. Confusion matrices of glmmLasso without random effects for single visual complexity ratings within training (left) and test data (right) in study 3

Table 8.

Evaluation measures of different models for the prediction of single visual complexity ratings for both training and test data within study 3

Method	Corr. Class	*MAE*	*MSE*	*MAE* M	*MAE* M categorywise						
					Category						
					1	2	3	4	5	6	7
Training data											
olmm train	**.47**	**0.68**	1.05	**0.71**	0.22	0.83	0.86	0.97	0.68	0.64	0.79
RF train	.42	0.89	1.71	0.94	0.19	1.17	1.21	1.03	0.88	0.92	1.20
glmmLasso train RE Sub Stim	**.47**	**0.68**	**1.03**	0.72	0.21	0.86	0.83	0.94	0.60	0.67	0.94
glmmLasso train RE Sub	.44	0.76	1.23	0.81	0.25	0.94	0.99	1.01	0.60	0.79	1.05
glmmLasso train no RE	.42	0.90	1.76	0.99	0.18	1.17	1.41	1.21	0.52	0.96	1.49
Test data											
olmm test RE Sub	.45	0.73	1.14	**0.76**	0.26	0.88	0.95	0.99	0.55	0.70	1.00
olmm test averaged	.39	0.91	1.68	0.94	0.21	1.13	1.27	1.27	0.79	0.77	1.17
RF test	.38	0.97	1.95	1.03	0.20	1.26	1.25	1.16	0.92	0.98	1.44
glmmLasso test RE Sub Stim	**.46**	**0.71**	**1.10**	**0.76**	0.25	0.91	0.81	0.96	0.49	0.73	1.17
glmmLasso test RE Sub	.42	0.76	1.20	0.81	0.26	0.98	0.90	0.99	0.57	0.78	1.17
glmmLasso test no RE	.37	1.00	1.97	1.07	0.22	1.24	1.42	1.23	0.73	1.07	1.56

Note. Corr. Class: Correct classification rate, *MAE*: mean absolute error, *MSE*: mean standard error, *MAE* M: *MAE* macroaveraged according to Baccianella et al. (2009), *MAE* M categorywise: MAE for the response levels of the seven-point Likert scale separately, RE: random effect, Sub: Subject, Stim: Stimulus, the methods and overall measures with the best performance in both training and test data are highlighted in bold

5.5 Discussion

Within this study, the effects of two fundamental factors of visual complexity, quantity and structure, were investigated within a laboratory eye tracking experiment. Thereby, the correlational findings from study 2 regarding ratings of visual complexity should firstly be experimentally substantiated if possible. Moreover, findings from a selection of ocular parameters can reflect attentional aspects and might thus allow for insights regarding the cognitive processing of visual complexity. Finally, ocular and computational measures are used to predict both mean and single visual complexity ratings.

Conclusion

The results from this study showed that both the quantitative aspect number of elements as well as the structural variable symmetry significantly affected the visual complexity ratings. While stimuli were rated as more complex when they contained more elements, a higher level of symmetry went along with lower ratings of visual complexity. This is in line with both correlational findings from study 2 and previous research literature, where both a quantitative as well as a qualitative dimension were seen as crucial for the perception of visual complexity (Chipman, 1977; Gartus & Leder, 2017; Ichikawa, 1985). Next to the two main effects, an ordinal interaction effect of both factors was found. This might be related to a different impact of symmetry on complexity ratings for larger numbers of elements. Pictures with only one element served as a control condition and were excluded from the statistical analyses. Within these, the central image axes used for the definition of symmetry categories may have been of less relevance for the subjects' perception. Instead, they might have largely used the centre of the single object as a reference point for symmetry, resulting in a rather high level of 'perceived' symmetry in contrast to the computational one. Therefore, the perceived symmetry level may have differed from the computed one.

Ocular parameters

Next to subjective ratings, it could be shown that the number of elements as well as symmetry also affected the selected ocular parameters. With regard to the number of fixations, main effects for both factors were revealed, where a larger number of elements and a lower degree of symmetry led to a higher number of fixations. This corresponds to findings by Bradley et al. (2011) and Madan et al. (2017), who similarly identified a positive relation between visual complexity and the number of fixations.

For scanpath length, an interaction effect of both experimental factors was identified. On the one hand, scanpath length was larger for pictures with medium symmetry compared to symmetrical pictures and increased with a growing number of elements for these two symmetry levels. On the other hand, scanpath length for asymmetrical pictures was largest for pictures with five elements and then decreased with a growing number of elements. This may point towards a switch of gaze behaviour between very simple, asymmetrical and thus unstructured and more complex images. Participants may have used a rather sequential gaze behaviour for most pictures, where they successively gazed at various objects within the picture. This is also described within Treisman and Gelade's (1980) feature integration theory, which supposes that focal attention, becoming manifest in the serial scanning of successive locations, is necessary for the integration of features for perception. The larger number of elements and decrease in symmetry would then lead to an increased scanpath length, which could also be observed within the data. For asymmetrical pictures with many elements however, this gaze behaviour might have been associated with too much effort, also with regard to the limited presentation duration of 6000ms. Instead, they may have applied a holistic gaze behaviour for these rather complex pictures, which may have resulted in decreasing scanpath lengths.

This may be related to what Oliva (2005) describes as the “gist of a scene”, meaning that observers can get a grasp of a scene including information about basic features within very short time. This is underlined by Potter (1976), who found that an average scene can be understood within as little as 100ms. This “gist” refers not only to low-level features of an image, but includes both perceptual and conceptual levels according to her. This aspect is also stressed within Torralba, Oliva, Castelhano, and Henderson's (2006) contextual guidance model, which suggests that a holistic representation of a scene can be built quickly enough to guide the subsequent deployment of attention as well as the eye movements. Another possible explanation for this effect is related to the concept of perceptual grouping (Wagemans, Elder et al., 2012) and the role of symmetry within it. Perceptual grouping describes the fact that “observers perceive some elements of the visual field as ‘going together’ more strongly than others”

(Wagemans, Elder et al., 2012, p. 1178), with symmetry as an important factor that facilitates the grouping of elements. In this regard, results from a strongly controlled experiment by Machilsen, Pauwels, and Wagemans (2009) empirically underline the relevance of mirror symmetry for perceptual grouping. A last possible aspect that might have contributed to the emergence of this rather unexpected result pattern is related to aspects of short-term memory, namely Miller's (1956) magical number seven plus or minus two. This is the number of information chunks that according to him can be held within the short-term memory and thus characterises the memory span within a fixed-capacity model. Although more recent research suggests a lower number of about four elements or chunks (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997), the visual information load of objects was also shown to affect capacity beyond to the sole number of elements (Alvarez & Cavanagh, 2004). Even though the experimental task did not explicitly require subjects to memorize items, memory capacity might still have influenced gaze behaviour of participants. In this context, memory capacities for limited number of objects, taken together with the impeded perceptual grouping for asymmetrical stimuli or a rather holistic processing of the scene, may explain the decreasing scanpath length for asymmetrical pictures with more than five objects.

Finally, for the measure spatial density, a similar number of elements by symmetry interaction effect was found. As for scanpath length, spatial density was larger for pictures with medium symmetry compared to perfectly symmetrical pictures and increased with a growing number of elements. For asymmetrical pictures however, spatial density increased with growing number of elements until a maximum for nine elements, but decreased for 13 elements. Similar possible explanations for this result pattern can be proposed as for the scanpath length. It remains however unclear, why the peak for spatial density was at asymmetrical pictures with nine compared to five elements for scanpath length.

Prediction

With the help of the Lasso regression model, very accurate predictions of mean visual complexity ratings for the stimulus images could be achieved with an R^2 of close to 1 and very small MSE values. Within the Lasso feature selection, both automated as well as ocular parameters were selected, which might suggest that both relate to the perception and rating of visual complexity and may thus be helpful for the prediction of visual complexity ratings. A large coefficient was particularly found for Homogeneity, a measure of visual balance established by Hübner and Fillinger (2016) in order to assess how scattered the elements within a picture are. Further large coefficients included compression measures such as TIFF filesize, combinations of edge detection and compression measures such as Mean x SD Edge Perimeter GIF as well as structural measures such as diagonal symmetry. In comparison to previous research works such as Gartus and Leder's (2017) studies, in which a maximum of explained variance of .82 within a linear regression model and .89 within a random forest could be achieved, the excellent prediction accuracy within this study may appear surprising, particularly given the smaller sample size and less stimuli compared to Gartus and Leder (2017). One reason for the better performance may consist in the stimulus material that was used. While the number of the number of elements, symmetry and type of elements were systematically varied according to the experimental design, the pictures of black and white shape patterns that were used might be described as generally less complex and diverse compared to the two stimulus sets by Gartus and Leder (2017). It might be hypothesized that larger levels of diversity within the stimuli may go along with a larger variety of influences on the complexity ratings, which cannot complete be accounted for with the help of computational measures. On the other hand, the implemented computational and ocular measures may of course also play an important role with regard to the accuracy of the model. While firstly, Hübner and Fillinger's (2016) Homogeneity, which was strongly related to the visual complexity ratings within this study, was not included by Gartus and Leder (2017), the integration of ocular parameters may also have contributed to the improved prediction performance.

With regard to the models for single visual complexity ratings, three rather explorative approaches were taken by means of different methods. This was due to the relatively complex requirements imposed by the clustered data structure, ordinal scaling of the criterion and the large number of potential predictors. These can hardly be perfectly met using well-established standard methods. First of all, the variables selected within the previous Lasso regression model for mean visual complexity ratings were included within an ordinal mixed model with a random intercept for subjects. Secondly, a random forest approach for classification was used. While this does not consider the ordinal structure and does not allow including random effects, random forests often offer good performance (e.g. Gartus & Leder, 2017; Hastie et al., 2017) and can also consider complex non-linear relationships. Finally, generalized linear mixed-models with Lasso feature selection (glmmLasso) were implemented. These meet the requirements of ordinal scaling of the criterion, clustered data and feature selection very well.

The results showed similar correct classification accuracies for ordinal mixed-models with a random intercept for subjects and glmmLasso with random effects for both subjects and stimuli of between 45 and 47 percent and an MAE_M of 0.71 and 0.76. Since the evaluation measures are similar between training and test data, overfitting seems to play only a minor role within these. Remarkable, glmmLasso yields good performance with only one (Homogeneity) or two variables (TIFF filesize and Homogeneity) selected, with 46% of correct classifications for test data and an MAE_M of 0.76 when both random effects for subjects and stimuli are included and 42% and an MAE_M of 0.80 when only subjects are considered, respectively. Interestingly, no ocular parameters were selected by the glmmLasso method within this study. This may appear surprising at first, since the computational measures TIFF filesize and Homogeneity do not vary between trials but only between pictures. It thus might have been expected that for the accurate prediction of single ratings, ocular parameters should have proven as essential. While ocular parameters were selected within the Lasso model of mean visual complexity ratings and are therefore considered within the ordinal mixed model and additionally showed large variable importances within the random forest, these seemed to play no important role within the glmmLasso regression. Two possible explanations for this could be seen. First of all, it might be argued that there were only minor differences in visual complexity ratings between subjects within this study, so that image-specific computational measures alone were already sufficient for the prediction of single visual complexity ratings. This is however not in line with the findings of decreasing prediction accuracies, when subject information is not considered within the models. These much rather suggest that particularly the included random effects for subjects can account for interindividual differences in the rating of visual complexity. In this context, the percentages of correct classification below 50% mean that the majority of ratings were not classified in the correct rating category. This may appear disappointing at first, suggesting that ratings cannot be adequately be predicted from the available data. Within ordinal data, the correct classification rate alone however does not show the whole picture. MAE evaluation measures below one for all methods except for the random forest and glmmLasso without any random effect indicate that rating predictions on average are not further than one category level apart from the actual rating, with much smaller errors particularly for category one, which encompassed a large proportion of all ratings. This is also illustrated within the confusion matrices. All in all, at least when subject information is taken into account, prediction models based on computational (and potentially also ocular) parameters can provide a relatively good approximation for individual visual complexity ratings.

Limitations

A first limitation of this study refers to the stimuli that were utilized. Very basic black and white shape patterns were used in order to avoid any possible confounding by other aspects. This however implies that a large number of other factors could not be addressed. Among these are for example aspects of the “colour”-dimension discovered in study 2 but also further aspects such as the density of elements or the variety of elements, belonging to the “quantitative” dimension. The integration of all possible factors would have generated a very complex experimental design. Thus, no conclusions about effects of these can be drawn from this study. Since the luminance of the stimuli was not controlled within this study, this may bias the use ocular measures such as the pupil size as indicators of visual complexity. Therefore, only pupil measures that control for lighting conditions such as the IPA were taken into account.

Moreover, the viewing durations within the experiment were controlled and set to 6000ms in order to avoid different viewing durations. These would certainly have appeared within a self-paced experiment. Eventually, complex images might have been observed longer than less complex images, as shown for example by Shigeto, Ishiguro, and Nittono (2011). This might have biased the visual complexity ratings. For the perception of visual complexity within naturalistic settings, the potential effects of different viewing duration on complexity perception might be an important issue, which yet could not be considered within the laboratory setting of this study.

Finally, the assessment of visual complexity ratings on a seven-point Likert scale may be a double-edged issue. One the one hand, the use of Likert scales is relatively common in research on visual complexity. Gartus and Leder (2018) for example used a five-point Likert scale to assess the impression of visual complexity while Nadal et al.

(2010) for example used a nine-point Likert scale. Additionally, the approach of using a seven-point Likert scale proved to be functional within study 2. On the other hand, the ordinal scaling of the rather small range of ratings complicates the analysis and particularly the prediction of single values. In contrast to mean ratings, methods for the prediction of ordinal data are necessary, which are not yet as widely established. Possibly, the use of a slider scale with a continuous range of response values for example would permit the use of classical linear mixed effects models. However, the use of slider scales showed to negatively affect the distribution of values and to increase response times, why it is not recommended (Funke, 2016).

Outlook

Within this experiment, the number of elements and symmetry, representing the two basic visual complexity factors quantity and structure, were experimentally investigated. Within future research, a systematic investigation of other factors as well as potential covariates might help to further clarify the construct and examine relations also within different domains. Moreover, a larger variation within the stimuli's complexity range might help to precisely address the sensitivity of scanpath patterns and spatial density for different complexity levels and particularly the effects of single influencing variables such as symmetry. Concerning the investigation of prediction models for single visual complexity ratings, a larger variation within stimuli as well as a larger sample size may prove to be advantageous in future studies. Since ratings within this study were largely clustered in the lower range of the 7-point Likert scale with most ratings in the lowest category one and almost none within category seven, the latter could be hardly predicted with appropriate accuracy. Additional stimuli within the higher complexity range and a larger sample size might thus help to establish more accurate models accounting for the whole range of the scale. Next to the implemented methods, further approaches are conceivable. With regard to feature selection or dimension reduction for example, results from methods such as recursive feature selection, principle component analysis or the Boruta algorithm (Kursa, Jankowski, & Rudnicki, 2010) may be compared to the reported ones while with regard to the actual prediction models, methods such as mixed-effects random forests (Hajjem, Bellavance, & Larocque, 2014) might be examined in order to take into account both clustered data structure as well as non-linear relationships between predictors and visual complexity ratings. On the other hand, non-linear transformations could also be applied to predictor variables as for example done by Gartus and Leder (2017), who could improve prediction performance by applying a power transformation to computed mirror symmetry values.

Additionally, varying picture presentation durations within experimental studies might facilitate further inferences about underlying cognitive processes and dimensions (Ichikawa, 1985).

While the use of eye tracking methodology appeared promising for the closer investigation of the construct visual complexity, it may also help to gain further insights regarding the perception of visual complexity for more naturalistic stimuli. This may carry special relevance within the field of human computer interaction, for example regarding the design of user interfaces, where visual complexity may also be related to mental workload (Harper et al., 2009). For the purpose of optimizing the design of user interface with regard to visual complexity, particularly the prediction of visual complexity ratings from computational and ocular parameters might be of great use.

6. Study 4: Application: Visual Complexity in user interfaces

Within the previous study, strongly controlled black and white shape patterns were used for the investigation of visual complexity. This study aims to examine if the previous findings from study 3 can be transferred to an applied setting. In order for that, the identified relations between influencing variables and ratings of visual complexity as well as ocular and computational measures will be examined. In particular, screenshots of websites are used as stimuli. Within these, further influencing variables are considered which may contribute to the perception of visual complexity within this context. The findings of this study can help to underline the impact of influencing variables within the applied context of user interfaces and be helpful with regard to their design. Furthermore, the approach of this study can find practical application for example in the prediction of complexity ratings from both computational measures as well as ocular parameters, which could also complement classical usability testing. Finally, the investigation of relations between visual complexity and mental workload is of particular interest for the assessment of the role of the former construct within human-machine interaction.

6.1 Background

The previous study supported the significance of both a quantitative as well as a structural factor for the perception of visual complexity. As previously discussed, this is in line with findings from basic research. With regard to user interfaces however, findings on influencing variables are rather contradictory. While Miniukovich and Angeli (2014) for example suggested a three-dimensional structure of the construct with the factors information organization, information amount and information discriminability, Miniukovich et al. (2018) proposed a different structure with four dimensions (quantity of information, variety of visual form, spatial organization, perceivability of detail). For a more detailed discussion of influencing variables of the visual complexity of user interfaces as well as their dimensionality, see paragraph 2.3.2. In order to add clarity and compare the dimensional structure of visual complexity within different domains, study 2 was previously conducted. This revealed that the relation between global visual complexity ratings and its influencing variables as well as their factorial structure were largely comparable between photographs and website screenshots with the three factors quantity, structure and colour being related to global visual complexity ratings. While study 3 then experimentally investigated the influence of a quantitative and a structural dimension using black and white shape patterns, experimental investigations for user interfaces such as websites are still missing to the best of my knowledge. Within this study, it is examined if insights from basic research can be successfully transferred to an applied context.

Of the three complexity factors identified in study 2, the quantitative and the structural ones were experimentally investigated in study 3. The third, colour-related factor has however not been addressed yet. Factor scores of this dimension were shown to contribute to the global visual complexity ratings while regressions with single variables revealed an influence of the variety of colours and colour contrast particularly for photographs. Within website screenshots, these single variables however appeared less relevant. As also discussed in paragraph 2.2.3, findings from previous literature regarding the effect of colour are similarly inconstant. While results by Reinecke et al. (2013) and Nadal et al. (2010) point towards an influence of colourfulness on visual complexity, other authors such as Ciocca et al. (2015) suggest that colour plays no important role with regard to complexity. All in all, this leaves the relationship between visual complexity and colour still rather unclear. In order to shed light on this issue, this study will address the effects of colourfulness within an experimental approach. Thereby, it will be clarified if colourfulness can affect the perception of visual complexity particularly within website screenshots or if it has only little or no impact.

In addition, it may be hypothesized that further factors are of relevance in particular with regard to the perception of user interfaces. This may relate particularly to the fact that unlike for basic shape patterns, subjects have certain pre-experience and expectations about the design of user interfaces such as websites, which may also affect the perception of visual complexity. In this regard, Roth et al. (2010) for example showed that users' mental models of web pages determined where certain types of objects such as the logo or name, the navigation area or the search field were expected on a web page (see Figure 57).

Figure 57. Placement of types of objects for an online shop web page, taken from Roth et al. (2010)

Accordingly, if objects are placed according to the mental model of the user, it can be assumed that these may be perceived as prototypical while if they are placed at unexpected locations, the web page should be perceived as non-prototypical. The prototypicality of the web pages can then be hypothesized to affect the perception of visual complexity as also reasoned in paragraph 2.2.5.3. Due to the familiar structure of prototypical stimuli, the perceptual grouping may be facilitated compared to non-prototypical ones. In this regard, Kimchi and Hadad (2002) for example found within an experiment by using either upright or inverted letters as familiar or unfamiliar visual configurations that past experience contributes to the early perceptual grouping of elements into configurations. Effects of prototypicality could moreover be shown by Tuch et al. (2012) with regard to aesthetic judgements, however to the best of my knowledge no systematic investigations have yet directly addressed the role of prototypicality for the perception of visual complexity.

Furthermore, the previous study identified effects of visual complexity factors on multiple gaze parameters such as the number of fixations, scanpath length and spatial density. While the former measure showed linear relations with the number of elements and symmetry as expected, the latter two revealed interaction effects, which may point towards a different information processing style for more complex stimuli. In this context, it is of particular interest to use rather complex naturalistic stimuli such as websites for the investigation of effects on ocular parameters in order to take into account a different complexity range of stimuli as compared to study 3. Results can then contribute to a better understanding of visual attention in human-computer interaction.

Regarding the relevance of visual complexity within human-machine interaction, study 1 revealed effects of video complexity on control room operators' mental workload within a CCTV surveillance task. This was evident not only in subjective ratings, but also showed influence on performance measures and physiological indicators of mental workload. Within human-computer interaction, visual complexity effects on mental workload have however hardly been investigated. There are yet good reasons to assume an impact, for example the limited capacities for information processing, as also discussed in paragraph 2.3.1. Therefore, the relation between visual complexity and mental workload is addressed within this study.

Finally, the previous study revealed very promising results regarding the prediction of visual complexity ratings. Thus, mean ratings of visual complexity could be predicted with a very high accuracy using a combination of both computational and ocular parameters. Moreover, complexity ratings for single trials were predicted initially with satisfactory accuracy. Within this study, it will be investigated whether the prediction approach can be extended to screenshots of websites. Within these, an accurate prediction of visual complexity can offer several benefits. First of all, due to its relation with aesthetical judgements (see paragraph 2.2.5.2) and its presumed effect on mental workload, it appears relevant particularly within the interaction with technology. Complementing usability testing for example, it can reveal additional insights regarding the design of user interfaces, particularly when focussing on the first impression of the interface. The prediction of visual complexity ratings can moreover be used for a quick screening of user interface designs instead of rather time-consuming behavioural assessments. Combined with online webcam-based eye tracking (e.g. Semmelmann & Weigelt, 2018), which gains increasing popularity, this approach could be used for remote online studies or even live feedback for websites, which could be conducted with less effort than typical laboratory studies or testing. In this regard, the present study may depict a valuable first contribution.

6.2 Research questions

1. Do number of elements, symmetry, colourfulness and prototypicality affect visual complexity ratings of website screenshots as well as the ocular parameters number of fixations, scanpath length and spatial density?
2. Is the visual complexity of website screenshots related to the perception of mental workload?
3. How well can mean and single visual complexity ratings for website screenshots be predicted from computational and ocular parameters?

6.3 Method

For the investigation of the research questions, a laboratory eye tracking experiment with realistic website screenshots was conducted. This is described in detail within the following paragraphs.

6.3.1 Participants

41 subjects participated within the experiment, of which one had to be excluded because the experiment could not be finished. Of the remaining 40 subjects, 19 were females (47.5 %) and 21 males (52.5%) with a mean age of 26.0 years (*SD* = 7.6). 36 (90.0%) of the subjects were students, while four (4.0 %) were employed. The study lasted approximately 45 minutes and participation was refunded with 10€. All participants had normal or corrected-to-normal vision.

6.3.2 Experimental design and procedure

For the investigation of the previously stated research questions, a repeated-measures 3 (number of elements) x 2 (symmetry) x 2 (colourfulness) x 2 (prototypicality) experimental design was used. Moreover, an additional factor website type with three categories company, news and shopping was included in order to represent the variety of the most popular websites. This was however not of theoretical interest and thus not further considered within the analyses. However, the three selected websites were modified according to the previous experimental design, which is also visualized in Figure 58.

Independent variables:			
Number of elements	Symmetry	Prototypicality	Colourfulness
• 3 elements • 6 elements • 9 elements	• symmetrical • asymmetrical	• prototypical • not prototypical	• high colourfulness • low colourfulness

Dependent variables:
• Subjective ratings of visual complexity • Ocular parameters: Number of fixations, scanpath length, spatial density

Figure 58. Experimental design for study 4

An independent manipulation of all factors was a central aspect for the appropriate creation of the according stimuli in order to ensure the interpretability of potential effects. First of all, the number of elements had three levels with either three, six or nine elements within a website. One element was defined as a cluster of information content such as a news item consisting of an image with related text. Since elements within company, news and shopping websites strongly differ, examples for elements from all three website types are given in Figure 59, Figure 60 and Figure 61.

Figure 59. Example for one element in a website of type company

Figure 60. Example for one element in a website of type news

Figure 61. Example for one element in a website of type shopping

The second factor symmetry had two levels with either high or low symmetry. It was manipulated with respect to the vertical symmetry axis, which in contrast to the horizontal axis is of greater relevance within websites (Seckler, Opwis, & Tuch, 2015; Tuch et al., 2010). The manipulation of symmetry did only affect the element clusters within the website but not its headline in order to maintain a realistic appearance of the site. It was then controlled by means of Elawady et al.'s (2017) method for the quantification of symmetry within naturalistic images. This gives a value between zero for minimum and 100 for the maximum symmetry. The mean symmetry for the symmetrical screenshots was 0.43 (*SD* = 0.10), while the mean symmetry for asymmetrical stimuli was 0.18 (*SD* = 0.10). Two examples are visualized in Figure 62.

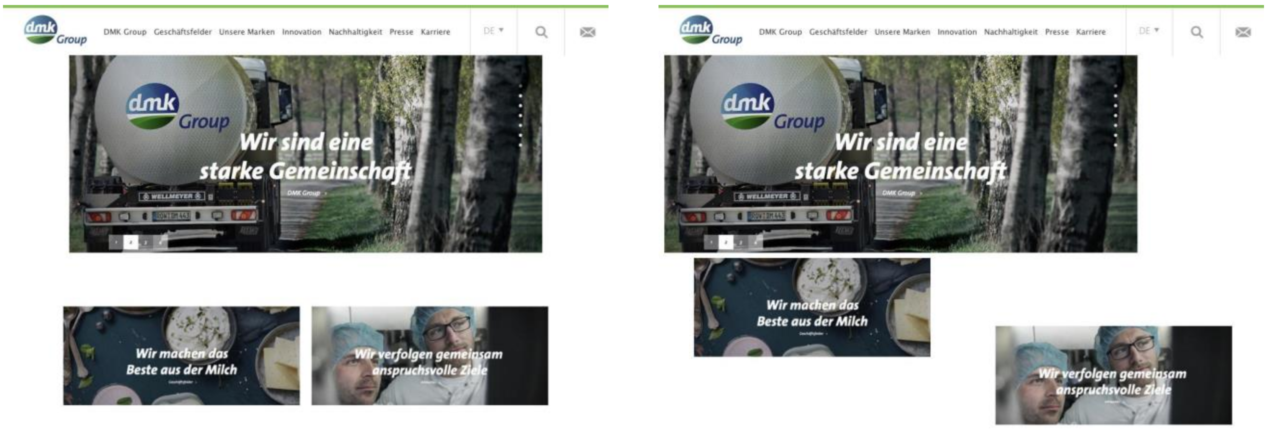


Figure 62. Symmetrical (left) and asymmetrical example (right) of a company website

Similar to symmetry, the next factor colourfulness also had two levels, high and low colourfulness. It was manipulated by increasing or decreasing the hue of the image and controlling the calculated level of colourfulness according to Hasler and Suesstrunk's (2003) method. This calculates the colourfulness of an image within the RGB colour space. According to their classification, an image with a value of zero is not colourful, while a value of 59 is quite colourful and 109 is extremely colourful. As for the manipulation of symmetry, differences between the two levels of colourfulness should be recognizable while both levels should still appear realistic and neither too colourful nor too colourless. Within the stimuli used, the colourful stimuli had an average colourfulness of 46.83 (*SD* = 8.50), while the less colourful images had an average colourfulness of 26.04 (*SD* = 6.67). Examples for screenshots with low and high colourfulness are shown in Figure 63.

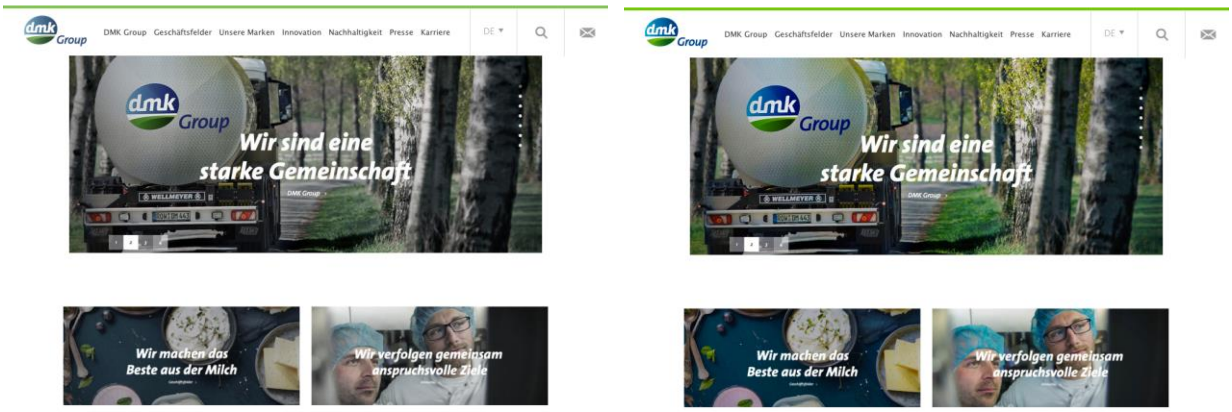


Figure 63. Low colourfulness (left) and high colourfulness example (right) of a company website

Finally, the fourth factor prototypicality again had two levels: prototypical and not prototypical. The manipulation of this factor was based on Roth et al.'s (2010) findings (see also Figure 57). Within their study, the authors identified locations where subjects expected certain items, such as logo, search field and newsletter link. Accordingly, elements of the menu bar such as the logo, navigation area, search function and contact as well as the newsletter registration were placed either at the expected location or on the opposite side of the image (see for example Figure 64). Since unlike symmetry or colourfulness, prototypicality can hardly be quantified by means of computational methods, a manipulation check was integrated within the study in order to ensure the validity of the manipulation.

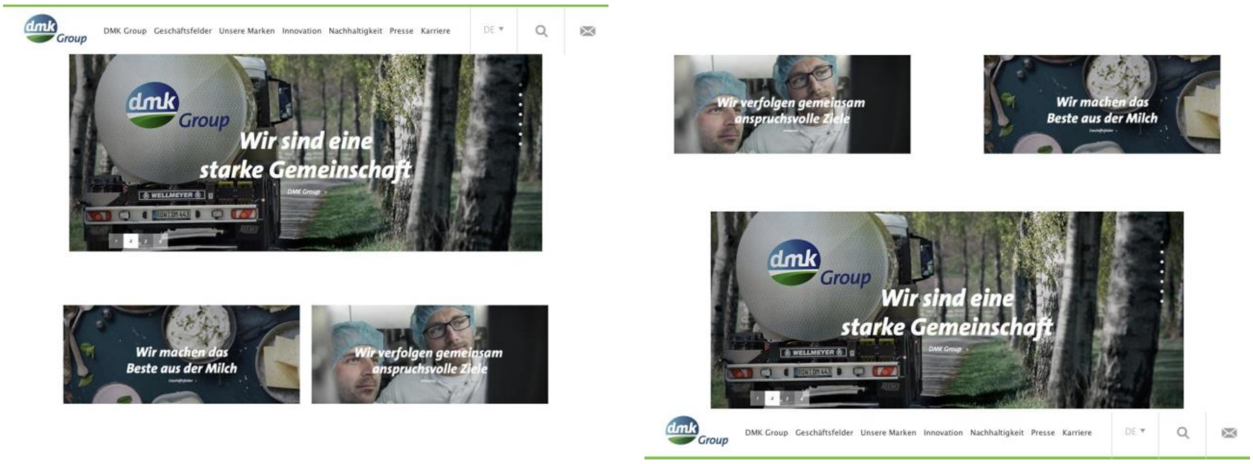


Figure 64. High (left) and low prototypicality example (right) of a company website

As an additional factor and in order to create a larger variation within stimuli, different types of websites were considered. According to Roth et al. (2010), the majority of the 100 most visited websites of Germany, Austria, Switzerland and the USA belonged to the categories company pages, social network sites, news portals, online shops and search engines. Social network sites usually show only very limited content without login, while the design of search engine site is usually also very rudimentary. Therefore, similarly to Roth et al. (2010), the three categories online shops, news pages and company pages were selected while for each type, a rather unknown German website was chosen as a model for the realistic creation of stimuli. As a company site, the website www.dmk.de was selected, while www.newlook.com/de was chosen as a shopping site. Finally, www.kurier.at served as a basis for the creation of news page screenshots. Since there were no specific hypotheses regarding the type of website, effects for these were not analysed.

The procedure of the experiment began with the instruction of participants that they could abort the experiment at any time without consequences. Moreover, they were explained that data was stored anonymously and used for research purposes, to which they gave written informed consent.

Subsequently, participants were explained within the experimental instructions that their task would be to carefully watch a number of website screenshots before giving ratings for each of these directly after presentation. After the following calibration of the eye tracker, the experimental procedure was started in *OpenSesame 3.2.6* (Mathôt et al., 2012). First of all, participants could get accustomed with the experimental procedure within three practice trials containing stimuli similar to the actual website screenshots. The following main part of the experiment comprised the presentation of 72 stimulus pictures, which are described in detail within paragraph 6.3.3. Each picture was presented for 6000ms, preceded by a fixation cross and followed by a questionnaire for the assessment of visual complexity, liking, prototypicality and mental workload ratings, which is also described in paragraph 6.3.3. All ratings were self-paced. The study was structured into three blocks consisting of 24 trials with screenshots for each website type. Both blocks and individual trials were presented in a randomized order. Between the three blocks, participants could take a short brake in order to move and relax. At the end of the experiment, the familiarity of subjects with the original websites prior to the experiment was inquired within a questionnaire. This showed that the website www.dmk.de was familiar to two subjects, while the website www.newlook.com/de was also familiar to two subjects. The news page www.kurier.at was familiar to four subjects. All eight experimental blocks with the corresponding trials of websites that were familiar to a certain subject were excluded from further statistical analyses in order to avoid possible confounding with familiarity.

6.3.3 Material

Website Screenshots

Within this study, website screenshots were used as stimuli. For the creation of these, existing but rather unknown websites were used as a basis in order to make the final stimuli appear as realistic as possible[4]. Still, these should not be known to the subjects to avoid effects of familiarity. Further selection criteria were German language and current design which admitted the manipulation of the factors.

[4] Stimulus creation was accomplished in close collaboration with master student Yi Ding

This resulted in the selection of the company website www.dmk.de, the shopping website www.firstlook.com/de and the news website www.kurier.at, which served as models for the further adaption according to the experimental design. This was accomplished with the help of the software Axure RP (Axure Software Solutions, 2018). Within this, the number of elements, prototypicality and symmetry were manipulated by in- or excluding and rearranging elements as described within paragraph 6.3.2. Subsequently, the colourfulness of the images was adjusted. Thus, each combination of factors levels is consequently represented by one screenshot of the company site, one screenshot of the news website and one screenshot of the shopping site.
The images were exported as JPG-files with a resolution of 1024 x 768 pixels. Examples of the three types of websites with three elements are depicted in Figure 65.

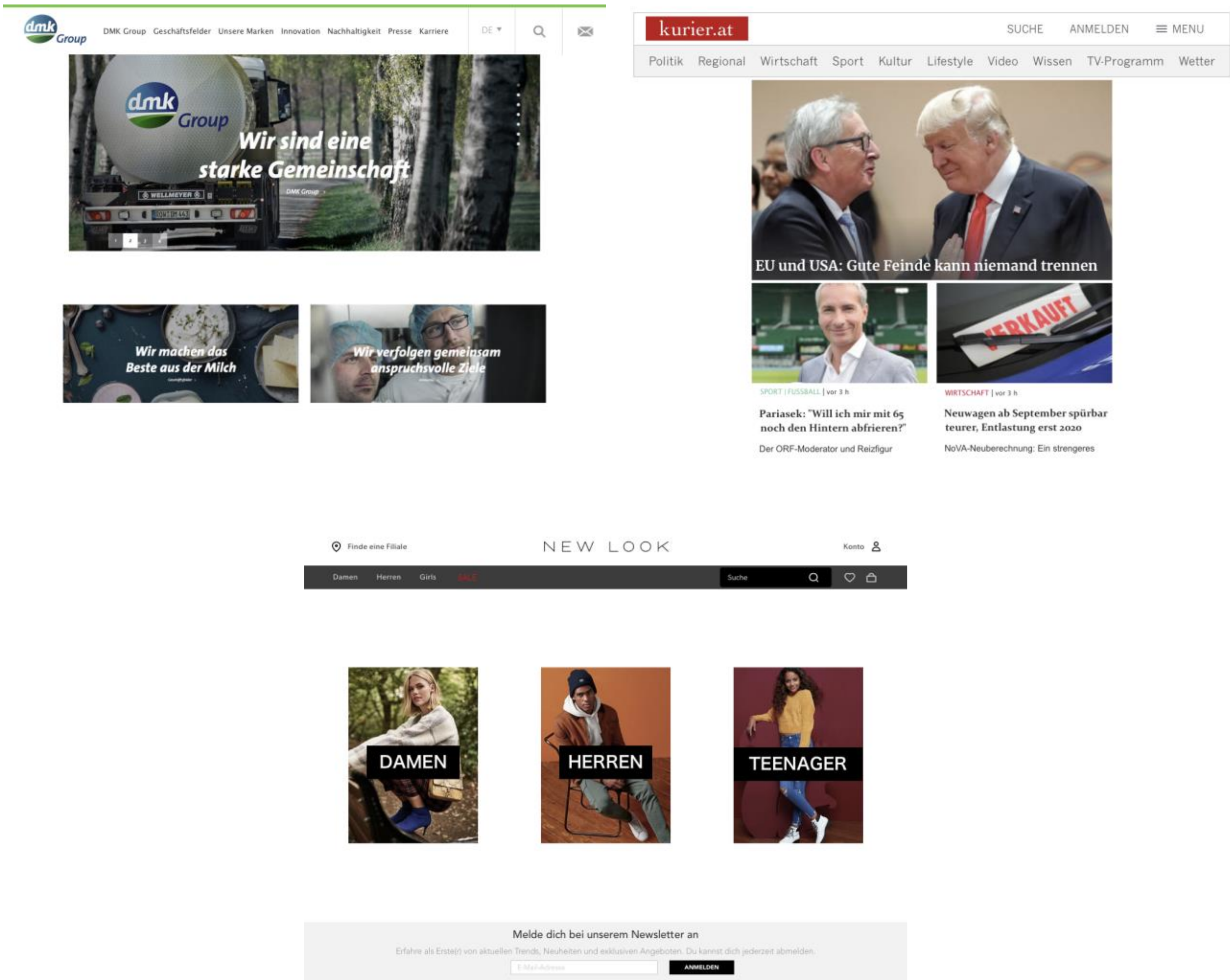


Figure 65. Examples of websites screenshots of different types of websites with three elements, symmetrical, prototypical, low colourfulness

As in study 3, the computational measures described in paragraph 2.4 were calculated for all stimuli. This resulted in a total of 183 measures and combinations of measures, including colour-related measures, which had not been considered within study 3.

Questionnaire

Within the questionnaire following each experimental trial, participants were asked to rate the visual complexity as well as the liking and prototypicality of the picture in 7-point scales ranging from 1 = “very low” to 7 = “very high”. The latter was verbalised as “This website looks like a typical website” (“Diese Webseite sieht wie eine typische Webseite aus”), according to the definition of prototypicality as “the amount to which an object is representative of a class of objects’’ (Leder et al., 2004, p. 496) or of how typical an example is considered for a category (Rosch, 1973).

In addition, the mental effort during the observation of each website screenshot was assessed. Therefore, Poitschke's (2011) adaptation of the German scale for the assessment of subjectively perceived effort (“Skala zur Erfassung subjektiv erlebter Anstrengung (SEA)” by Eilers, Nachreiner, and Hänecke (1986), which again is based on the rating scale of mental effort (RSME) by Zijlstra and van Doorn (1985). The unidimensional scale is often used for the measurement of mental workload (Verwey & Veltman, 1996) and has shown to be sensitive to changes in mental workload (e.g. Lin & Cai, 2009; Mulder, Dijksterhuis, Stuiver, & Waard, 2009) despite its relative simplicity and fast administration. The scale was presented visually (see Figure 66) and subjects entered the according value using the keyboard.

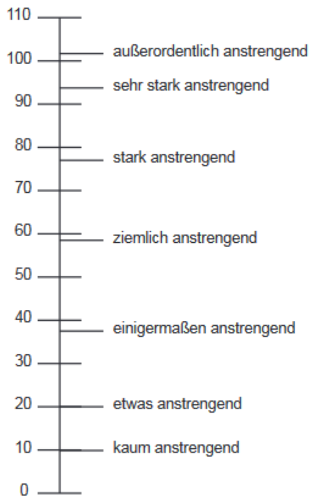


Figure 66. Adaptation of the scale for the assessment of subjectively perceived effort (SEA) by Poitschke (2011)

Apparatus

The same apparatus was used as within study 3. Details can thus be found in paragraph 5.3.3. From the collected eye tracking data, 44 ocular parameters were calculated. A picture size of 1024 × 768 pixels at an eye-to-screen distance of approximately 98 cm resulted in a visual angle of 16.3° × 12.3°.

6.3.4 Statistical Analysis

As in study 3, mixed models were used for the statistical analyses of the data. Details can thus be found within paragraph 5.3.4. The analysis of visual complexity ratings and the ocular parameters number of fixations, scanpath length and spatial density was supplemented with repeated measures correlations between visual complexity and mental workload ratings, which were calculated with the help of Bakdash and Marusich's (2017) package *rmcorr* in R (R. Core Team, 2018). Within the result plots, all error bars depict the 95% confidence interval.

6.4 Results

Subsequently, the results from this study are reported. After the manipulation check for prototypicality, visual complexity ratings are analysed with regard to the impact of several influencing variables. Subsequently, the findings on the relation between visual complexity and mental effort ratings are described. Furthermore, effects of the experimental manipulations on the ocular measures number of fixations, scanpath length and spatial density are investigated. Finally, both computational and ocular parameters are used for the prediction of mean and single visual complexity ratings.

6.4.1 Rating data

First of all, a manipulation check for prototypicality was conducted in order to ensure the validity of both prototypicality levels, so that the more prototypically designed websites are also perceived as more prototypical. According to a paired samples t-test, this was the case with prototypical screenshots (M = 3.58, SD = 1.58) rated as more prototypical than the non-prototypical screenshots (M = 2.23, SD = 1.40), $t(1343) = 26.23$, $p < 0.0001$. Data are also visualized in Figure 67.

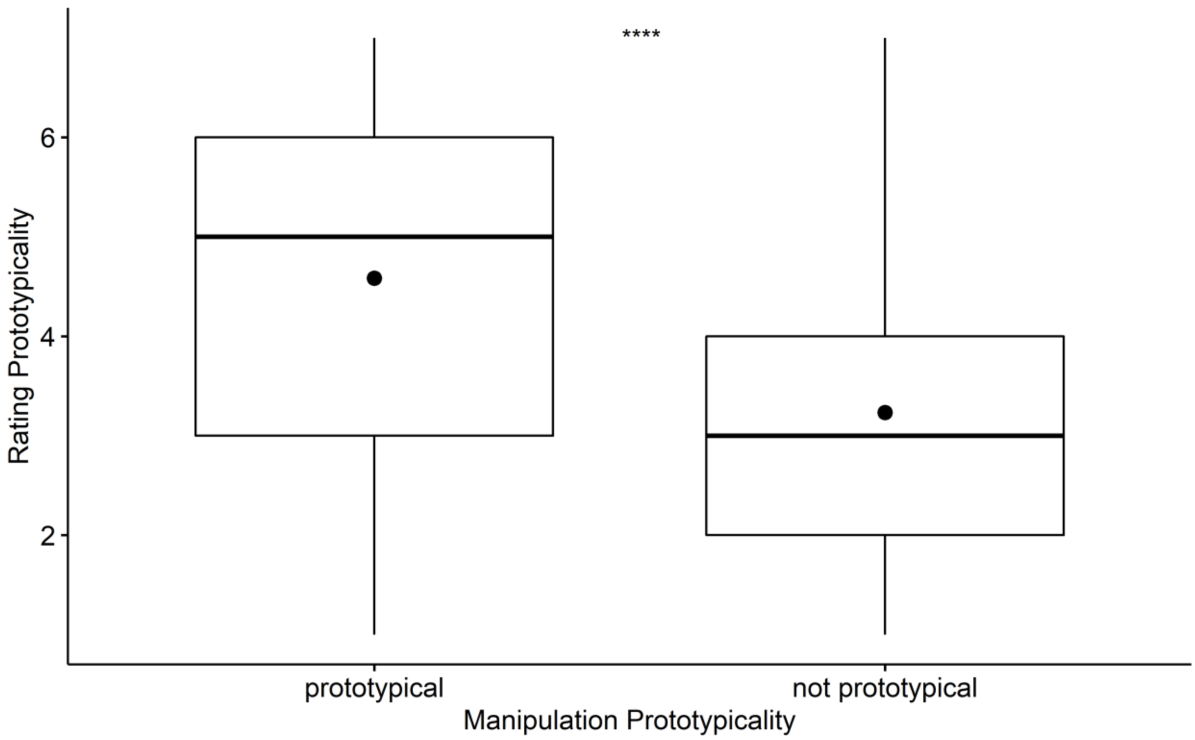


Figure 67. Manipulation check for prototypicality of website screenshots. The point within the boxplot depicts the mean and the line the median.

Consequently, effects of all four potential influencing factors on ratings of visual complexity were investigated using cumulative link mixed models with random effects for subjects, images and types of websites. First of all, a significant main effect for the

number of elements could be found, $\chi^2(2) = 179.93$, $p < .0001$. Tukey post hoc tests revealed that pictures with nine elements were rated as more complex compared to pictures with six, $\beta = 0.76$, $SE = 0.13$, $p < .0001$, and three elements, $\beta = 3.84$, $SE = 0.15$, $p < .0001$. Similarly, pictures with six elements were rated as more complex than pictures with three elements, $\beta = 3.08$, $SE = 0.14$, $p < .0001$.
Furthermore, the factor symmetry significantly affected the rating of visual complexity, $\chi^2(1) = 15.72$, $p < .0001$, with asymmetrical screenshots being rated as more complex than symmetrical ones. The two main effects number of elements and symmetry are depicted in Figure 68.

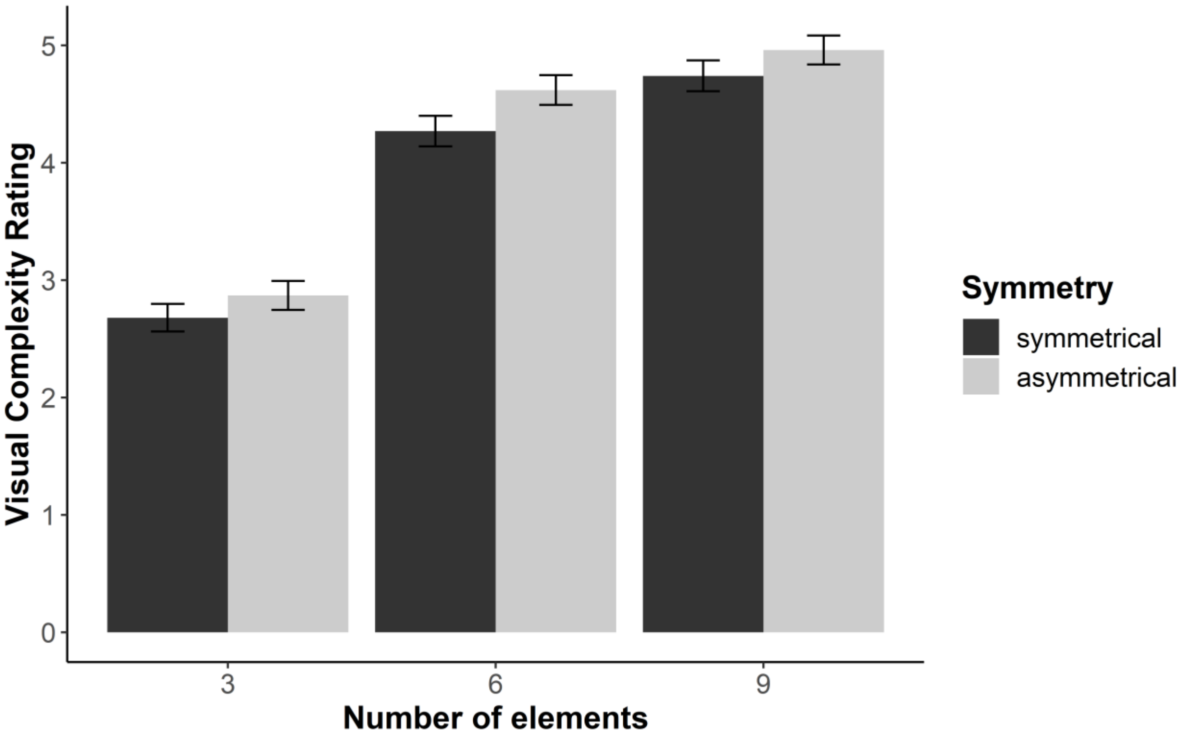


Figure 68. Effects of number of elements and symmetry on visual complexity ratings in study 4

Moreover, prototypicality had a significant effect on visual complexity ratings as well, $\chi^2(1) = 27.91$, $p < .0001$, with non-prototypical screenshots being rated as more complex than prototypical ones. This is depicted, together with the main effect of number of elements, in Figure 69. Colourfulness had no significant effect on visual complexity ratings, $\chi^2(1) = 0.08$, $p = .78$. There was also a significant ordinal interaction between symmetry and prototypicality, $\chi^2(1) = 4.87$, $p < .05$, which is visualized in Figure 70. This did however not restrict the interpretability of the two main effects.

The marginal R^2 was .33 and the conditional R^2 was .64 for the underlying regression model. The coefficients for this can be found within appendix 9.20.

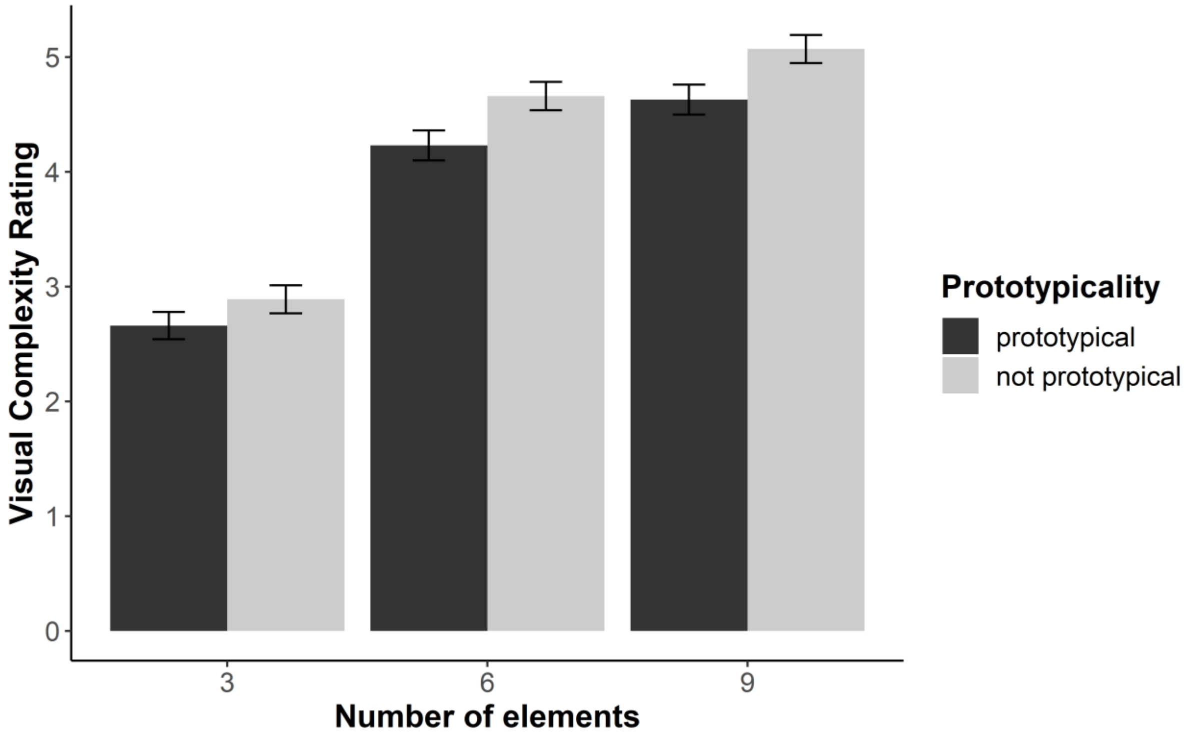


Figure 69. Effects of number of elements and prototypicality on visual complexity ratings in study 4

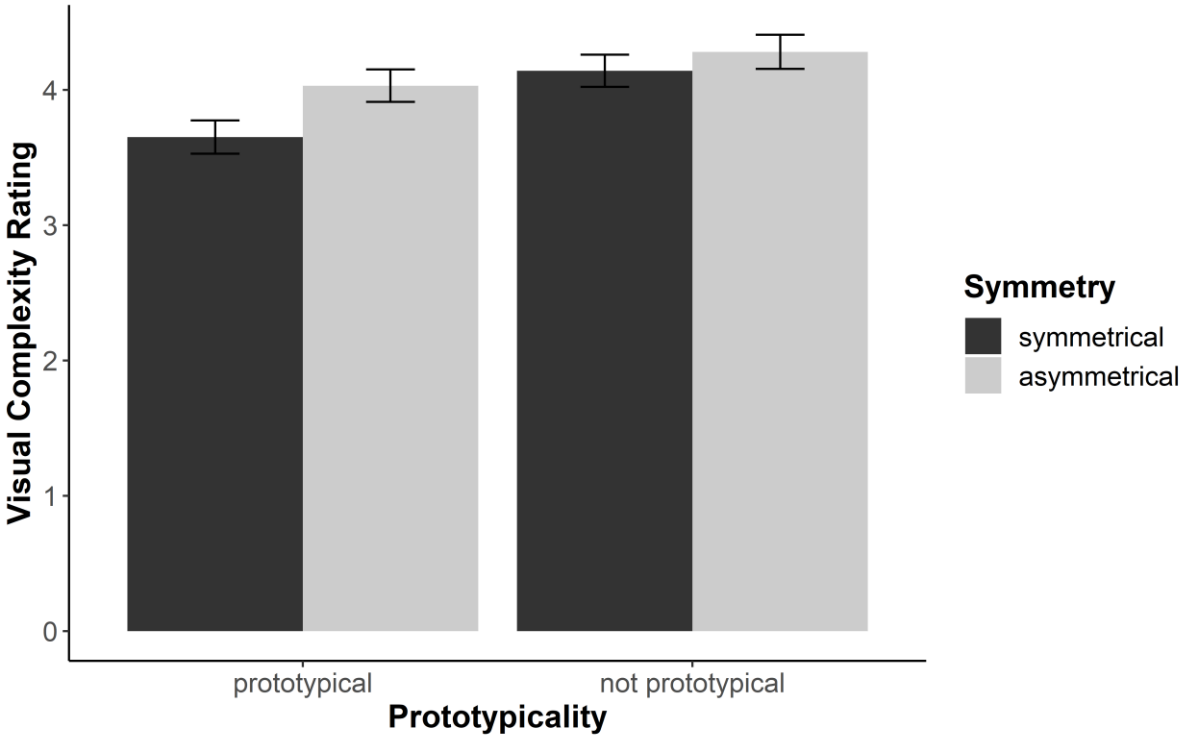


Figure 70. Ordinal interaction effect of symmetry and prototypicality on visual complexity ratings in study 4

6.4.2 Relation between visual complexity and mental workload

In order to examine the relation between visual complexity and mental workload, I calculated the repeated measures correlation between visual complexity ratings and the subjectively experience effort. This showed a significant relation, $r(2643) = .60$, $p < .0001$, which is also visualized in Figure 71.

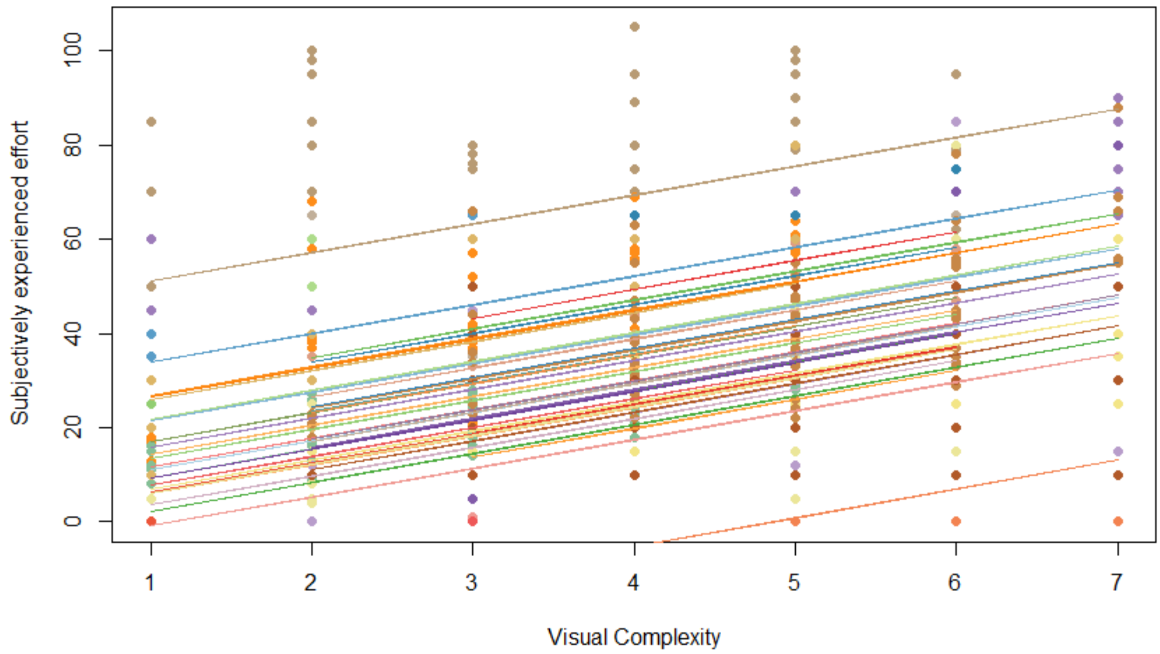


Figure 71. Visualization of the correlation between visual complexity and subjectively experience effort as a measure of mental workload. Every line represents one subject.

6.4.3 Ocular Parameters

As in study 3, effects of the experimental factors on the ocular parameters number of fixations, scanpath length and spatial density were investigated.

Number of Fixations

Firstly, the number of elements significantly affected the number of fixations, $\chi^2(2) = 55.37$, $p < .0001$. Tukey post hoc tests revealed a lower number of fixations for website screenshots with three elements than for those with nine elements, $\beta = -0.78$, $SE = 0.12$, $p < .0001$, and those with six elements, $\beta = -0.74$, $SE = 0.12$, $p < .0001$. However, there was no significant difference between screenshots with six and nine elements with regard to the number of fixations, $\beta = 0.03$, $SE = 0.12$, $p = .9557$.
Moreover, the factor symmetry had a significant effect on the number of fixations with more fixations for symmetrical screenshots, $\chi^2(1) = 10.63$, $p < .01$. This is visualized in

Figure 72 together with the main effect of the number of elements. Finally, prototypicality affected the number of fixations as well, with more fixations for prototypical screenshots, $\chi^2(1) = 12.53$, $p < .001$, which is depicted in Figure 73. Colourfulness had no significant effect on the number of fixations, $\chi^2(1) = 0.17$, $p = 0.68$, neither had the interactions between factors, p >.05. The marginal R^2 for the underlying regression model was .02, while the conditional R^2 was .46. More information about this can be found within the appendix 9.21.

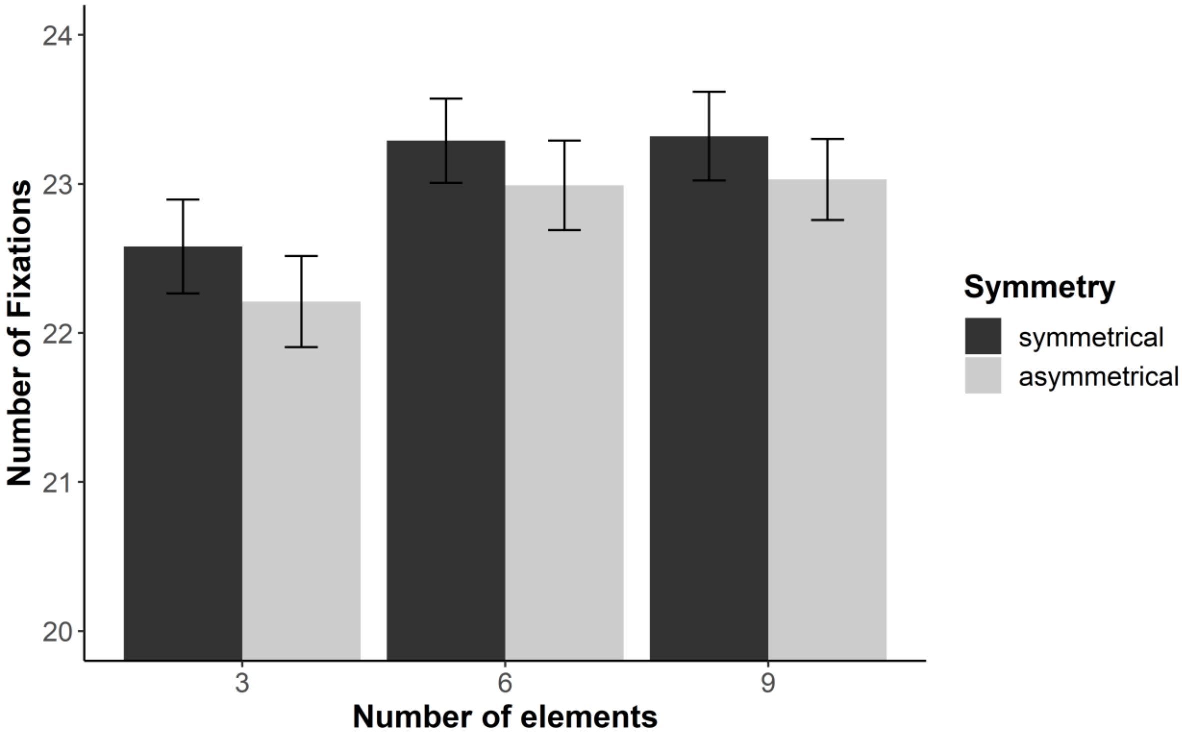


Figure 72. Number of fixations by number of elements and symmetry in study 4

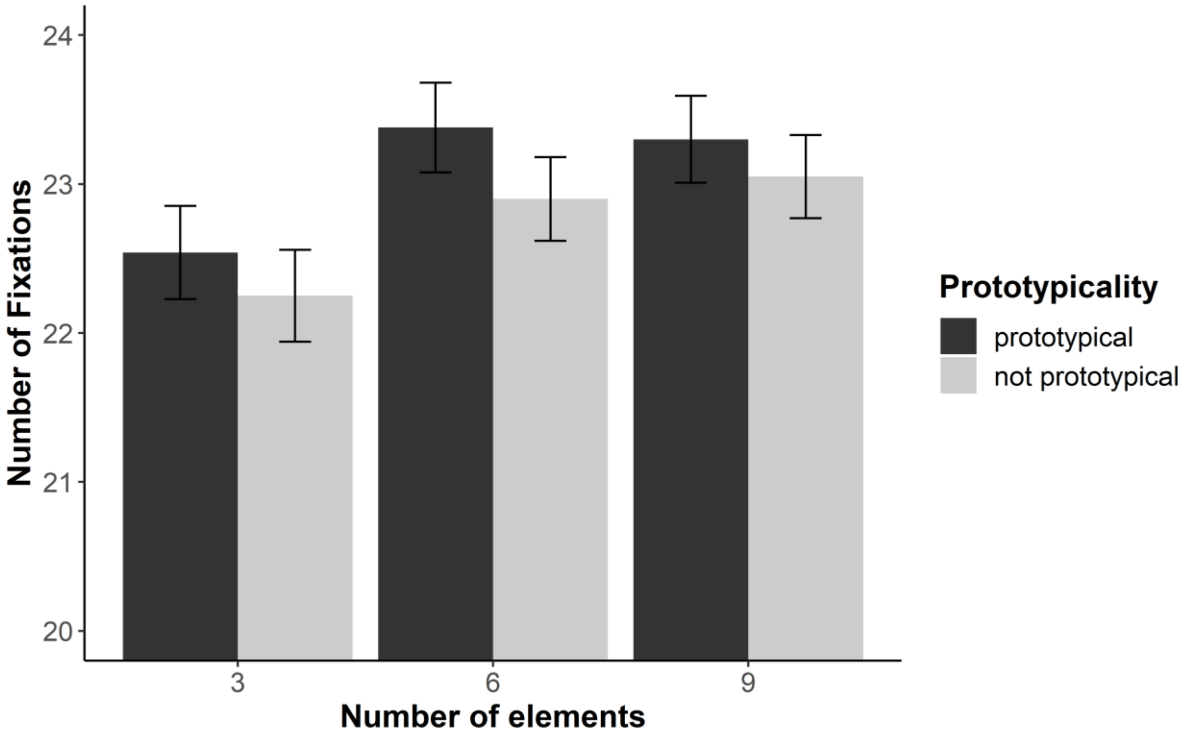


Figure 73. Number of fixations by number of elements and prototypicality in study 4

Scanpath Length

Both symmetry, $\chi^2(1) = 6.31$, $p < .05$, as well as prototypicality, $\chi^2(1) = 4.04$, $p < .05$, had a significant effect on the scanpath length, with longer scanpaths for symmetrical and non-prototypical screenshots. Effect of both factors are visualized in Figure 74. The number of elements, $\chi^2(2) = 0.83$, $p = .66$ and colourfulness, $\chi^2(1) = 0.00$, $p = .98$, had no significant effect on the scanpath length, neither had the interactions between factors, p >.05. The marginal R^2 was .02 and the conditional R^2 was .42 for the underlying regression model. More information about this can be found within the appendix 9.22.

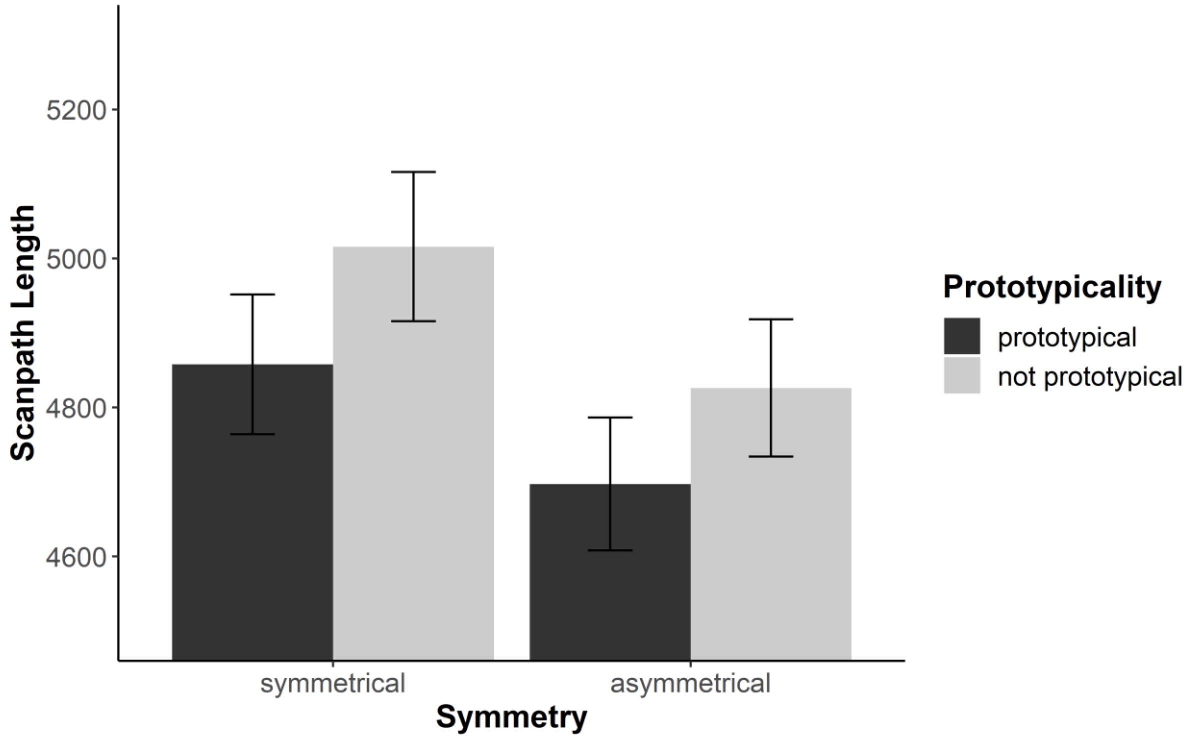


Figure 74. Scanpath length by symmetry and prototypicality in study 4

Spatial Density

Regarding spatial density, there was first of all a significant main effect of the number of elements, $\chi^2(2) = 206.43$, $p < .0001$.

Moreover, symmetry had an effect on spatial density, $\chi^2(1) = 5.43$, $p < .05$, as well as prototypicality, $\chi^2(1) = 19.61$, $p < .0001$ with larger spatial density for symmetrical and prototypical screenshots. The effect of symmetry is, together with the effect of number of elements, visualized in Figure 75.

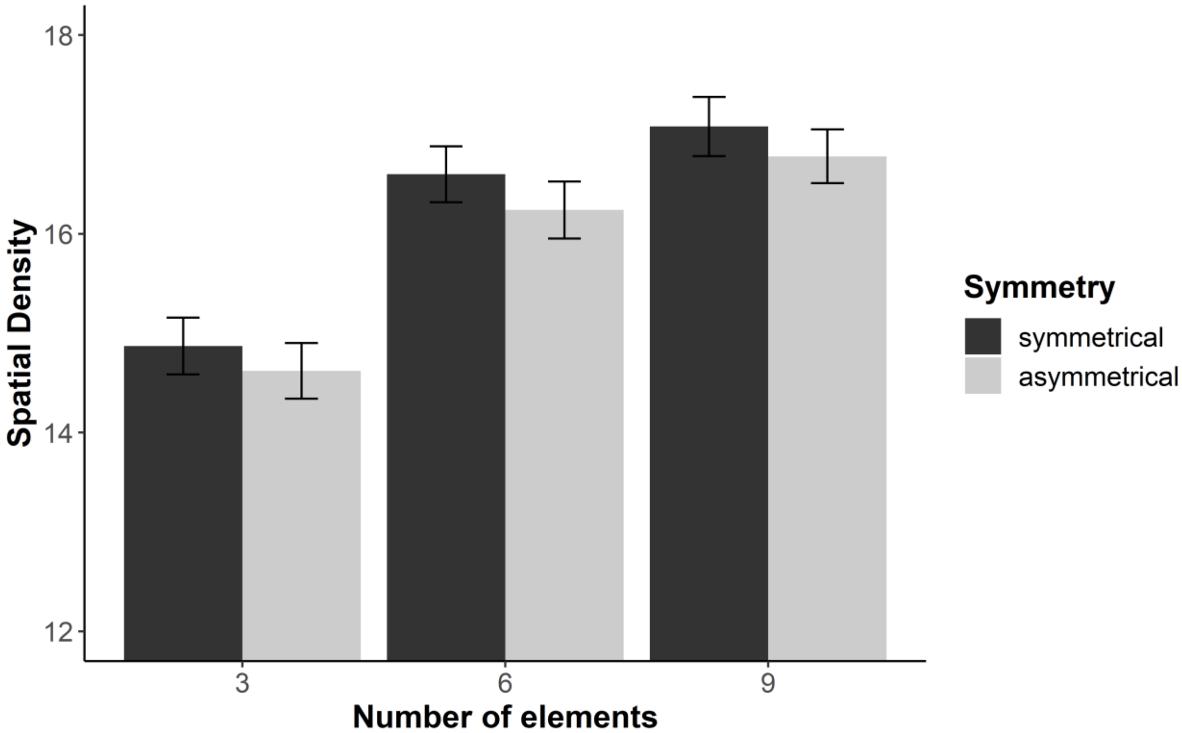


Figure 75. Spatial Density by number of elements and symmetry in study 4

There were also two significant ordinal interaction effects, one between number of elements and prototypicality, $\chi^2(2) = 7.76$, $p < .05$, and one between symmetry and prototypicality, $\chi^2(1) = 4.18$, $p < .05$. Both did however not restrict the interpretability of the main effects. The former effect is visualized in Figure 76 and the latter in Figure 77. Colourfulness had no significant effect on spatial density, $\chi^2(1) = 0.44$, $p = .51$. The marginal R^2 was .10 and the conditional R^2 was .31 for the underlying regression model, which is documented in more detail within appendix 9.23.

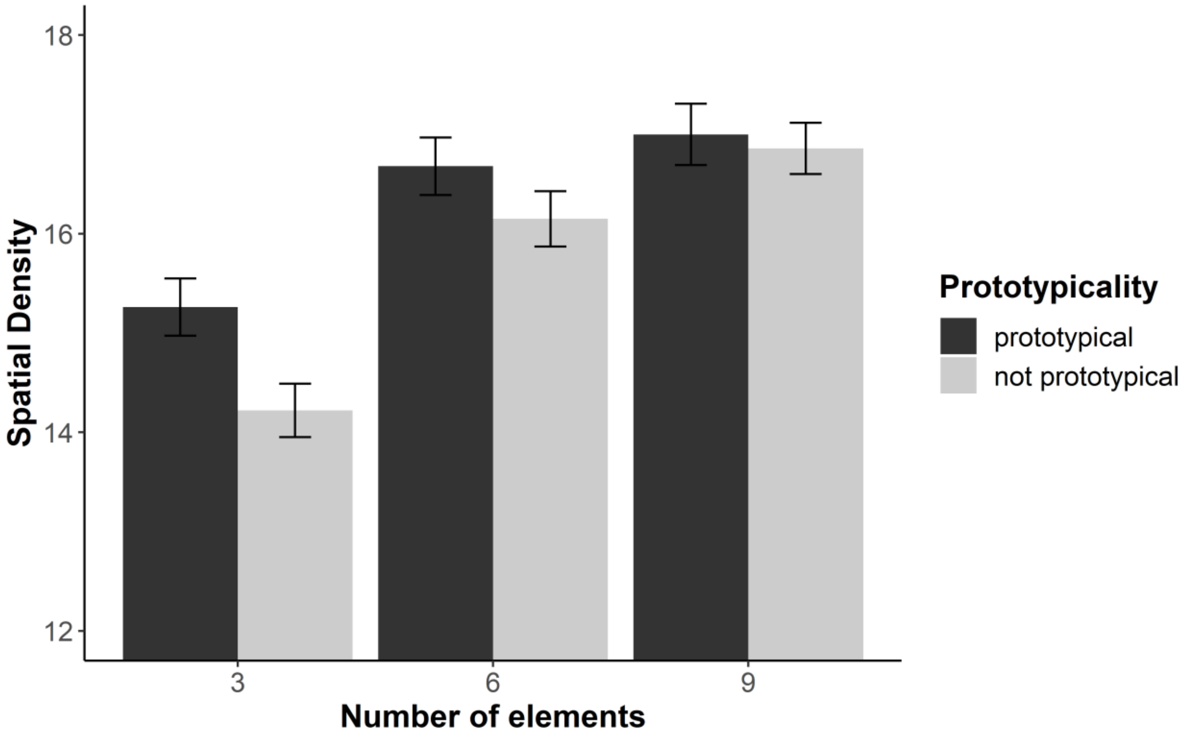


Figure 76. Spatial Density by number of elements and prototypicality in study 4

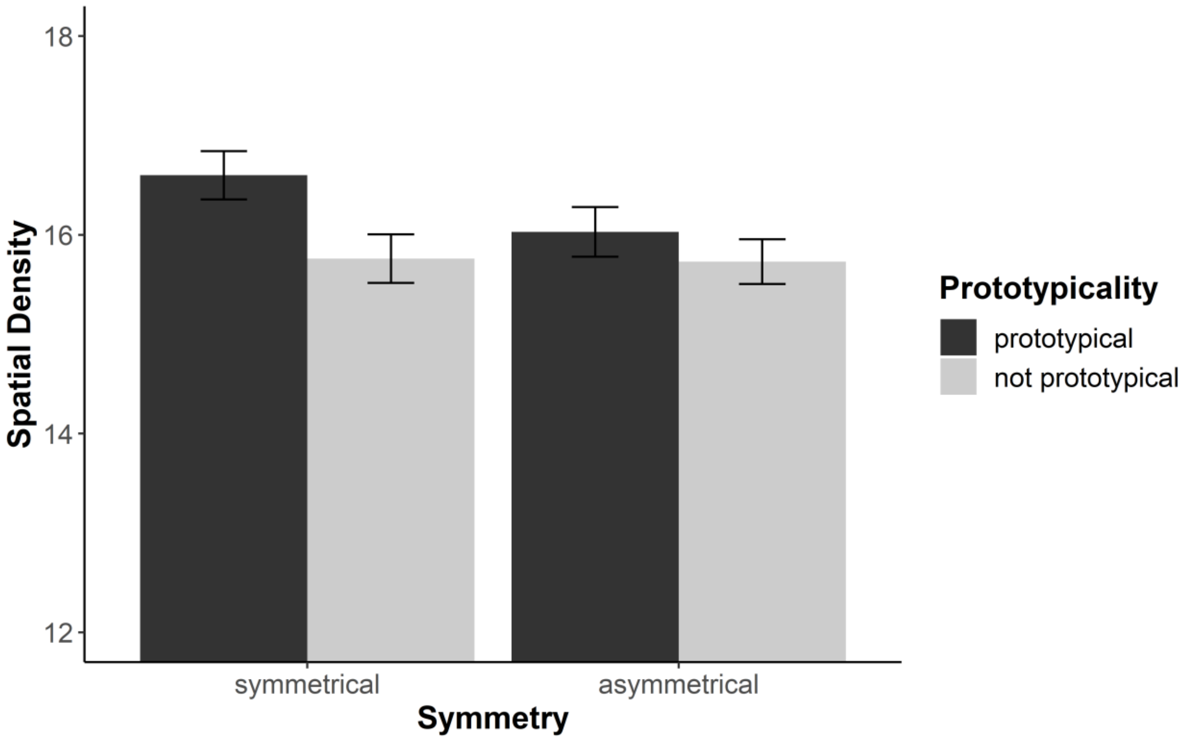


Figure 77. Spatial Density by symmetry and prototypicality in study 4

6.4.4 Prediction of visual complexity ratings

Finally, prediction models for both mean and single visual complexity ratings based on ocular and computational parameters were examined similar to the approach within study 3. For addressing mean ratings, LASSO regression (Tibshirani, 1996) was used. Within this, a variety of 183 computational measures (as described in paragraph 2.4) and 44 ocular measures (as described in paragraph 2.5.3) was considered. Firstly, the mean visual complexity ratings and ocular parameters were calculated for each image, resulting in 72 values. Subsequently, all variables were centered and scaled. With scaling is essential for LASSO regressions, this procedure also helps to prevent scaling problems and improves interpretability. Then, data were randomly split into an 80% training and 20% test set, so that prediction models could be examined with previously unknown data with regard to overfitting.

Consequently, a cross-validated lasso regression model was fit with the help of the R-package *glmnet* (Friedman et al., 2010). Within this, the penalty parameter λ was selected based on the criterion of the mean-squared error (MSE) according to the "optimal" model based on 100 penalty values Figure 78.

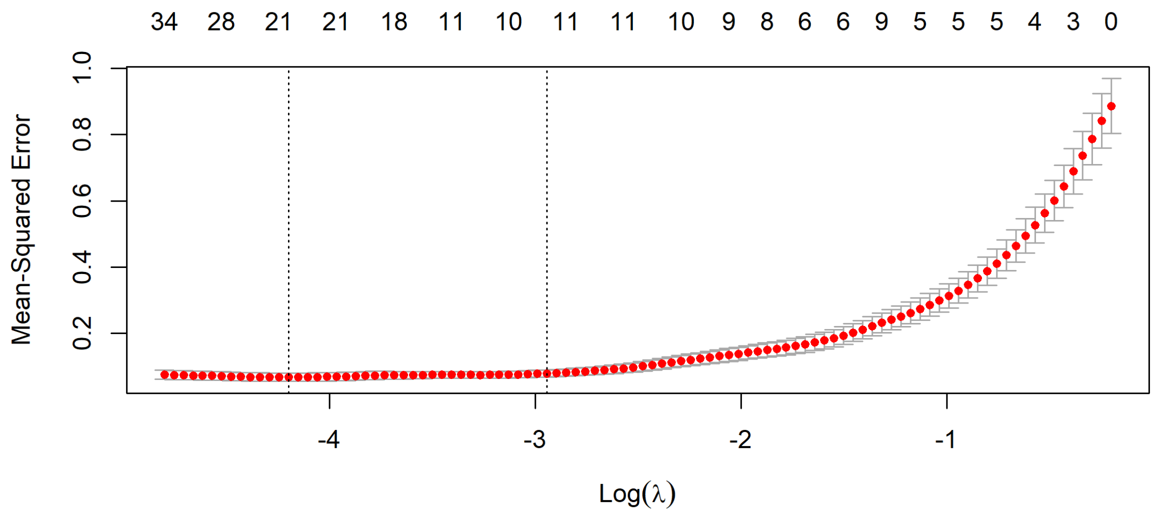


Figure 78. Mean-Squared Error (*MSE*) in relation to λ from lasso technique in study 4. The two dotted lines represent the optimal (left) and tolerance (right) fit lambda

Within the LASSO regression, 21 of the 256 variables were included within the final model. These are reported with their coefficient estimated in Table 9.
Of these, the upper five variables denote ocular parameters as described in more detail within paragraph 2.5.3. The others are computational parameters, among these combinations of compression (e.g. GIF, JPEG; TIFF) and edge (e.g. Perimeter, Canny, RMS) measures, including both mean and standard deviations of the pixel distribution within an image. Other selected variables relate to structural aspects of the image, such as symmetry or visual balance (such as the APB measure by Wilson & Chatterjee, 2005), segmentation and decomposition methods such as the number of quads as well as measures for visual clutter by Rosenholtz et al. (2007) such as Feature Congestion, Colour Congestion and Orientation Congestion.

Table 9.

Selected variables from lasso regression with coefficients for prediction of mean complexity ratings in study 4

Variable	**Coefficient estimate**
(Intercept)	3.99
Mean Velocity	-0.036
Mean Drift	0.056
Stationary Entropy	0.035
SD Nr. of Nodes	-0.081
PERCLOS	-0.011
SD Edge Phase Congruency GIF	0.049
SD Edge Phase Congruency TIFF	0.0017
SD Edge Phase Congruency PNG	0.0000000000000015
Mean x SD Edge Perimeter GIF	0.13
Mean x SD Edge Perimeter JPEG	0.0040
Vertical Symmetry	-0.17
Quads 4x4	0.35
RMS SD	0.096
APB Horizontal Inner Outer	0.013
APB Vertical	-0.022
APB Vertical Inner Outer	-0.083
Nr. of Segments	0.14
Average of Elawady et al.'s (2017) five largest symmetries	-0.12
SD for Colour Congestion clutter map	-0.0072
SD for Orientation Congestion clutter map	-0.12
Feature Congestion	0.21

Note. Due to the large differences in size, I offended the suggestions of the APA to report two decimal places but instead decided to report two valid places for each coefficient.

The correlations between the averages of 21 predictors and the visual complexity rating for all screenshots from the trial data are visualized in Figure 79.

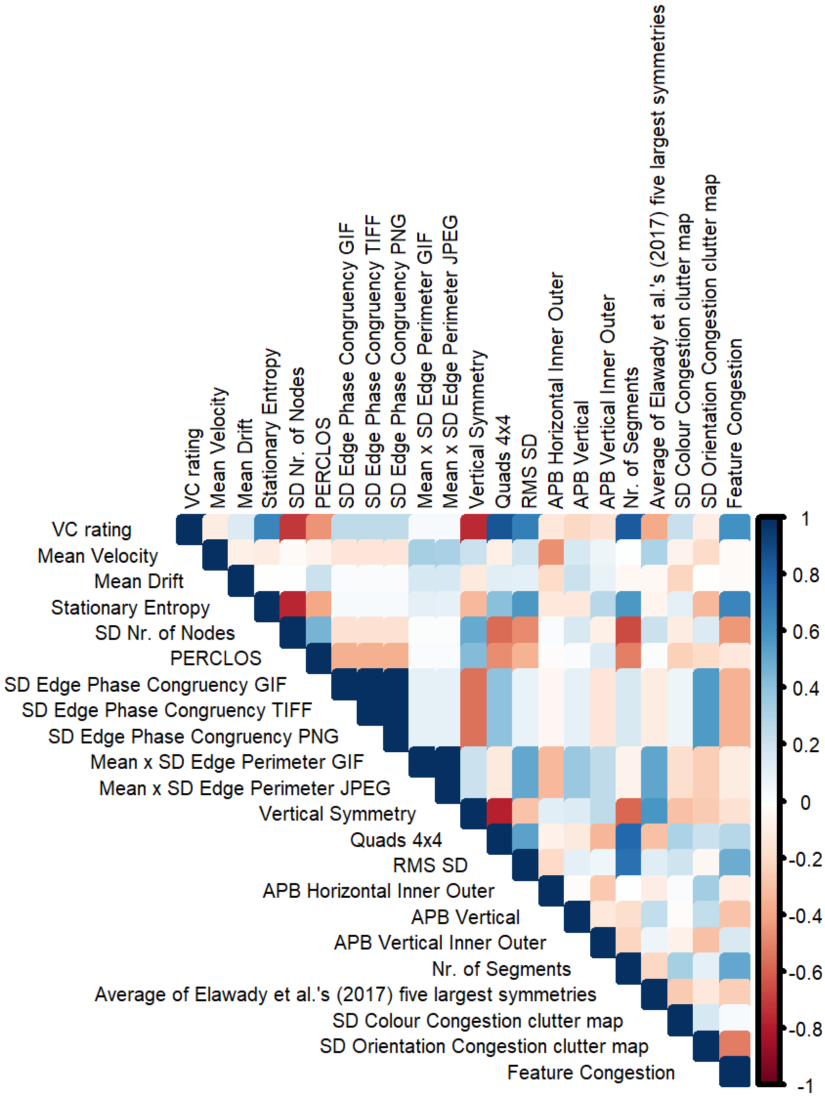


Figure 79. Correlations of selected predictors and the mean visual complexity rating in study 4

The model gave an R^2 of .97 and a mean-squared error (*MSE*) of 0.068 within the training data. It was then used for the prediction of the previously unknown test data. A plot of the actual mean visual complexity ratings versus the according predictions from the regression model is depicted in Figure 80. The R^2 within the test data was .91 and *MSE* 0.098.

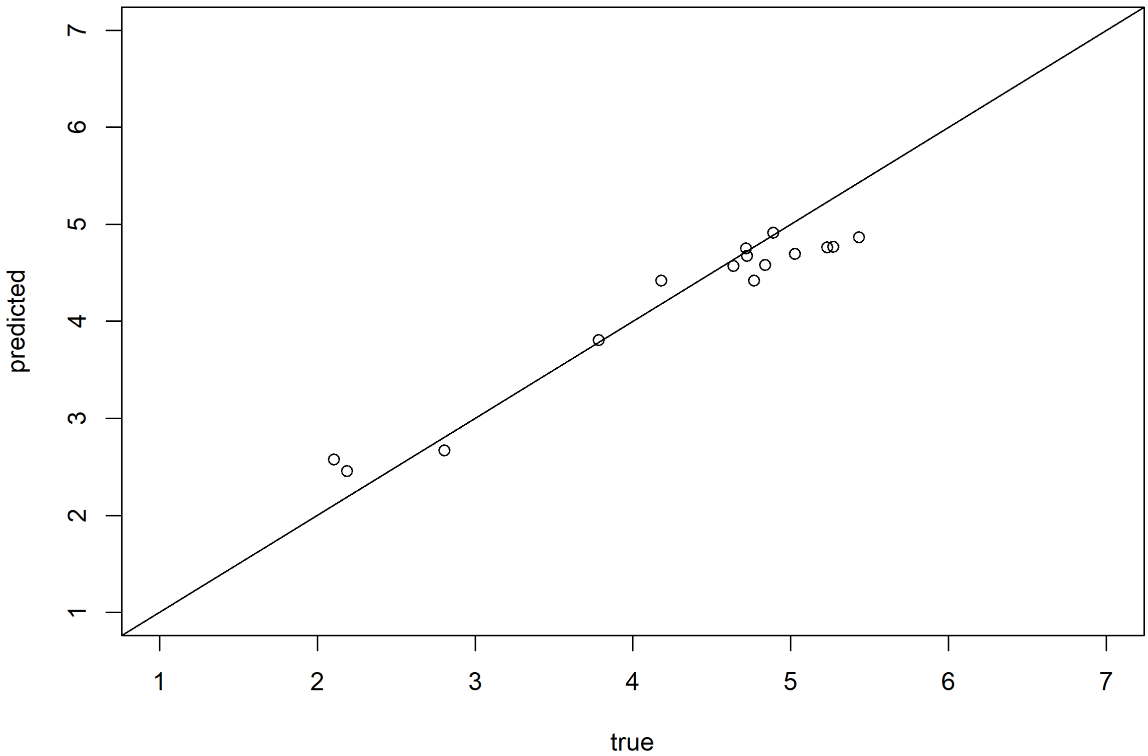


Figure 80. Predicted vs. actual values for mean visual complexity ratings in study 4

Subsequently, prediction models for individual visual complexity ratings were examined as within study 3. First of all, from the variables selected within the previous Lasso regression model for mean visual complexity ratings, SD Edge Phase Congruency TIFF, SD Edge Phase Congruency PNG and Mean x SD Edge Perimeter JPEG were removed due to collinearity. The remaining 18 variables were then included within a mixed-effects ordinal regression with a random intercept for subjects. This was calculated with the help of the function olmm from R-package *vcrpart* (Bürgin & Ritschard, 2017) for the training dataset, consisting of 80% of the total observations. Within the training data, this achieved a correct classification rate of .43. Detailed information concerning the regression model can be found within appendix 9.24.
This regression model was then used for the prediction of single visual complexity ratings within the previously unknown test data. When the subject vector is taken into account, a correction classification rate of .41 could be achieved. If this was ignored and the population-averaged response probabilities were considered instead, the

correct classification rate was .39. Confusion matrices for training and test data, the latter both with subject vector considered and ignored, are visualized within Figure 81. Next to the correct classification rate, MSE, MAE and Baccianella et al.'s (2009) macroaveraged MAE were calculated, which are reported in in Table 10.

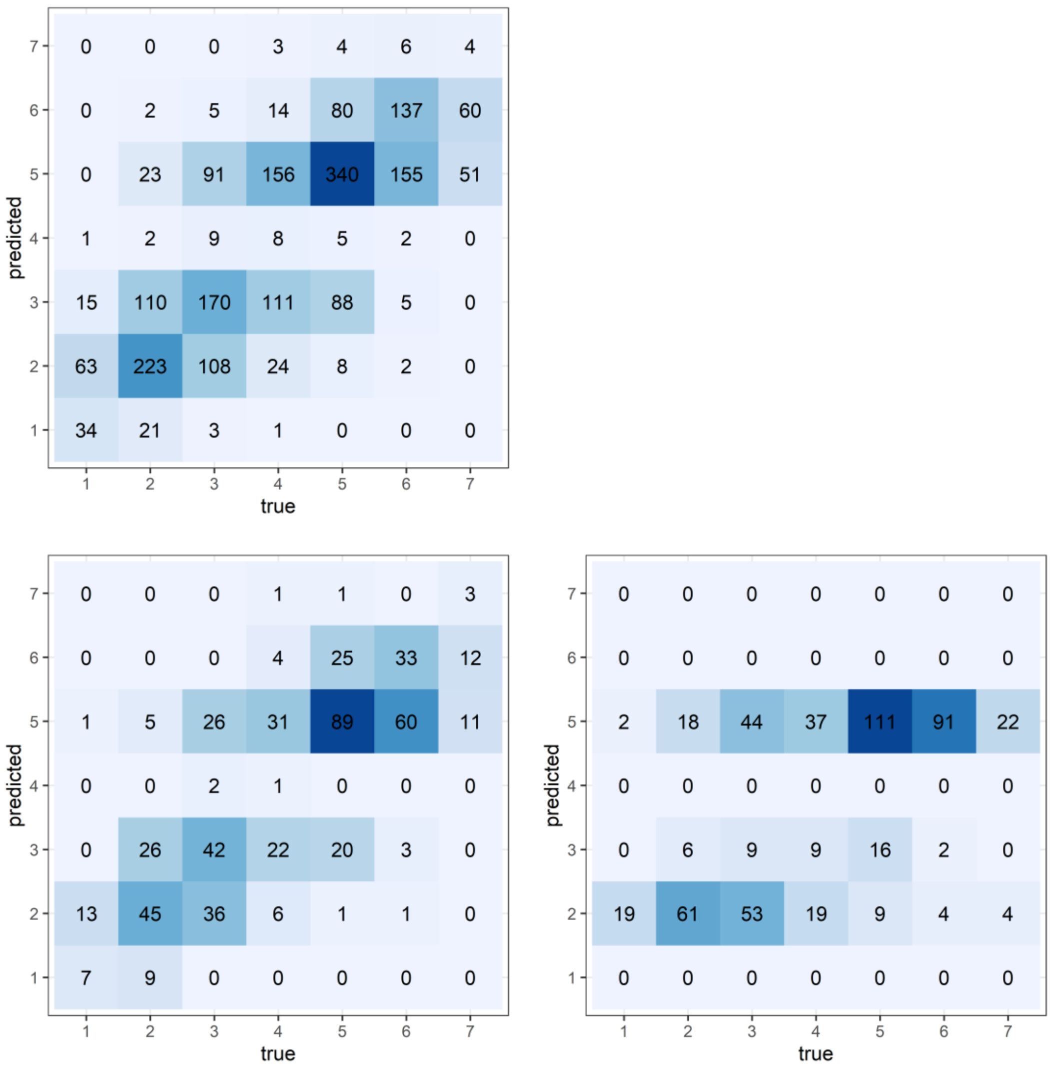


Figure 81. Confusion matrices of ordinal mixed regression for single visual complexity ratings within training data (top) and test data with subject vector considered (bottom left) and ignored (bottom right) in study 4

Subsequently, a random forest model was fitted for the training data with the help of the R-package *ranger* (Wright & Ziegler, 2017), using all available variables. The parameters *number of variables randomly sampled as candidates at each split* (mtry), *minimum node size* and *sample size* were tuned within a grid search in order to improve prediction performance. This gave the best results with a correct classification rate of .33 for an mtry of 20, a minimum node size of 10 and a sample fraction of .60 with 500 trees. The variable importance values (Hastie et al., 2017) for the top 30 variables in the model are visualized within Figure 82.

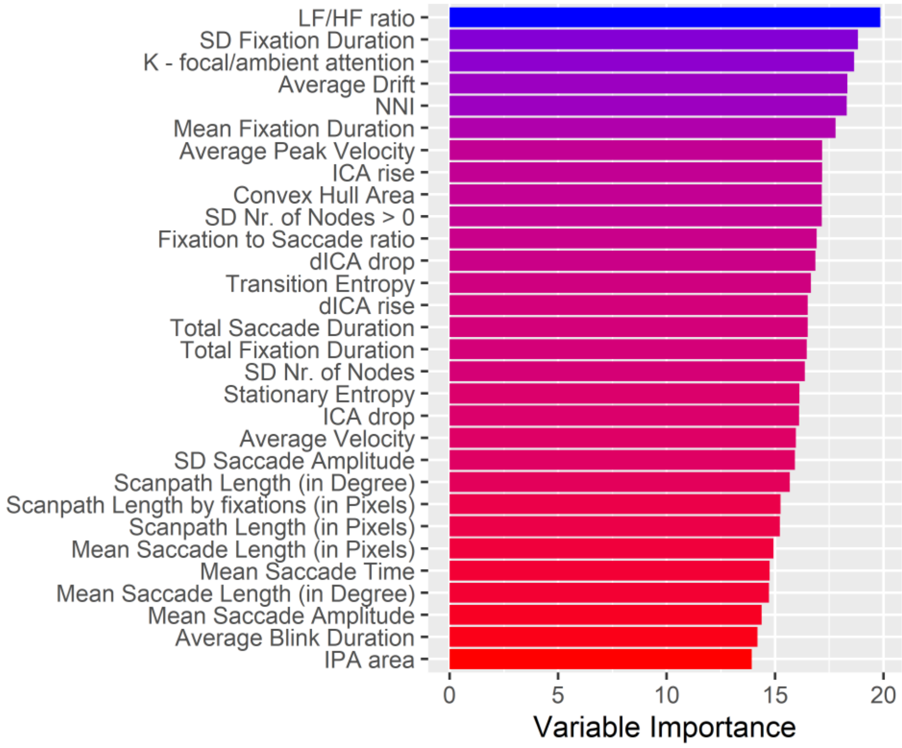


Figure 82. Variable importance values in the final random forest model for the prediction of single visual complexity ratings in study 4

The final random forest allowed for a correct classification rate of .33 within the test data. All evaluation measures for both training and test data are reported in Table 10. Confusion matrices are visualized in Figure 83.

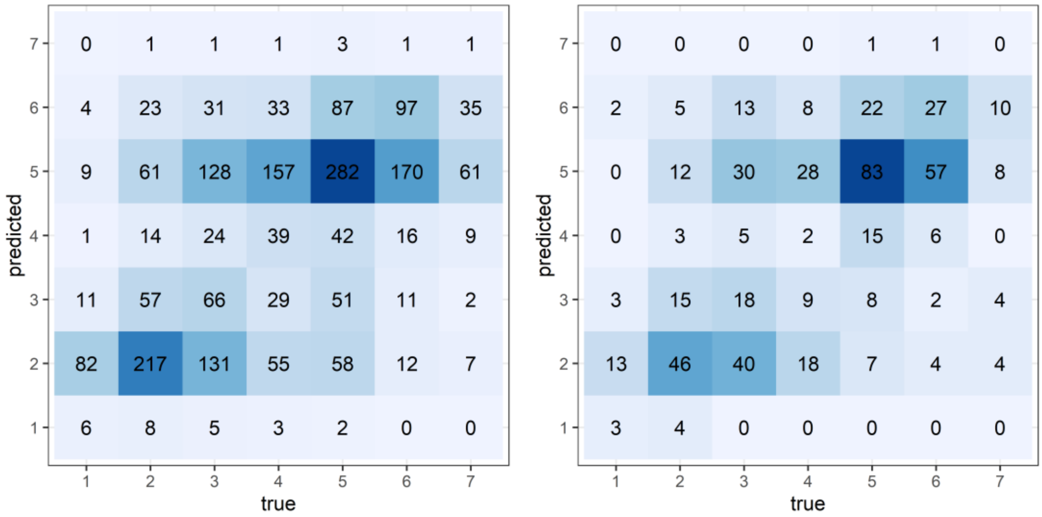


Figure 83. Confusion matrices of random forest for single visual complexity ratings within training (left) and test data (right) in study 4

As the last method, generalized linear mixed models with Lasso variable selection (glmmLasso) were applied within a similar process as within study 3. After the optimal penalty parameter λ was determined based on the Bayesian information criterion (BIC) (Schwarz, 1978), the final models were built based on the optimal parameters.

First of all, a model with two random effects for both subjects and stimuli was calculated. The final model included the variables Scanpath Length (by Fixations), Vertical Symmetry, Quads 4x4, Quads 8x8, RMS SD, RMS Mean x SD, RMS JPEG Size, Nr. of Segments, Näsänen Complexity as well as Orientation, Feature Congestion map Filesize and Orientation map Filesize as referring to Rosenholtz et al. (2007). Detailed information regarding the model and coefficients can be found in appendix 9.25. With the help of this model, 44% of the ratings in the training and 45% of the test data could be correctly classified. A detailed overview of all evaluation measures for both training and test data is reported in Table 10. Confusion matrices are visualized in Figure 84.

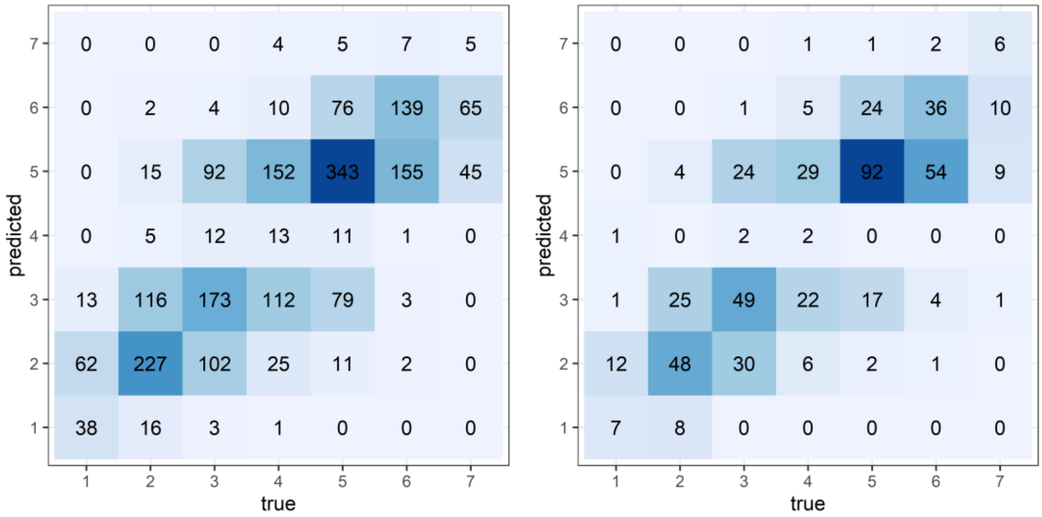


Figure 84. Confusion matrices of glmmLasso with random effects for subjects and stimuli for single visual complexity ratings within training (left) and test data (right) in study 4

Secondly, a model with only a random intercept for subjects was calculated, which encompassed the variables Quads 4x4, Quads 8x8, RMS JPEG Size, Näsänen Complexity and Orientation. Detailed information regarding the model and coefficients can be found in appendix 9.26. This achieved a correct classification rate of .41 within the training data. Within the test data, this model gave a correct classification rate of .41. A detailed overview of all evaluation measures for both training and test data is reported in Table 10. Confusion matrices are visualized in Figure 85.

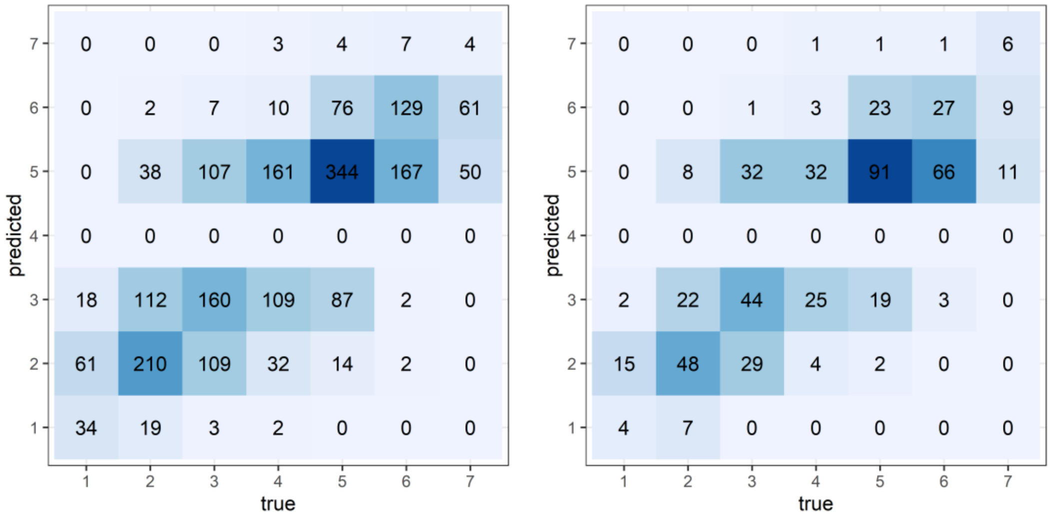


Figure 85. Confusion matrices of glmmLasso with a random effect for subjects for single visual complexity ratings within training (left) and test data (right) in study 4

The third glmmLasso model encompassed no random effects in order to investigate the prediction performance without considering any additional information apart from the automated and ocular measures. Thereby, the eleven variables Number of Fixations, Vertical Symmetry, Quads 4x4, Quads 8x8, RMS SD, RMS JPEG Size, Number of Segments, Spatial Frequency, Näsänen Complexity, Orientation and Feature Congestion were included within the final model, which achieved a correct classification rate of .33 within the training data. Detailed information regarding the model and coefficients are reported within appendix 9.27. Within the test data, this model gave a correct classification rate of .36. An overview of all evaluation measures for both training and test data is reported in Table 10. Confusion matrices are visualized in Figure 86.

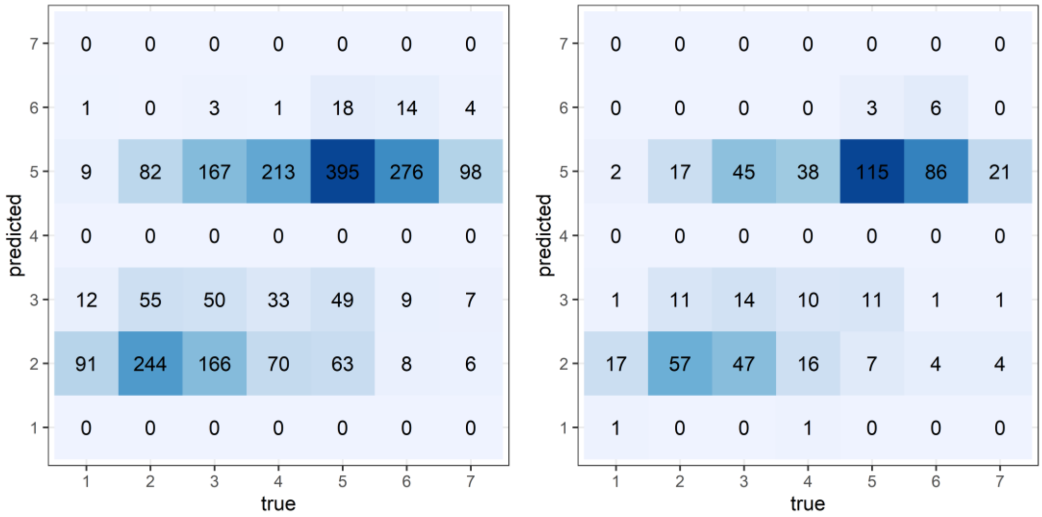


Figure 86. Confusion matrices of glmmLasso without random effects for single visual complexity ratings within training (left) and test data (right) in study 4

Table 10.

Evaluation measures of different models for the prediction of single visual complexity ratings for both training and test data within study 4

Method	**Corr. Class**	***MAE***	***MSE***	***MAE* M**	***MAE* M categorywise**						
					Category						
					1	2	3	4	5	6	7
Training data											
olmm train	.43	0.76	1.18	0.85	0.85	0.56	0.83	1.12	0.56	0.61	1.41
RF train	.33	1.10	2.33	1.23	1.44	0.98	1.34	1.18	0.80	0.93	1.97
glmmLasso train RE Sub Stim	**.44**	**.73**	**1.12**	**0.81**	0.78	0.51	0.81	1.11	0.55	0.59	1.35
glmmLasso train RE Sub	.41	0.83	1.30	0.88	0.86	0.66	0.91	1.16	0.57	0.61	1.40
glmmLasso train no RE	.33	1.05	2.06	1.23	1.38	0.79	1.32	1.22	0.58	1.09	2.24
Test data											
olmm test RE Sub	.41	0.76	1.18	0.86	0.81	0.59	0.85	1.17	0.51	0.75	1.31
olmm test averaged	.34	1.02	2.01	1.24	1.29	0.71	1.33	1.29	0.43	1.16	2.46
RF test	.33	1.07	2.28	1.28	1.38	0.95	1.36	1.37	0.56	0.95	2.38
glmmLasso test RE Sub Stim	**.45**	**.73**	**1.15**	**0.82**	0.81	0.53	0.78	1.17	0.49	0.74	1.23
glmmLasso test RE Sub	.41	0.78	1.22	0.87	0.90	0.62	0.91	1.14	0.51	0.78	1.19
glmmLasso test no RE	.36	.98	1.93	1.22	1.29	0.73	1.29	1.28	0.34	1.08	2.54

Note. Corr. Class: Correct classification rate, *MAE*: mean absolute error, *MSE*: mean standard error, *MAE* $_M$: *MAE* macroaveraged according to Baccianella et al. (2009), *MAE* $_{M\ categorywise}$: MAE for the response levels of the seven-point Likert scale separately, RE: random effect, Sub: Subject, Stim: Stimulus, the methods and overall measures with the best performance in both training and test data are highlighted in bold

6.5 Discussion

Within this study, an experimental approach was used in order to investigate the impact of several influencing variables on visual complexity ratings as well as on ocular parameters within the applied context of website screenshots. Stimuli were created by adapting three real websites of the type shopping, company and news according to the experimental design. These were presented to 40 subjects, whose ratings and gaze behaviour was assessed. Ocular parameters were also combined with computational measures in order to predict both mean and single visual complexity ratings. This constitutes an extension of the approach used within study 3, where only basic black and white shape patterns served as stimuli. Within the following, a conclusion of the findings from this study is given before limitations are discussed and an outlook is presented.

Conclusion

Similar to study 3, results from this study again revealed a strong influence of both the number of elements as well as symmetry on visual complexity ratings. Thereby, asymmetrical screenshots and screenshots with more objects were rated as more complex. Moreover, the prototypicality of the website screenshots also had a considerable impact on the visual complexity ratings with less prototypical websites being rated as more complex. For the manipulation of prototypicality, the structural configuration of different website elements such as the logo, navigation area or search function was adapted, so that these were either placed at expected or unexpected locations. According to the results of the manipulation check, this approach appeared to successfully influence the perception of prototypicality, which may be due to the influence of subjects' mental models. These are based on previous experience and can allow for inferences and predictions for example regarding the typical design of websites (Roth et al., 2010). As shown within the study, prototypicality can also influence the perception of visual complexity. This is particularly interesting with respect to Leder et al.'s (2004) model of aesthetical experience. Within this, the implicit memory integration, which also encompasses the processing of prototypicality, sequentially follows the perceptual analysis, which includes the processing of complexity (see also image section in Figure 87).

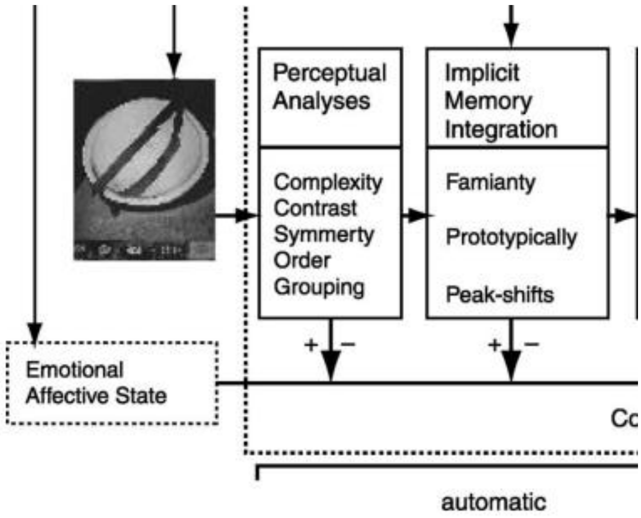


Figure 87. Image section of Leder et al.'s (2004) model of aesthetic processing

Although the authors stress that “it is important to note that the model does not depict a strict serial flow of information.” (p. 493), the presumably bidirectional relations between visual complexity and prototypicality are notable and may suggest that the perception of visual complexity does not merely rely on bottom-up processing, which is based on stimulus features. The visual features in prototypical and non-prototypical websites screenshots do not differ, these are only differently arranged. Instead of pure bottom-up processing, influences based on previous experience and resulting expectations that go beyond a purely perceptual analysis also seem to affect the perception of visual complexity.

With regard to the effects of colourfulness, study 2 suggested the relevance of an additional colour-related dimension within visual complexity. While both colour contrast and the variety of colours were significantly related to visual complexity for photographs, a regression revealed less clear coherences for website screenshots. In order to address this issue in more detail, the influence of colourfulness was experimentally investigated within this study. Thereby, no significant effect on visual complexity ratings could be found, for which there are mainly two possible explanations. The first option is that colourfulness has no or only a very small impact on the visual complexity perception of websites, which might also be in line with the correlational findings from website screenshots within study 2, while significant relations were found for photographs. Another possible explanation for this finding is that the manipulation of colourfulness was too subtle so that differences were hardly perceived by the subjects. The manipulation followed Hasler and Suesstrunk's (2003) method for the quantification of colourfulness with mean colourfulness values of 46.83 for colourful images, which according to their classification is near to the mark of 59 meaning “quite colourful” and 26.04 for less colourful images, which is closer to zero meaning “not colourful”. Although a difference between the calculated colourfulness values of both groups can be seen, this may have had too little effect regarding the actual perceptual impression. Of course, the differences between colourfulness groups could have easily been adjusted more strongly. However, the created websites screenshots should also still look realistic and neither appear extremely colourful nor too colourless. Since the differences between the two colourfulness conditions are rather low, effects of colourfulness might have been observed if stronger differences or different methods for the manipulation would have been used.

Interaction Symmetry x Prototypicality

In addition to the main effects, two significant interactions were identified. Firstly, the ordinal interaction between symmetry and prototypicality suggested that symmetry had a larger effect on visual complexity ratings for prototypical images than for non-prototypical images. This might be related to the association between symmetry and prototypicality. While both factors were manipulated independently by focussing on the element clusters within the website for symmetry and the menu items for prototypicality, particularly asymmetrical prototypical stimuli may rarely appear in reality, since prototypical design of a website might often also include the symmetry of its elements. This aspect might partially counteract the manipulation and lead to a reduced perception of prototypicality in relation to the other stimuli, which might have increased the visual complexity rating and thus explain the interaction effect.

Relation between visual complexity and mental workload

Next to the effects of influencing factors on visual complexity ratings, the relation between the latter and mental workload were analysed. The strong positive correlation between both is also in line with the findings from study 1, where video complexity significantly affected the mental workload of operators. In sum, the results from this study can be interpreted as underlining the role of visual complexity in the context of human-machine interaction, since the association was also shown to be particularly valid within user interfaces such as websites. The results are additionally considerable, since subjects within this study had no further experimental task apart from the observation of screenshots. It might be hypothesized that effects of visual complexity might have had even larger effects on mental workload when subjects had to perform a search task for example.

Eye tracking

From the gathered eye tracking data, effects of the influencing factors were analysed regarding three parameters: number of fixations, scanpath length and spatial density.
Within the number of fixations, it could be shown that three factors significantly affected this measure. First of all, more fixations were found for screenshots with a larger number of objects. The difference between three and six as well as between three and nine elements were significant, while no significant difference was found between six and nine elements. This may relate to the smaller differences in visual complexity ratings between six and nine elements. One possible reason for these findings may be the limited capacity of the short-term memory. While Miller (1956) proposed a capacity of seven plus or minus two information chunks, more recent literature suggests rather lower amounts (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997), with visual information of the single objects affecting the capacity limits as well. The reaching and exceeding of short-term capacity limits might thus reduce the increment of visual exploration of the website screenshots and cause the lower differences in the number of fixations between six and nine elements.
Moreover, both symmetry and prototypicality affected the number of fixations. A larger number of fixations was found for both prototypical and symmetrical websites screenshots. This might appear rather surprising at first, since a larger number of fixations was found for asymmetrical images within study 3, using rather simple shape patterns. Within this study however, the result pattern was reversed. This may relate to the higher complexity of the website screenshots compared to the relatively simple shape patterns used in study 3. Along with this, different manipulations of symmetry were applied. Within study 3, mirror symmetry along the image axes was used for the manipulation of the rather basic stimuli, while this was hardly viable for naturalistic stimuli such as website screenshots in study 4. Thus, Elawady et al.'s (2017) method was used for the symmetry assessment. With regard to the different types of symmetry

considered within the two studies, the results can be related to a number of findings from previous research literature. First of all, Locher and Nodine (1989) differentiated between static and dynamic symmetry. According to the authors, static symmetry describes “the exact duplication of structural elements about an axis of symmetry” (Locher & Nodine, 1989, p. 476). This perfectly describes the stimuli from the symmetrical category within study 3, since these were mirror-symmetric with regard to several axes. Dynamic symmetry however “is achieved by differentially weighting and counterweighting distributions of compositional elements about an imaginary axis of symmetry which serves as a fulcrum” (Locher & Nodine, 1989, p. 476). It can thus be seen as a more deliberate form of symmetry. Findings of the authors point towards a restriction of visual exploration by static symmetry, while dynamic symmetry enhances it. Locher and Nodine (1989) explain this by the fact that stimuli with static symmetry contain fewer unique elements. Due to the mirror symmetry, both sides of an axis are identical, therefore elements are not unique. This is however not the case for dynamic symmetry. The authors suggest that the latter encourages visual exploration out of curiosity and in order to reduce information uncertainty. Moreover, using naturalistic images such as street and natural scenes, Kootstra, Boer, and Schomaker (2011) showed that particularly early fixations often fall within symmetrical areas of a picture. This firstly emphasizes the role of symmetry within early cognitive processes, but also complements Locher and Nodine's (1989) by incorporating results from gaze data. In sum, the previous references can be helpful in order to explain the at first sight contradictory results of study 3 and 4 and in particular might give a reason for the larger number of fixations for symmetrical screenshots within the latter, which may indicate enhanced visual exploration. Moreover, they might also help to explain the effect of prototypicality with more fixations for prototypical screenshots. Prototypicality, similar to dynamic symmetry, might similarly encourage the visual exploration of the website, since it provides a familiar framework for the observer in contrast to non-prototypical screenshots.

With regard to the scanpath length, significant effects of symmetry and prototypicality were found as well. While scanpaths were longer for symmetrical screenshots, scanpaths for prototypical images were shorter. The positive relation with symmetry may be explained by an enhanced visual exploration for (dynamic) symmetrical screenshots similarly to the effects on the number of fixations. The finding of shorter scanpaths for prototypical images in combination with a larger number of fixations however seems surprising at first sight. Yet, the arrangement of elements within the non-prototypical condition, where for example the navigation menu is placed at the bottom of a website, might have triggered longer saccades because of the larger distance with other elements as compared to the prototypical condition. However, no significant effects of number of elements and colourfulness on scanpath length were found. For the number of elements, a positive effect on scanpath length would have been expected. This could be justified first of all by the findings from study 3, where this pattern was identified. Moreover, within serial processing as according to Treisman and Gelade's (1980) feature-integration theory for example, more objects would also require a more extensive consecutive scanning and thus longer scanpaths in order to construct a mental representation of the picture. The higher complexity level of stimuli within this study however might have worked against this effect, while both symmetry and prototypicality revealed an impact on scanpath length.

Finally, multiple effects could be found regarding the measure spatial density. First of all, the spatial density was larger the higher the number of elements in the screenshot was. This larger spatial density likely indicates a larger spatial distribution of fixations due to the higher number of elements, which occupy broader areas of the screenshot. Moreover, symmetry had a significant effect on spatial density with larger spatial density for symmetrical screenshots. Similar to the effects of the number of fixations as well as regarding the results for scanpath length, this might again be explained by the enhanced visual exploration of the stimulus due to dynamic symmetry (cf. Locher & Nodine, 1989). Moreover, there was an ordinal interaction between symmetry and prototypicality, showing a larger effect of prototypicality for symmetrical stimuli. Further research is needed in order to explain this interaction effect.

Prototypicality also affected spatial density with larger values for prototypical screenshots. Similar to the findings on the number of fixations and effects of symmetry, this could also reflect an enhanced visual exploration. Within an ordinal interaction effect of number of elements and prototypicality however, it was revealed that the influence of prototypicality was more pronounced among stimuli with fewer objects while there was only little effect of prototypicality for screenshots with nine elements. This could relate to the spatial distribution of elements within the screenshots. Due to the larger number of elements, the spatial distribution of fixations was generally larger within screenshots with nine elements, while the impact of the placement of the logo, search field and newsletter link might have been diminished. A possible reason for this might be that the majority of the fixations in screenshots with nine elements might have focussed on the elements, while the relevance of and visual attention towards other areas such as the logo and navigation area may have increased for screenshots with less elements. This may be a possible explanation for the larger impact of prototypicality on spatial density for screenshots with a smaller number of elements, however other possible explanatory approaches remain to be investigated.

Prediction

Next to the experimental analyses within this study, prediction models for both mean and single visual complexity ratings based on computational and ocular parameters were examined.

Similar to study 3, a LASSO regression model was used for mean visual complexity ratings, which yielded very accurate predictions. Within this, 21 measures were selected, among these five ocular parameters and 16 computational measures. The selected ocular parameters were Mean Velocity, Mean Drift, Stationary Entropy, SD Number of Nodes and PERCLOS. Particularly with regard to Mean Drift, it might appear rather surprising that this was included within the model by the LASSO method, since it has rarely been studied within the context of visual perception to the best of my knowledge but is mostly rather seen as an artefact (Duchowski, 2017). Concerning computational measures, variables from different groups such as compression (e.g. GIF, JPEG; TIFF) and edge (e.g. Perimeter, Canny, RMS) measures, structural image aspects such as symmetry or visual balance, segmentation and decomposition methods such as the number of quads as well as measures for visual clutter by Rosenholtz et al. (2007) were selected. Thereby, particularly the variables Quads 4x4, Feature Congestion and Vertical symmetry showed larger regression coefficients. The accuracy of the model is comparable to study 3. Regarding the selected measures, there are some communalities but also differences. Within both studies, combinations of edge and image compression, symmetry and visual balance measures as well as quad numbers of different sizes from quadtree decomposition were included. Within this study, (sub-)measures for visual clutter by Rosenholtz et al. (2007) seemed to be relevant, however these were not calculated for study 3 since these are only applicable to colour images. Regarding ocular parameters, the selected parameters largely differ between the two studies, which may go back to the different types of stimulus images that were used.

Overall, the highly accurate prediction of mean visual complexity ratings is considerable, given that further potential aspects such as prototypicality contributed to the perception and rating of visual complexity, as found within the experimental part of this study. Since these are based on expectations or mental models of the observer, they can yet hardly be represented by computational parameters, however these might at least partially be assessed by ocular parameters. Averaging both these ratings as well as ocular parameters across participants however seemed to counterbalance interindividual differences and allowed for very accurate predictions of mean visual complexity ratings. Findings from using this methodology to real-world data in the future might thus provide hints regarding design decisions for the website depending on the predicted visual complexity level, for example regarding the number of presented items or their structure. This approach, focussing on mean visual complexity ratings, does however not consider interindividual differences within the perception and judgement.

This variation may go back to several aspects, including preference and previous experience with certain types of websites, demographic aspects, personality traits or interestingness. While larger experience with a certain type of websites such as shopping websites may lead to generally lower ratings of visual complexity for this type, demographic and personality traits might also affect the perception of complexity. Chamorro-Premuzic, Burke, Hsu, and Swami (2010) for example showed that age, gender and personality factors such as openness and conscientiousness were related to preference for complex artworks. It might however be hypothesized that interindividual aspects might not only have an impact on the perception of artworks, but also on other stimuli such as website screenshots and in this regard add to the variance of visual complexity ratings. Additionally, interestingness is similarly related to visual complexity as discussed in paragraph 2.2.5.3 (e.g. Aitken, 1974; Berlyne & Boudewijns, 1971; Day, 1968). Particularly for naturalistic stimuli, interestingness may however not only be an attribute of the stimulus, but also relate to individual interests. For example, an online shop for plants is probably perceived as more interesting by an amateur gardener as compared to someone who is not interested in plants at all. It could thus be hypothesized that interindividual differences, for example with regard to the perceived interestingness of stimuli, might also affect their visual complexity ratings.

In order to account for these differences between individuals with regard to the complexity rating of specific website screenshots, prediction models for single visual complexity ratings were investigated similarly to study 3. Again, the three approaches of an ordinal mixed-model, random forest and glmmLasso were evaluated. Within these, the highest correct classification rate of 44% in the training and 45% in the test data and lowest MSE as well as MAE were found for the glmmLasso method with random effects for both subjects and stimuli. Within this, the ocular parameter scanpath length and the computational measures Vertical Symmetry, Quads 4x4, Quads 8x8, RMS SD, RMS Mean x SD, RMS JPEG Size, Nr. of Segments, Näsänen Complexity as well as Orientation, Feature Congestion map Filesize and Orientation map Filesize were included by the algorithm. The glmmLasso approach with a random intercept for subjects yielded similar accuracies as the ordinal mixed-model regression with the same random effect, while the subject-averaged ordinal mixed-model, the random forest and glmmLasso model without any random effect gave the lowest correct classification rates of between 33 and 36% and highest MSE and MAE values. These results suggest that the integration of additional information, particularly regarding differences between subjects but also between stimuli, can improve the accuracy of the prediction model as compared to when this information is not taken into account. All in all, accuracies for the rating data within this study are slightly lower than within study 3. This may reflect the larger number of potential influences on applied stimuli such as website screenshots, which can hardly be assessed by means of computational or ocular parameters. Particularly high MAE_M values and thus low accuracies were found for categories 4 and 7. For category 7, this may relate to the relatively smaller number of ratings on this level. For category 4 however, this appears rather surprising. While there are slightly less ratings within this category as compared to the adjacent categories, this alone is unlikely to explain why category 4 is comparatively rarely predicted by most measures as can be seen within the confusion matrices. A more plausible explanation might be that some of the ratings within this category reflect a central tendency bias and are thus less related to the according range of computational and ocular measures.

The latter played a slightly larger part compared to study 3 with scanpath length selected in the first glmmLasso model with two random effects and number of fixations selected within the third glmmLasso model without random effects, while various ocular

parameters showed large variable importance values within the random forest model. Still, due to the data structure, it might have been expected that these carried more relevance. However, similar to study 3, it is possible that the random intercepts for subjects in the first and second glmmLasso model already accounted for much of the explainable variation between different ratings of the same stimulus in combination of information about the stimulus from computational parameters.
Morevoer, due to the balanced experimental design that was implemented with the help of controlled stimuli and that was based on three basic templates of a shopping, company and news website, the overall variation both of ratings but also of computational and ocular measures was probably lower than when considering a range of different real website screenshots. With regard to the latter, the likely larger variation within predictors might positively affect prediction performance. Yet all in all, the presented methodology provides a good starting point for a transfer to uncontrolled, real-world stimuli such as screenshots from real websites or user interfaces.

Limitations

Of course, there are several limitations to this experiment. First of all, many further possible influencing variables such as the variety of elements, density of elements and visual balance might still influence visual complexity ratings and also affect ocular parameters. The implemented experimental design with four factors however was already rather complex. An integration of additional factors would both have complicated the creation of controlled stimuli even more and also significantly extended the length of the experiment due to a multiplication of the number of stimuli from 72 to at least 144. Moreover, the interpretability of main effects might have been reduced because of potential interactions among the factors. In sum, a selection of four supposedly relevant factors was considered within this experiment, in which the number of elements, symmetry and colourfulness represented the three visual complexity factors quantity, structure and colour identified within study 2. Additionally, the influence of prototypicality was investigated. In order to consider a broader range of possible factors, further experiments are necessary in which selections of factors could be systematically investigated. Thereby, a more fine-grained rating scale might be applied for the assessment of visual complexity ratings as discussed within the subsequent outlook-section.

Next to the selection of influencing variables, a broader range of ocular parameters could have been considered within the experimental analyses. In order to cover different aspects of gaze behaviour however, the three parameters number of fixations, scanpath length and spatial density were chosen. While all three are related to visual search (Goldberg & Kotval, 1999), the number of fixations might rather reflect the number of components or elements that the observer processes (Goldberg & Kotval, 1999), which is also in line with the identified effect of the number of elements. Scanpath length according to Goldberg and Kotval (1999) however rather reflects the efficiency of the visual search behaviour, which relates to saccade amplitudes. Findings from this study, specifically the shorter scanpaths for prototypical screenshots, are in line with Goldberg and Kotval's (1999) proposition of shorter scanpaths indicating more efficient search behaviour, while longer scanpaths for symmetrical screenshots may indicate more extensive visual exploration as discussed above. Finally, the third parameter spatial density was taken into account, which refers to the spatial distribution of fixations, indicating the spatial coverage of a stimulus or interface within visual search (Goldberg & Kotval, 1999). The three selected parameters thus cover different aspects of gaze behaviour, as also revealed within the results. Considering the existence of many further potentially relevant parameters however, a larger number was analysed for the prediction of visual complexity ratings. Nevertheless, the experimental analyses of all 44 parameters would have gone far beyond the scope of this work. Still, a detailed analysis of additional parameters might provide further insights into attentional aspects and the search behaviour of observers.

Furthermore, a limitation of the study`s experimental design consists in the strictly controlled stimuli. These allowed for precise analyses of effects of influencing variables on visual complexity ratings as well as on ocular parameters. However, this approach also restricts to a certain extent the variation of stimuli that would have appeared when using various screenshots from a number of real websites as naturalistic stimuli. This might have provided benefits regarding the prediction performance due to a possibly larger variation both in computational as well as in ocular parameters. Moreover, a broader range of stimuli might also have implicated the selection of both very simple and also very complex screenshots, which again might have improved the prediction accuracy particularly within the extreme ranges due to the larger number of very low and very high complexity ratings. Next to the controlling of the experimental design, the presentation time was also strictly controlled. This provided advantages concerning the comparability of ocular measures as well as visual complexity ratings by preventing possible confoundings due to different viewing durations. Within a naturalistic context however, different viewing durations may appear as a result of differences in visual complexity. Assessing viewing durations within subsequent studies might therefore also shed light with regard to the perception of visual complexity, either by comparing visual complexity judgements for systematically manipulated levels of viewing durations or within free-viewing tasks. The latter was already examined for example by Shigeto et al. (2011), who identified longer viewing durations for more complex stimuli. Finally, no performance measures were assessed within this study. While subjective ratings clearly suggested relations with mental workload, these could have been further underlined by the integration of performance data, as also shown for example in study 1. However, this would have required a different experimental task, most likely also in combination with a dynamic sequence or at least with variable viewing durations, in order to assess reaction times for example. Since it would have contradicted the strictly controlled investigation of influencing variables within this work as described before, it was not implemented within this work. Still, it remains an interesting issue that could be addressed within future research.

Outlook

Several aspects of this study offer potential for following investigations. These are discussed within the following.

First of all, the experimental investigations of the effect of colourfulness on visual complexity did not reveal clear results. While no significant effect of colourfulness on visual complexity ratings could be found, this might also be due to the relatively subtle differences between the two conditions. Within future research, differences between the levels of colourfulness should therefore be enlarged or a broader variation of colourfulness conditions should be considered in order to ensure that the stimuli are adequate for the identification of possible effects on the perception of visual complexity. Therefore, three or more levels of colourfulness ranging from low to high colourfulness could be used within the manipulation in order to allow for a detailed differentiation of possible effects. Additionally, by calibrating the monitor colours, their precise display could be ensured. This would also contribute to the reproducibility of findings.

Furthermore, next to colourfulness, other aspects such as the number of colours or colour contrast might also be taken into account. The analysis of relations between ratings of these variables and visual complexity showed inconclusive results within study 2. While both variables were significantly related to visual complexity ratings for photographs, this was not the case for website screenshots. However, factor scores of the colour dimension had a significant impact for both types of stimuli. Therefore, the influence of colour on the perception of visual complexity should be investigated more closely within subsequent research. This might encompass not only a manipulation of colourfulness but also of the number of colours or colour contrast, although it might be assumed that these are often interrelated particularly when naturalistic images or screenshots serve as stimuli.

Moreover, further quantitative and structural influencing variables of visual complexity might also be investigated within future studies by experimental means. This might particularly refer to aspects such as the variety of elements or density of elements. Although these are often related with each other in naturalistic stimuli, as suggested in study 2, independent manipulations might provide additional insights into their effects and possible interactions.

As already indicated in the paragraph limitations, further ocular parameters could additionally be considered within subsequent studies. This could provide additional insights into the viewing behaviour in relation to visual complexity and conclusions about the attentional processes involved.

Focussing on the relation with mental workload, performance measures as well as physiological indicators might be integrated into future studies in order to achieve a holistic image of effects on the mental workload state of users. Furthermore, associations of visual complexity with usability and user experience as well as aesthetical judgements and emotions could be addressed within future research. Thereby, the impact of visual complexity on cognitive and affective processes of the user could be assessed more comprehensively.

Focussing on the prediction of visual complexity ratings, subsequent research might capture the approach of combining both computational and ocular parameters and apply it to a larger number of real-world user interfaces and websites. Since this study underlined both the sensitivity of ocular parameters for manipulations of influencing variables as well as their applicability in combination with computational measures for the prediction of visual complexity ratings, the logical next step would be to aim for less controlled stimuli. Thereby, a large pool of stimuli encompassing a broad variation in different regards could easily be created and employed within research. This would also help to create and improve a general prediction model with a higher prediction accuracy also within extremely low and high visual complexity ranges. Thereby, a more fine-grained rating scale could be used for the assessment of the subjective perception of visual complexity. In contrast to the seven-point Likert scale used within this study, for example a slider bar with up to 100 discrete values might provide more detailed ratings and facilitate the analysis of data, which would not have to be considered as ordinally scaled. This would however probably come at the price of higher response times (Funke, Reips, & Thomas, 2011).

As a next step, visual complexity might also be investigated within dynamic scenarios. This might comprise video sequences as within study 1 for example. Here, it might be hypothesized that video complexity incorporates both the visual complexity of single video frames as well as properties of their temporal sequence. Moreover, the investigation of dynamic scenarios might be particularly interesting with regard to the interaction with user interfaces. This often involves several steps, as for example when interacting with websites. The interaction of users typically starts for example at the main page, from where they navigate through the website structure according to their goals (see for example Tan & Wei, 2006). The visual complexity of single pages visited throughout the navigation process may be hypothesized to contribute to the goal-achievement of the users, e.g. the finding of specific information or the completion of a transaction as well as mental workload or affective processes. Within the investigation, performance measures such as the required time for a navigation process or for finding certain information or products might be integrated. These could serve as additional indicators of mental workload next to subjective or physiological measures. Moreover, since situational awareness can also be related to performance (Endsley, 2019), the assessment of mental representations of user interfaces might similarly be of interest in order to draw inferences about potential effects of visual complexity. All in all, considering visual complexity within a dynamic sequence might be particularly interesting for future research, particularly regarding the design of user interfaces for example.

In conclusion, the results of this study stress the role of both quantitative and structural influencing variables on the subjective perception of visual complexity as well as the impact of prototypicality, which may relate to mental models built on the previous experience of subjects. Moreover, effects on ocular parameters such as scanpath length underline the suitability of their use within the investigation of visual complexity. Using both computational and ocular parameters, visual complexity ratings could be predicted with acceptable accuracy within regression models.

These findings and the clear relation of visual complexity with mental workload underlined the relevance of the construct within human-machine interaction.

The achieved results can be helpful both within research but also be adapted for the design of interfaces. Within the former, the investigation of influencing variables contributed to the theoretical understanding of the construct. Beyond that, the integration of eye tracking methodology, for example by including ocular parameters next to computational ones for the prediction of visual complexity ratings, opens new possibilities in research on visual complexity. With regard to the latter, the insights from this work might serve as points of reference for example for the design of user interfaces with regard to quantitative and structural aspects as well as the effects on users' mental workload.

7. Overall discussion

Within the subsequent overall discussion, the insights from the conducted studies are concluded and connected at first, before they are combined with findings from previous literature into a research model of visual complexity in human-machine interaction. Subsequently, both limitations and implications are pointed out. Finally, a summary and outlook of issues for further research is given.

7.1 Conclusion of findings

Four studies were conducted within the scope of this dissertation, focussing on the investigation of the role of visual complexity within human-machine interaction.

The first study focussed on the role of autocycling frequency and video complexity within a CCTV surveillance task in control rooms. Results showed that both factors affected the mental workload of operators not only by means of subjective ratings, but also with regard to performance and physiological measures. In the context of previous research works on mental workload in control rooms, this provided experimental support for considering complexity more closely within human-machine interaction. This had previously also been suggested by Pikaar et al. (2015), who described the potential influence of visual properties and complexity factors on the cognitive processes of the observer within the concept of *scenes*, which they defined as logical or meaningful sets of visual information.

Although video complexity is not the same as visual complexity (which mostly refers to static stimuli), it can still be assumed that visual complexity of single frames is an essential feature of video complexity and that similar influencing factors such as the number of elements affect both. All in all, the findings of complexity effects on mental workload underline the importance of the further investigation of the construct for the domain of human-machine interaction.

Therefore, study 2 firstly aimed at gaining a better theoretical understanding of the construct visual complexity and its influencing variables. By examining both basic IAPS photographs (Lang et al., 2008) as well as website screenshots, potential influencing variables could be structured into the three factors quantity, structure and colour. Together with their significant relations with visual complexity ratings, this provides the basis for a better theoretical understanding of the construct and supports its general validity within several domains.

Of the three factors identified within study 2, the influence of the two factors quantity and structure was then supported experimentally within study 3, which allows for causal inferences regarding their impact. Manipulations of these, applied on basic black and white shape patterns, revealed significant effects on both visual complexity ratings and various ocular parameters, which stresses the suitability of the latter for research on visual complexity. Moreover, good accuracies could be achieved by combining both ocular and computational measures for the prediction of visual complexity ratings.

A similar approach was used on applied stimuli in study 4. With website screenshots serving as stimuli, it was revealed that next to quantity and structure, prototypicality had an impact on visual complexity ratings as well as ocular parameters. Additionally, a clear relation between visual complexity and mental workload was identified on the base of subjective ratings. Again, the prediction performance particularly for mean visual complexity ratings was very good.

Within these four studies, the research gaps identified within the research agenda in paragraph 2.6 were thus addressed, as described within the following.

First of all, the gathered findings regarding the influencing variables and their factorial structure contribute to a better theoretical understanding and therefore help to formulate a generally acknowledged definition of visual complexity, which is based on the influence of quantitative, structural and colour-related aspects. The impact of influencing variables was also further stressed within experimental investigations, which underline the causality of the relations while building on findings from previous literature such as Chipman (1977) and Ichikawa (1985) and extending these to more naturalistic stimuli. This additionally contributed to the theoretical foundations of the construct and provides a reliable basis for further research.

Furthermore, the eye tracking methodology provided insights into the attentional aspects and gaze behaviour involved within the perception of visual complexity, while having been rarely used within research on visual complexity yet. Significant effects of visual complexity influencing variables on various ocular measures suggest that these are highly suitable for the investigation and identification of visual complexity effects and are therefore worth being incorporated within future research works.

Finally, single visual complexity ratings have been firstly predicted by combining both computational and ocular parameters. In contrast to previous research, the approach of integrating ocular parameters next to computational ones allowed both for the prediction of individual ratings while providing good accuracies for the prediction of mean complexity ratings as well, which extends previous findings that are based merely on computational measures.

7.2 Model of visual complexity in human-machine interaction

Based on the results gathered within the framework of this dissertation project as well as the findings and models from previous research literature (see for example paragraph 2.2.4), an integrative research model of visual complexity in human-machine interaction is developed. This considers the role of influencing factors of visual complexity as well as the effects of visual complexity and its relations with other constructs such as mental workload. Moreover, associations with ocular parameters as well as effect of previous experience are included. The model is presented in Figure 88. Within the model, latent variables are depicted as circles and manifest variables as squares as it is common within latent variable models. The relations within the model as represented by arrows should not necessarily be seen as strictly unidirectional. Very likely, this research model is far from complete. However, the intention of this model is to sum up the current state of research focussing on visual complexity and related constructs within the context of human-machine interaction to the best of my knowledge in order to encourage future research in this field as well as to draw attention towards the relevance of the construct.

The references linked within the depiction of the model are the following:

1: Ichikawa (1985)

2: Leder et al. (2004)

3: Torralba et al. (2006); Awh et al. (2012)

4: Itti, Koch, and Niebur (1998); Torralba et al. (2006)

5: Berlyne (1971); Berlyne et al. (1963); Tuch et al. (2011; 2012)

6: Zajonc (1968, 2001)

7: Nadkarni and Gupta (2007)

Subsequently, theoretical foundations and findings, which serve as the foundations for the model, are described in detail.

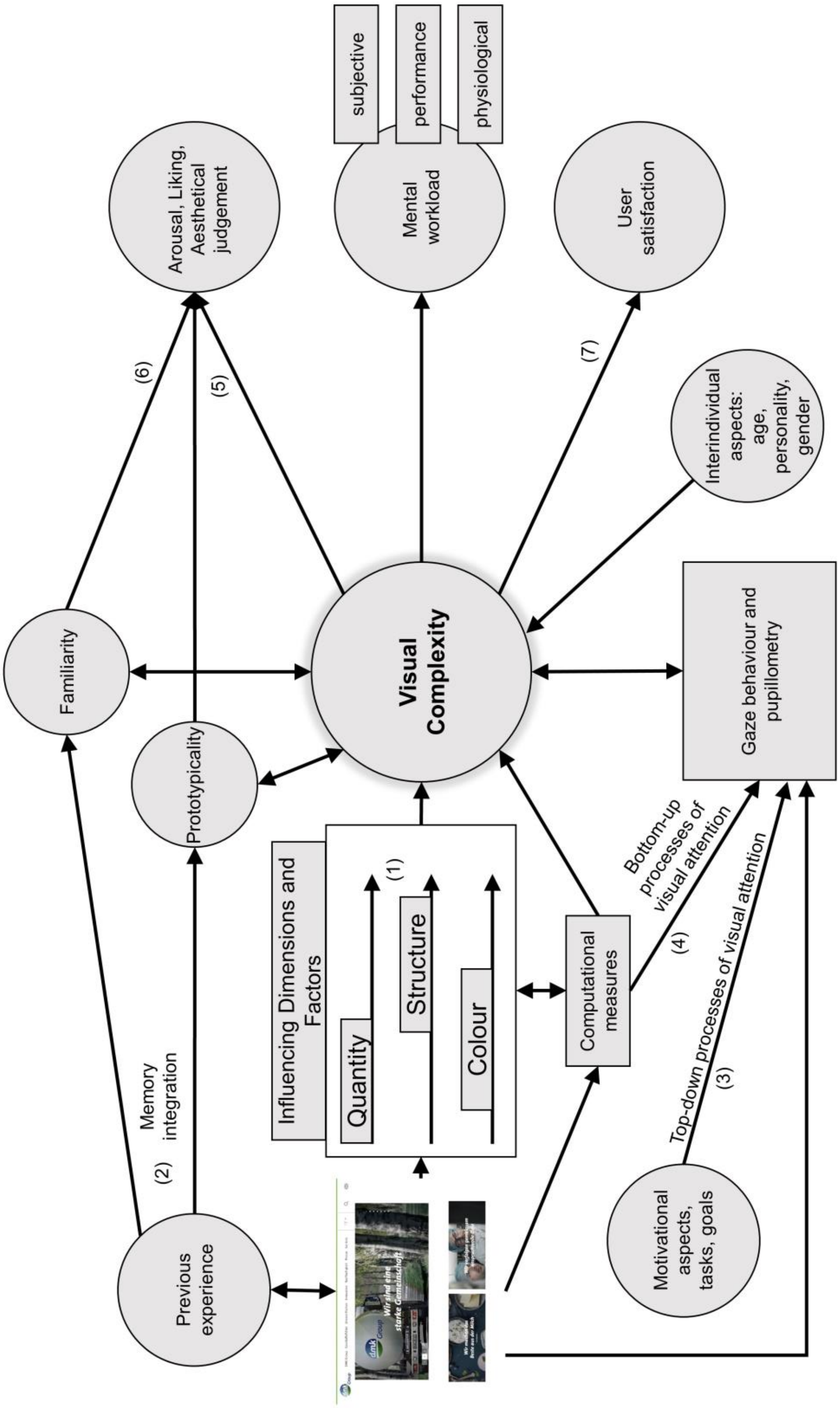


Figure 88. Integrative research model of visual complexity in human-machine interaction

With regard to the **influencing factors**, Ichikawa (1985) proposed a tentative model for the judgement of pattern complexity. This reflects his results regarding the two cognitive processes of visual complexity perception, with a fast process for the detection of quantitative features and a slower process for the detection of structural features, as depicted in Figure 89 and previously described in more detail in paragraph 2.2.3.

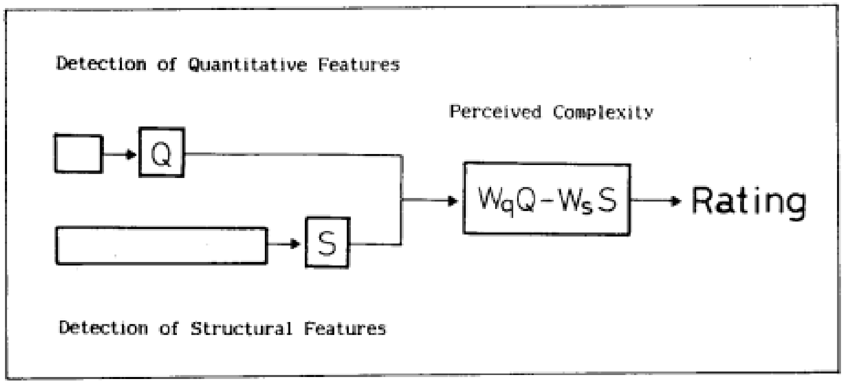


Figure 89. Ichikawa's (1985) tentative model for the judgement of pattern complexity

Results from studies 2, 3 and 4 support the role of both quantitative as well as structural properties for the perception of visual complexity. While study 2 showed that factor scores of the quantitative and the structural factor were significantly related to visual complexity ratings, study 3 and study 4 underlined their relevance based on experimental investigations. Moreover, study 2 suggests that a third, colour-related factor has an impact on visual complexity perception. For this however, unlike for the other two factors, no previous research literature exists concerning the timeframe of processing. Regarding the stimulus material, study 3 used basic shape patterns similar to Ichikawa's (1985) research, while photographs and screenshots of websites were used within study 2 and 4. Since the gathered results are thus based on different types of stimuli, the model not only applies to pattern complexity as Ichikawa's (1985) findings but refers to visual complexity in general.

The three identified factors of visual complexity each comprise a number of influencing factors. Study 2 assessed their dimensionality and found that the number of elements, the variety of elements and the density of elements contribute to the quantitative factor while the structural factor contains the influencing variables organization, symmetry and visual balance. Finally, the variety of colours and colour contrast are the constituents of the colour-related factor. This selection of influencing variables however might not be exhaustive, since other aspects that were not included within the analysis could additionally be related to visual complexity and its identified factorial structure. Yet, since the selection of potential influencing variables for the investigation was based on an extensive review of visual complexity literature, the most important aspects were most likely included. For each of the three identified factors of visual complexity, the most relevant influencing variable was then selected for subsequent experimental investigation. Among these, the number of elements and symmetry significantly affected visual complexity ratings within study 3. Additionally, colourfulness (which is assumed to combine both the variety of colours and colour contrast) was investigated within study 4, but did not show a significant effect on visual complexity ratings. This may be due to the relatively small differences in the manipulation.

In sum, the factors quantity, structure and colour and their related influencing variables affect the perception of visual complexity as shown within the results of studies 2, 3 and 4. Additionally, previous researchers such as Ichikawa (1985) had found support for the influence of quantity and structure and addressed the temporal sequence of their processing. Therefore, the factors and related influencing variables depict the first essential part of the integrative research model of visual complexity.

At the same time, **computational measures** can be calculated for stimulus images. Partly, these can also represent influencing variables of visual complexity, for example for quantitative aspects such as the number of elements or density of elements (for example when using segmentation methods), structural ones such as symmetry (Bauerly & Liu, 2008; Elawady et al., 2017) or colourfulness (Hasler & Suesstrunk, 2003; Yendrikhovskij et al., 1997). Other measures however, such as image compression measures, edge measures or decomposition measures are not directly related to specific factors. However, it can be assumed that these are particularly affected by quantitative aspects, since the number and density of objects most likely has a larger impact on these compared to structural or colour-related aspects. Previous research had revealed that subjective ratings of visual complexity are strongly related to computational measures (e.g. Gartus & Leder, 2017; Tuch et al., 2009). Within studies 3 and 4 of this dissertation project, computational measures similarly such as compressed file sizes, edge density and decomposition measures contributed to the prediction accuracy for visual complexity ratings. Therefore, these are included within an early stage of the research model, since they arise directly from a certain stimulus image while being related to both influencing factors as well as visual complexity.

Next to the previously described influencing factors, **previous experience** can also affect the perception of visual complexity. Within Leder et al.'s (2004) model of aesthetic experience, this is associated with implicit memory integration, which again includes the processing of familiarity and prototypicality (see Figure 4). As shown within study 4, prototypicality can also affect the perception and subjective judgement of visual complexity. The definition of prototypicality as “the amount to which an object is representative of a class of objects’’ (Leder et al., 2004, p. 496) implies that it relies on previous experience with different objects of a certain class, which is necessary in order to judge the degree to which a specific object is representative for this class. The role of previous experience also becomes evident within the ‘mere-expose’ effect (Zajonc, 1968, 2001). The mere-exposure paradigm relies on the repeated exposure of a subject to a certain stimulus. Research with this has shown that repeated experience with stimuli enhances positive affect towards the stimulus itself but also towards similar stimuli that were not exposed (Zajonc, 2001). However, as shown by Berlyne (1970), Saegert and Jellison (1970) as well as Smith and Dorfman (1975) for example, the effect of repeated exposure on liking may interact with the visual complexity of stimuli. Berlyne (1970) for instance showed that the liking of complex stimuli increased with exposure while the liking of simple stimuli tended to decrease. Cox and Cox (1988) therefore suggested that complexity might act as a moderator of repetition effects on liking. This relation however may also be bidirectional. Snodgrass and Vanderwart (1980) for example found a significant negative correlation between familiarity and visual complexity ratings within their picture set and discussed possible explanations for this finding. In sum, previous literature revealed that previous experience or exposure with stimuli can affect both the prototypicality of as well as the familiarity with these. This effect can go beyond previously observed stimuli and extend also to similar ones. With regard to familiarity, relations with visual complexity have been identified within previous research. For prototypicality, it was revealed within study 4 that it can affect the perception of visual complexity. Similar to familiarity, this relation is not necessarily unidirectional, with complexity levels also possibly influencing the rating of prototypicality. Previous research has moreover relatively clearly underlined the role of both familiarity and prototypicality with regard to aesthetical appraisal (e.g. Leder et al., 2004; Tuch et al., 2012). This relation may at least partly be moderated by visual complexity. The described relations between previous experience, prototypicality, familiarity, visual complexity and aesthetical appraisal are included into the research model in order to depict an overview and integrate research findings regarding the interrelations of the constructs.

Next to previous experience, **interindividual differences** such as age, gender or personality may similarly influence the perception of visual complexity. Within this regard, Chamorro-Premuzic et al. (2010) for example found that the personality dimensions openness to experience, extraversion and conscientiousness significantly correlated with the preference for classified complexity of visual art. Moreover, significant relations with age, sex, education and visits to museums were found. These findings go back to a study with a large sample of $N = 3254$. Earlier research from Eisenman (1967, 1968) similarly found effects of sex, birth-order [sic] and personality attributes on visual complexity preferences. With regard to websites, Reinecke and Gajos (2014) used a large dataset of ratings and found interaction effects between visual complexity and age, gender, country as well as education with regard to visual appeal. For example, the authors found that older persons and those with a lower education level preferred more complex websites. With regard to gender, females for example disliked simple websites more than males. Similarly, Wang, H.-F., Wang, P.-Y., Liao, C.-C., and Lin, Y.-Y. (2014) identified gender effects in children with regard to preference for visual complexity of websites within their study. Focussing on personality, Martin, Sherrard, and Wentzel (2005) found that the trait of sensation seeking influenced the preference for visual complexity. While high sensation seekers preferred complex visual designs, low sensation seekers preferred simple designs.

In conclusion, findings from previous research literature showed that interindividual aspects such as gender, age, and personality affect the *preference* for complexity. Relatively little research has yet directly addressed their influence on the *perception* of visual complexity, for example that personality affected the subjective ratings of visual complexity. Yet, it might be hypothesized that interindividual differences in preference for complexity partly go back to a different perception and judgement of visual complexity. Therefore, next to previous experience, interindividual aspects such as gender, age and personality are included into the research model as potential influencing factors of visual complexity. However, this association remains a preliminary hypothesis until further research has investigated the relation more closely.

Furthermore, as could be shown by studies 3 and 4 of this dissertation as well as within previous research by Madan et al. (2017) and Bradley et al. (2011), **gaze behaviour and pupillometry** can reflect the perception and processing of visual complexity. The resulting ocular parameters can relate to different aspects of visual attention, among these particularly bottom-up as well as top-down effects. As explicitly discussed in paragraph 2.5.1, representatives of bottom-up processing argue that features of the stimulus such as saliency determine the visual processing and thereby also the visual attention and eye movements (Itti et al., 1998; e.g. Itti & Koch, 2001) while top-down theories assume that visual attention can be directed via voluntary control, for example depending on motivational aspects such as specific goals or task demands (e.g. Folk et al., 1992; Posner, 1980). In this regard, Henderson (2003) differentiates between different mechanisms of the visual-cognitive system that control gaze in order to fixate informative and important image regions. First of all, he focusses on stimulus-based gaze control, which includes image features such as spatial frequency, edge density, contrast and saliency for example. Next to that, knowledge-driven aspects are relevant for gaze control according to him. This draws on short- and long-term memories regarding visual, spatial and semantic information about the observed scene as well as previously observed similar scenes next to goals and plans of the observer. This knowledge-driven control of gaze is of particular relevance for the observation of meaningful scenes during the execution of tasks as compared to visual saliency (e.g. Henderson, Weeks, & Hollingworth, 1999; Land & Hayhoe, 2001; Turano, Geruschat, & Baker, 2003). Among the knowledge-driven aspects of gaze control fall episodic scene knowledge (“information about a specific scene that can be learned over the short term in the current perceptual encounter”, Henderson, 2003, p. 500) as well as scene-schema knowledge (“generic semantic and spatial knowledge about a particular type of scene”, Henderson, 2003, p. 501) and task-related knowledge. In sum, these stress the role of knowledge for gaze behaviour in natural scenes, among which user interface or website screenshots may also be counted.

The contextual guidance model (Torralba et al., 2006) integrates both local salience features as well as global context features and thus combines bottom-up saliency, scene context and top-down mechanisms at an early stage of visual processing. This is assumed to rely on both a global and a local pathway. These two are depicted in Figure 90 and together create a scene-modulated saliency map, which in contrast to a pure bottom-up saliency map integrates global image features as well as top-down influences such as the specific search task and thus can help to identify image regions for visual exploration.

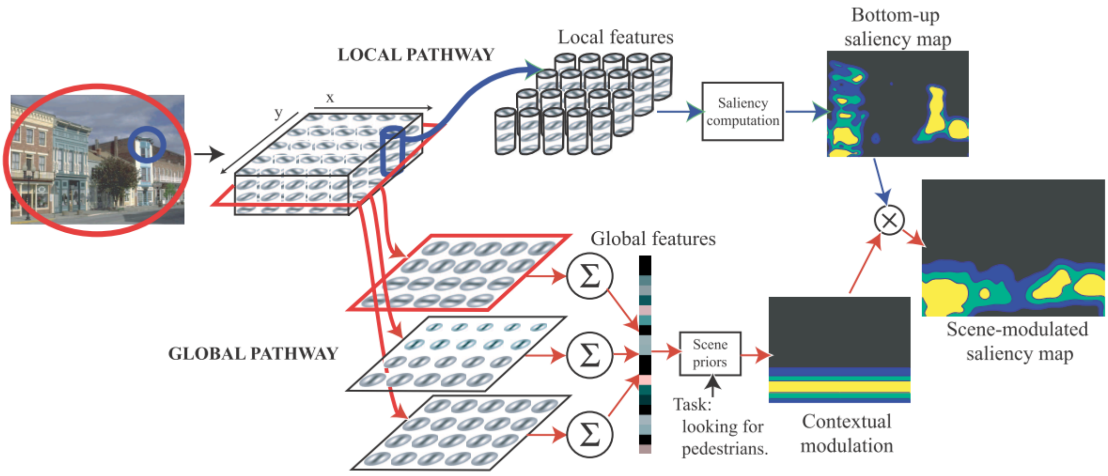


Figure 90. Contextual guidance model by Torralba et al. (2006)

Within the experimental investigation of Torralba et al.'s (2006) model using eye tracking methodology, the authors found that their contextual guidance model predicted the first five fixations significantly better compared to the saliency-only model (73% vs. 58% accuracy). This finding particularly underlines the role of top-down influences and global image features for gaze behaviour. Similarly, Awh et al. (2012) suggest to include a third aspect next to bottom-up (such as saliency) and top-down influences (such as current goals), which they call selection history. According to the authors, this represents past experience with previously attended items in a certain context as well as reward history (for example when the visual selection of certain items in associated with previous reward), but can also encompass other effects of past experience.

In conclusion, with regard to the processing and perception of visual complexity in human-machine interaction, different types of processes can have an impact, particularly when interacting with user interfaces. Next to bottom-up features such as saliency and top-down effects of specific goals, previous experience with regard to selection history or knowledge based on the context of the scene for example can have an impact on gaze behaviour. This should particularly carry relevance for user interfaces, since their design may often define global image features while users in many cases have already made previous experiences with similar interfaces. In this regard, the interrelation between theories of visual attention and visual complexity appears as particularly relevant, since the visual complexity as an image feature may play a role within visual attention theories. The integration of insights from the latter may in turn provide benefits in research on visual complexity and particularly with regard to gaze behaviour and ocular parameters. Therefore, the links between both are integrated into the research model in order to raise attention to the connections, which again may provide clues and stress the need for future research in this field.

Effects of visual complexity

As shown before, visual complexity is affected by a number of factors and is related to other constructs and measures. It can however also influence numerous other constructs. First of all, findings from previous literature suggest that visual complexity is positively related to the **arousal** of subjects. As discussed in detail within paragraph 2.2.5.1, Berlyne et al. (1963) for example provided support for an effect of complexity on arousal and later introduced the term “arousal potential” (Berlyne, 1971). According to him, this is associated with collative properties of the stimulus such as complexity as well as novelty and surprisingness next to psychophysiological (e.g. intensity) and ecological properties (e.g. gratification or discomfort). Later research, such as the findings of Marin and Leder (2013), supported the positive association between complexity and arousal for example for IAPS pictures (Lang et al., 2008) as well as for paintings. With regard to human-machine interaction, Deng and Poole (2010) similarly found a positive effect of webpage visual complexity on arousal ratings (see also their research model of webpage visual complexity depicted in Figure 5). Similarly, Tuch et al. (2011; 2009) for example found effects of visual complexity and arousal ratings as well as physiological measures. On the other hand, according to Madan et al. (2017), arousing stimuli may also be rated as more complex. But in sum, a broader literature base exists for the former direction of the relation.

Next to arousal, visual complexity can affect the impression of **pleasantness** or **(aesthetical) preference**. This again goes back to Berlyne (1971, 1974), as explicitly discussed in paragraph 2.2.5.2. He suggested that medium levels of visual complexity are generally preferred, since these go along with an optimal amount of arousal. While the suggested inverted u-shape has been supported by a number of findings (e.g. Farley & Weinstock, 1980; Imamoglu, 2000; Saklofske, 1975; Vitz, 1966) also within the field of human-computer interaction (Chassy et al., 2015; Geissler et al., 2006; Güçlütürk et al., 2016), numerous other studies rather suggested a linear relation instead (for a detailed discussion see paragraph 2.2.5.2). This has also been shown for websites for example by Tuch et al. (2011; 2012). Despite the contradictory results regarding the shape, the general relation between both constructs has been supported multiple times.

Next to arousal and pleasantness, Nadkarni and Gupta (2007) showed that visual complexity can affect **user satisfaction**. Within their theoretical model of perceived website complexity (see Figure 6 and detailed description within paragraph 2.2.4), they conclude their empirical findings that objective website complexity affects perceived website complexity, while this relation is moderated by user familiarity. Moreover, they found that the relation between perceived website complexity and user satisfaction is moderated by task goals. This finding particularly underlines the role of complexity within human-computer interaction.

Finally, studies 1 and 4 of this dissertation underlined the relations between visual complexity and **mental workload**. While findings in both studies are based on subjective ratings of mental workload, study 1 additionally presented results from performance measures such as reaction time and hits as well as physiological measures, for example ECG.

All in all, the effects of visual complexity on arousal, pleasure, user satisfaction and mental workload particularly underline the relevance of considering visual complexity within the domain of human-machine interaction. Since these aspects can play a role for the human-centered design of human-machine interfaces, they are integrated within the previously presented research model of visual complexity in human-machine interaction (see Figure 88).

7.3 Limitations and outlook

Within the range of this dissertation, of course a number of aspects could not be addressed in detail. First of all, of course not all potential influencing variables of visual complexity could be exhaustively investigated within the conducted experiments for

example. While a selection of the presumed most relevant variables of the visual complexity factors determined in study 2 were experimentally investigated within study 3 and 4, variables such as the variety of elements were not considered more closely. In order to achieve a fully comprehensive picture, further potential influencing variables of visual complexity such as variety and density of elements and particularly their interactions could be experimentally addressed within future research. Although this may lead to more complex experimental designs, a combined investigation of several influencing variables might provide insights regarding their interdependence. Additionally, while subjective ratings of visual complexity showed linear main effects across the presented studies, some ocular parameters suggested various interactions between these as for example concerning the number of elements and symmetry. These findings further underline the possible benefits of the combined consideration of multiple influencing variables which might, as for example suggested by findings from eye tracking, allow for conclusions regarding the cognitive and attentional processes involved in the processing of stimuli of different visual complexity.
Moreover, the effects of interindividual aspects such as personality, gender, age or previous experience could not be investigated in detail. As discussed within paragraph 7.2, previous research literature found that for example gender, age and personality factors can influence the preference for different levels of complexity in visual art (e.g. Chamorro-Premuzic et al., 2010). While this effect might not only refer to art, but also be found in other stimuli, it could also be hypothesized that it is partly based on differences in the perception and individual judgement of visual complexity. Therefore, the assessment and investigation of the influence of different interindividual aspects such as age, gender and particularly personality should similarly be incorporated into subsequent research. This might allow for insights into interindividual differences in the cognitive processing of visual complexity. By incorporating different gaze parameters within the investigation, potential findings could further be underlined.
Apart from that, previous experience as related to familiarity or expertise with a certain type of user interface for example can also influence the processing of visual complexity as discussed in paragraph 7.2. While the influence of previous experience was controlled within the conducted experimental studies by using only previously unknown stimuli, the systematic investigation of familiarity or experience effects remains to be

addressed within subsequent studies. Within the domain of human-machine interaction, this might particularly be relevant for conclusions regarding the training of novices or differences between experts and novices with regard to the perception and interaction with user interfaces. For the investigation, experienced experts within a certain domain, who are accustomed to using a technical system and user interfaces could for example be compared with novices concerning the perception and judgement of visual complexity. In this regard, the analysis of gathered gaze data might also allow for insights into differences in gaze strategies between both groups. Within previous research, effects of expertise on gaze patterns have for example been found within surgery (e.g. Law, Atkins, Kirkpatrick, & Lomax, 2004), biology (Jarodzka, Scheiter, Gerjets, & van Gog, 2010), programming (Bednarik, 2012) and even collaborative tasks (Liu et al., 2009). Moreover, according to a meta-analysis of 65 references on the influence of expertise on the comprehension of visualizations by Gegenfurtner, Lehtinen, and Säljö (2011), experts compared to non-experts showed shorter fixation durations, longer saccades, more fixations on task-relevant areas and fewer on task-redundant areas. The authors discuss that this may relate to a superiority in parafoveal processing and the allocation of selective attention. These findings suggest that the incorporation of expertise, which is based on previous experience, can have a significant impact on the gaze behaviour of subjects. The integrated investigation of expertise and visual complexity within future research might thus help to find out how expertise facilitates the perceptual coping with different (and particularly higher) levels of visual complexity. Next to expertise, effects of the familiarity of stimuli could be addressed within experimental investigations by presenting previously unknown stimuli more or less often to subjects before assessing their judgements of visual complexity and gaze behaviour. This would allow for conclusions regarding the effects of mere exposure of stimuli on the perception of visual complexity while not requiring subjects to have previous experience with stimuli, which would allow the independent and highly controlled assessment of familiarity effects.

Furthermore, a limited number of stimulus domains was considered within the investigation of visual complexity in this dissertation project. Among these were video sequences in study 1, photographs and website screenshots in study 2, black and white shape patterns in study 3 and again website screenshots within study 4. Since results appear consistent between the studies, the conclusions previously drawn from these seem relatively reliable. However, this does not necessarily mean that these can be directly transferred to any type of visual stimulus like drawings or other kinds of visual art for example, which have not been investigated within the framework of the presented studies but were taken into account within previous research works. This also applies to other types of user interfaces such as car displays or interfaces of other machines or electronic devices. Although their design often appears rather reduced compared to websites for example, differences in visual complexity can have effects on the user or driver as well, as for example shown by Yoon et al. (2015). They found that visual complexity aspects such as the quantity of components also affected the visual search performance in an automotive instrument cluster. In this regard, future research might investigate if the dimensional framework identified within this dissertation is also valid for other types of user interfaces and if the prediction of complexity ratings from computational and ocular measures provides similarly good results. This also holds valid for dynamic sequences such as the navigation between multiple subpages of a user interface or the observation of videos within control rooms. Referring to the relevance of findings for real life, Tuch et al. (2009) stressed that “as the name implies, in real HCI situations (such as browsing the web) it is actual interaction, rather than passive viewing, which is most influential in shaping the overall user experience” (p. 713). Even though the main focus of the presented studies was on static images, this can still carry relevance with regard to real life. First of all, as suggested by Tuch et al. (2012), first impressions are formed very quickly during visual inspection, for example of websites, which may determine of the user decides to stay on the site or continues browsing to other sites. In this regard, the visual complexity of static website screenshots itself may play an important role. Moreover, findings from investigations with static pictures may also provide hints with regard to dynamic contexts. For example, variables such as the number of elements may have an impact for both static and dynamic visual complexity. Based on this assumption, the impact of video complexity on mental workload was for example assessed within study 1, using the number of different persons in the video as an indicator of complexity. However, the role of certain influencing factors of dynamic complexity was not addressed in detail, neither within study 1, nor within the framework of my dissertation. The issue of defining video complexity, which might be particularly helpful for example in the context of workplace design within CCTV control rooms therefore remains to be addressed within future research works. Similarly, dynamic sequences for example of multiple pages of an interface that are visited within the process of achieving a certain goal, for example when a website or software is searched for specific information, has (to the best of my knowledge) not yet been addressed in previous research either and could be a goal of future studies, also discussed in more detail within study 4. All in all, the perception of visual complexity in dynamic real-world scenarios remains an interesting aspect for future research on visual complexity.

Regarding the impact of visual complexity in real life, particularly concerning the design of human-machine interfaces for example, its effects on the perception of user experience and usability could be further explored in future studies. Usability is a core concept within human-computer interaction (Hornbæk, 2006) and describes the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (Deutsches Institut für Normung e.V., 2018, p. 8). Moreover, user experience has increasingly gained interest over the last years (Hassenzahl & Tractinsky, 2006; Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009). It is defined as the “user’s perceptions and responses that result from the use and/or anticipated use of a system, product or service“ (Deutsches Institut für Normung e.V., 2018, p. 11). While the association of visual complexity with these constructs has not yet been directly addressed, Nadkarni and Gupta's (2007) findings that website complexity has an effect on user satisfaction might therefore serve as a good starting point for further investigations. Within these, quantitative assessments may for example incorporate Brooke's (1996) system usability scale or the user experience questionnaire by Laugwitz, Held, and Schrepp (2008). The gathered insights may help to draw conclusions regarding the role of visual complexity with regard to the human-centered design of technical systems and eventually further underline the relevance of the concept beyond the identified effects on mental workload for example. The notion of considering visual complexity as a measure of usability was also taken up by Stickel, Ebner, and Holzinger (2010), however the authors did not address the relation explicitly, for example by assessing effects on usability.

Another aspect that could not be addressed within the framework of the conducted studies is the integration of neurophysiological methodology for the investigation of visual complexity. Next to gaze data, these can provide deeper insights into the neurocognitive foundations for the processing of visual complexity. Previous studies laid a base by using electroencephalography (EEG), identifying visual complexity effects both from event-related potentials (ERPs, e.g. Barkaszi, Czigler, & Balázs, 2013; Bradley et al., 2007) but also within time-frequency analyses. Findings of Bradley et al. (2007) for example suggest ERP differences in occipital and frontal areas within a time window between 150 and 250 milliseconds. Literature using time-frequency analyses showed complexity effects particularly in alpha and beta frequencies (Bruce, Delafield, Bonnie, Winwood, & Gale, 1972; Gale, Coles, & Boyd, 1971). Within a more detailed investigation, Gale, Spratt, Christie, and Smallbone (1975) found that the number of elements as well as their variety had effects on alpha and beta frequencies, while the number of elements affected theta frequency. Next to EEG, additional insights can be provided using functional magnetic resonance imaging (fMRI). A first step in this direction was made wihtin a study by Schlochtermeier et al. (2013), who addressed the role of visual complexity within emotional picture and word processing by using fMRI.

While all of these findings stem from basic research, studies from applied contexts such as human-machine interaction have not used neurophysiological methodology for research on visual complexity yet. Based on these previous results however, both EEG and fMRI methodology might provide additional insights into the neural processing of visual complexity in human-machine interaction, since these allow for both the localisation of neural activation within functional areas of the brain, but also give insights into the precise temporal sequence of its effects. Combined with an experimental approach such as the one used by Ichikawa (1985), who investigated the role of complexity aspects with different presentation durations, a detailed understanding of the processing of influencing variables of visual complexity as well as their temporal sequence could be achieved.

While the proposed research model of visual complexity in human-machine interaction integrates findings of the four conducted studies with previous literature and thus also provides a framework of relevant relations between visual complexity and other constructs, most of the corresponding results are based on correlations. This means that no definitive statements can be made regarding the directionality and causality of relations between constructs, such as visual complexity and arousal for example. Next to the use of advanced experimental study designs, at least as far these are viable, structural equation modelling might be an appropriate tool for the further examination and statistical analysis of the relations between latent variables. This approach might help to further validate the interrelations of visual complexity beyond the existing evidence and thus help to underline the relevance of the model.

Furthermore, there is a certain conceptual ambiguity between visual complexity and other constructs or terms. While visual complexity is often defined as “the level of detail or intricacy contained within an image” (Forsythe, 2009, p. 158; Snodgrass & Vanderwart, 1980, p. 183), the term (visual) perceptual load was for example used by Macdonald and Lavie (2011) to describe “the amount of information involved in the perceptual processing of the task stimuli” (p. 1780). The term load was also by other authors with a similar meaning. For example, Alvarez and Cavanagh (2004) referred to the visual information load of objects, which they found to have an influence on the capacity of short-term memory next to the number of objects. Moreover, Engström et al. (2005) for example used the term visual load when referring to a visually demanding secondary task within a driving simulator study. Similar to mental workload, the concept of visual load as suggested by Macdonald and Lavie (2011) thus relates to an available amount of (attentional) capacities that are consumed by a certain task and leave less resources available for task-irrelevant information (Lavie, 1995; Lavie & Tsal, 1994). This also determines if attentional selection happens at an early or late stage according to the authors. In contrast, the stress and strain model by Rohmert (1984), which is well established within the field of human factors, would however suggest a separation of the stresses coming from the outside (such as visual complexity for example) and the strains resulting from these within the individual (for example mental workload). Compared to this, Macdonald and Lavie's (2011) concept thus rather suggests an interplay of both levels comprised by the concept of visual or perceptual load, which also influences the locus of selective attention. This would however contradict a unidirectional process as for example implicated by Rohmert (1984). Consequently, the concept visual load offers the benefit of considering capacity limits (for example of the short-term memory), which can also affect the perception of visual complexity as for example reflected by the gaze behaviour (see for example the discussion of the results on ocular parameters for study 3 in paragraph 5.5). Therefore, visual complexity might not only influence mental workload unidirectionally, but in its perception and processing also depend on the available mental resources, such as the capacity of the short-term memory as shown by Alvarez and Cavanagh (2004). Within further investigations, this interplay would be worth to be examined in more detail. This might help to gain further insights into the cognitive processes involved in the perception of differently complex stimuli. By that, not only the relation between visual complexity and mental workload could be more closely addressed, but also a clearer distinction and definition of the concepts visual complexity and visual load might be facilitated. A first approach for addressing the described issue might consist in using a dual-task paradigm with both tasks occupying the visual modality as related to Wickens' (1984) multiple resource model. By increasing mental workload level in this modality, its influence on the perception and judgement of visual complexity could be investigated. A starting point for the closer investigation of the interplay between visual complexity and mental workload was done within an experiment (Ries, Wolf, Olschowski, Döllken, & Deml, 2018), which however used an auditory manipulation of mental workload, while effects on ocular parameters were examined. This idea could also be further pursued by using a visual secondary task and a more advanced manipulation of visual complexity, as for example through the variation of influencing variables as implemented within study 3 or 4 of this dissertation. Within the more detailed investigation of the role of mental resources and the interrelation between visual complexity and mental workload, the analysis of gaze behaviour might provide a better understanding of the cognitive processes involved. The partly opposing effects for ocular parameters within studies 3 and 4, which may arise from different demands for limited processing resources, may offer first insights in this respect. However, further research is needed to examine these post-hoc explanations in detail.

In conclusion, the four conducted studies enrich the existing body of research on visual complexity in human-machine interaction by addressing some current issues. Yet, of course not all open questions could be addressed in detail within the scope of this dissertation. The last paragraph therefore pointed out the limitations of the research works as well as starting points for subsequent investigations.

7.4 Conclusion and implications

Within the scope of this dissertation, four studies were conducted in order to address the role of visual complexity within human-machine interaction from three different perspectives. First of all, at a theoretical level, visual complexity was further scrutinized by examining the impact of different influencing variables as well as the factorial structure of the construct. In the second place, at the level of measurement, the relations with computational and ocular parameters were investigated, which may contribute to the assessment of a comprehensive image of visual complexity. Finally, at the level of impacts, effects on mental workload were shown both for the visual complexity of website screenshots but also for video complexity in a CCTV surveillance task from the context of a control room. Next to mental workload, previous research additionally showed associations between visual complexity and visual aesthetics, user satisfaction, performance and arousal (e.g. Marin & Leder, 2013; Nadkarni & Gupta, 2007; Tuch et al., 2009). Moreover, visual complexity could be relevant with regard to the user experience and usability of systems as discussed within the previous paragraph. All in all, the examined and hypothesized associations with the mentioned constructs particularly underline the relevance of visual complexity within the field human-machine interaction.

Beyond the domains addressed within this work, visual complexity can also have an impact on many further activities such as driving, both with regard to the road environment as well as the instrument cluster (Edquist et al., 2012; Yoon et al., 2015). Moreover, it might even affect ‘traditional’ workplace design in manufacturing for example where the visual demands of the work task (such as the difficulty in visual differentiation, which according to Nadal et al. (2010) also reflects a facet of visual complexity) can have an impact for example regarding the recommended working height (Deutsches Institut für Normung e.V., 2009). Next to the established effects of visual complexity in different application areas, the acquired findings on theoretical foundations may also have practical implications. In this regard, the impact of influencing variables as well as the factorial structure of the construct visual complexity was investigated within this dissertation. This can contribute to a better understanding and a universal definition of the construct, since both aspects allow for a better grasp of what visual complexity is and which factors contribute to its perception. Considering the differences in definitions (see paragraph 2.2.1) and proposed influencing variables and factors (see paragraph 2.2.3), these findings appear valuable both as a base for further research but also as points of reference for practitioners. Knowledge about the impact of influencing variables on the subjective impression of visual complexity may for example help to reduce or optimize the complexity level of user interfaces for example.

Furthermore, the findings on effects of visual complexity on the subjects' gaze behaviour demonstrated that the consideration of ocular parameters within research on visual complexity is very promising, since these can shed light upon the underlying attentional processes involved in the visual perception of complexity. On this basis, they can also serve as indicators of visual complexity and thus be used for the prediction of visual complexity ratings in addition to computational measures. The new approach of integrating both types of measures within this work provided good prediction accuracies for mean visual complexity ratings while also allowing to also consider interindividual differences and address single ratings instead of mean values. The achieved results are remarkable, particularly given the rather small sample size and number of stimuli. This might make the proposed prediction approach also interesting for use in the field of user research for example. Particularly when combined with interindividual attributes such as age, gender and personality as well as the level of previous experience, the use of eye tracking methodology may provide new insights into differences in the processing of visual complexity. The integration of interindividual aspects might also help to further improve prediction models of visual complexity ratings by explaining additional shares of variance.

Finally, the findings gathered within the scope of this dissertation were integrated with previous research into a research model of visual complexity in human-machine interaction. This illustrates the relations between visual complexity influencing variables, the construct itself, but also other aspects that are affected by visual complexity. Moreover, backgrounds of visual processing and attention are taken into account, providing relevant links with findings from basic research and underlying cognitive processes. The research model not only summarizes the state of research up to this date, but may also encourage the generation of hypotheses for future studies in this field.

In conclusion, next to other types of complexity such as cognitive, display and task complexity (Endsley & Jones, 2012), visual complexity can play a central role within different human-machine interfaces, such as control rooms, graphical user interfaces and driving. Thereby, it not only affects people during work, but within many scenarios of daily life. As suggested within this work and previous research literature, higher levels of visual complexity can lead to an increase in mental workload and have effects on other measures such as performance (e.g. Svensson et al., 1997) and liking (e.g. Berlyne, 1971). While the degree of tolerance towards visual complexity may differ between subjects, for example based on individual properties such as expertise, age or gender, there are also possible strategies for dealing with the visual complexity of user interfaces. First of all, findings regarding the impact of influencing variables on the perception of visual complexity can help to establish guidelines for reducing the visual complexity, for example by decreasing the number of elements, establishing structure or by grouping or "chunking" elements together. On the other hand, Donald Norman (2016) within his book "Living with Complexity" provides examples for cases, where a high level of (also visual) complexity is appropriate, such as the flight deck of a Boeing 787 (see Figure 91) and consequently suggests that complexity itself is neither good nor bad.

Figure 91. Flight deck of a Boeing 787, taken from Norman (2016) as an example of appropriate complexity

Instead, he argues that “Just as the owner of a cluttered desk sees order in its structure, we will see order and reason in complexity once we come to understand the underlying principles. But when that complexity is random and arbitrary, then we have reason to be annoyed.” (Norman, 2016, p. 4). According to him, “good design can help to tame the complexity, not by making things less complex – for the complexity is required – but by managing the complexity (Norman, 2016, p. 5). This aspect – good design – is an important aspect, to which the discipline of human factors can strongly contribute in order to make systems and products useful and usable.

8. References

Airgas (2020). Airgas | Home. Retrieved from https://www.airgas.com/*, last accessed 24.4.2020.

Aitken, P. P. (1974). Judgments of pleasingness and interestingness as functions of visual complexity. *Journal of Experimental Psychology*, *103*(2), 240–244. https://doi.org/10.1037/h0036787

Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531–552. https://doi.org/10.3758/BF03200732

Alemerien, K., & Magel, K. (2014). GUIEvaluator: A Metric-tool for Evaluating the Complexity of Graphical User Interfaces. In M. Reformat (Ed.), *The 26th International Conference on Software Engineering and Knowledge Engineering, Hyatt Regency, Vancouver, BC, Canada, July 1-3, 2013* (pp. 13–18). Knowledge Systems Institute Graduate School.

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010

Allsop, J., Gray, R., Bülthoff, H., & Chuang, L. (2017). Eye movement planning on Single-Sensor-Single-Indicator displays is vulnerable to user anxiety and cognitive load. *Journal of Eye Movement Research*, *10*(5), 1–15. https://doi.org/10.16910/JEMR.10.5.8

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103*(1), 62–70. https://doi.org/10.1037/h0036799

Aquino Shluzas, L. M., & Leifer, L. J. (2014). The insight-value-perception (iVP) model for user-centered design. *Technovation*, *34*(11), 649–662. https://doi.org/10.1016/j.technovation.2012.08.002

Arnheim, R. (1954). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA, Los Angeles, CA: University of California Press.

Arnoult, M. D. (1960). Prediction of Perceptual Responses from Structural Characteristics of the Stimulus. *Perceptual and Motor Skills*, *11*(3), 261–268. https://doi.org/10.2466/pms.1960.11.3.261

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). ATC complexity and controller workload: Trying to bridge the gap. In *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics: 23-25 October 2002, Cambridge, Massachusetts* (pp. 56–60). Menlo Park, California: The AAAI Press.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193. https://doi.org/10.1037/h0054663

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology*, *53*(4), 221–227. https://doi.org/10.1037/h0043921

AVSS (2007). i-Lids dataset for AVSS 2007: Advanced Video and Signal based Surveillance. Retrieved from http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html*, last accessed 30.12.2020.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Axure RP (Version 8) [Computer software] (2018): Axure Software Solutions, Inc. Retrieved from https://www.axure.com/*, last accessed 08.04.2019.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Evaluation Measures for Ordinal Regression. In *Ninth International Conference on Intelligent Systems Design and Applications: Isda 2009; Nov. 30, 2009 - Dec. 2, 2009, Pisa, Italy;* (pp. 283–287). Piscataway, NJ: IEEE. https://doi.org/10.1109/ISDA.2009.230

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*, 1–13. https://doi.org/10.3389/fpsyg.2017.00456

Barkaszi, I., Czigler, I., & Balázs, L. (2013). Stimulus complexity effects on the event-related potentials to task-irrelevant stimuli. *Biological Psychology*, *94*(1), 82–89. https://doi.org/10.1016/j.biopsycho.2013.05.007

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

BAuA (2019). Moderne Bildschirmarbeit. Retrieved from https://www.baua.de/DE/Themen/Arbeitsgestaltung-im-Betrieb/Neue-Technologien-und-Arbeitsformen/Moderne-Bildschirmarbeit/_functions/BereichsPublikationssuche_Formular.html?queryResultId=null&pageNo=0*, last accessed 03.04.2019.

Bauerly, M., & Liu, Y. (2008). Effects of Symmetry and Number of Compositional Elements on Interface and Design Aesthetics. *International Journal of Human-Computer Interaction*, *24*(3), 275–287. https://doi.org/10.1080/10447310801920508

Becker, K. (2016). *Becker Meditec Varioport Gebrauchsanweisung*. Karlsruhe.

Bednarik, R. (2012). Expertise-dependent visual attention strategies develop over time during debugging with multiple code representations. *International Journal of Human-Computer Studies*, *70*(2), 143–155. https://doi.org/10.1016/j.ijhcs.2011.09.003

Bennett, C. H. (1990). How to Define Complexity in Physics, and Why. In W. H. Zurek (Ed.), *Complexity, Entropy and the Physics of Information* (pp. 137–148). Redwood City California: Addison-Wesley.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279–286. https://doi.org/10.3758/BF03212593

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York, NY: Appleton-Century-Crofts.

Berlyne, D. E. (1974). *Studies in the new experimental aesthetics : Steps toward an objective psychology of aesthetic appreciation*. Washington, D.C.: Hemisphere Publishing Corp.

Berlyne, D. E., & Boudewijns, W. J. (1971). Hedonic effects of uniformity in variety. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *25*(3), 195–206. https://doi.org/10.1037/h0082381

Berlyne, D. E., Craw, M. A., Salapatek, P. H., & Lewis, J. L. (1963). Novelty, complexity, incongruity, extrinsic motivation, and the GSR. *Journal of Experimental Psychology*, *66*(6), 560–567. https://doi.org/10.1037/h0045767

Berlyne, D. E., Ogilvie, J. C., & Parham, L. C. (1968). The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *22*(5), 376–387. https://doi.org/10.1037/h0082777

Bertamini, M., & Makin, A. (2014). Brain Activity in Response to Visual Symmetry. *Symmetry*, *6*(4), 975–996. https://doi.org/10.3390/sym6040975

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*. (1939-1471(Electronic),0033-295X(Print)), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und erweiterte Auflage). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12770-0

Bradley, M. M., Hamby, S., Löw, A., & Lang, P. J. (2007). Brain potentials in perception: Picture complexity and emotional arousal. *Psychophysiology*, *44*(3), 364–373. https://doi.org/10.1111/j.1469-8986.2007.00520.x

Bradley, M. M., Houbova, P., Miccoli, L., Costa, V. D., & Lang, P. J. (2011). Scan patterns when viewing natural scenes: Emotion, complexity, and repetition. *Psychophysiology*, *48*(11), 1544–1553. https://doi.org/10.1111/j.1469-8986.2011.01223.x

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(40), 10056–10061. https://doi.org/10.1523/JNEUROSCI.1776-08.2008

Broadbent, D. E. (1958). *Perception and communication*. Elmsford: Pergamon Press. https://doi.org/10.1037/10037-000

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. *PloS One*, *9*(9), 1-10. https://doi.org/10.1371/journal.pone.0106953

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Bruce, C., Delafield, G., Bonnie, L., Winwood, M., & Gale, A. (1972). Stimulus complexity and the EEG: Differential effects of the number and the variety of display elements. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *26*(2), 155–170. https://doi.org/10.1037/h0082424

Bruce, V., Green, P. R., & Georgeson, M. A. (2006). *Visual perception: Physiology, psychology, & ecology* (4. ed., repr). Hove: Psychology Press.

Bruijn, D. de, Jansen, J., Lenior, D., Schreibers, K., & Pikaar, R. (2016). Human factor guidelines and workload in CCTV design. In D. de Waard, A. Toffetti, R. Wiczorek, A. Sonderegger, S. Röttger, P. Bouchner, & T. Franke (Eds.), *Proceedings of the Human Factors and Ergonomics Society (HFES) Europe Chapter Annual Conference, October 26-28, 2016 in Prague, Czech Republic* (pp. 209–219).

Buettner, R. (2017). The Relationship Between Visual Website Complexity and a User's Mental Workload: A NeuroIS Perspective. In F. D. Davis, R. Riedl, J. Vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Lecture Notes in Information Systems and Organisation. Information Systems and Neuroscience* (Vol. 16, pp. 107–113). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41402-7_14

Burge, W. K., Ross, L. A., Amthor, F. R., Mitchell, W. G., Zotov, A., & Visscher, K. M. (2013). Processing speed training increases the efficiency of attentional resource allocation in young adults. *Frontiers in Human Neuroscience*, *7*, 684. https://doi.org/10.3389/fnhum.2013.00684

Bürgin, R., & Ritschard, G. (2017). Coefficient-Wise Tree-Based Varying Coefficient Regression with vcrpart. *Journal of Statistical Software*, *80*(6), 1–33. https://doi.org/10.18637/jss.v080.i06

Buswell, G. T. (1935). *How people look at pictures: a study of the psychology and perception in art*. *How people look at pictures: a study of the psychology and perception in art*. Oxford, England: Univ. Chicago Press.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and Raw Task Load Index (TLX) Correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety, I* (pp. 481–485). London: Taylor & Francis.

Cambridge Dictionary (2019). COMPLEXITY | Bedeutung im Cambridge Englisch Wörterbuch. Retrieved from https://dictionary.cambridge.org/de/worterbuch/englisch/complexity*, last accessed 28.02.2019.

Campbell, F. W., Cooper, G. F., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *The Journal of Physiology*, *203*(1), 223–235. https://doi.org/10.1113/jphysiol.1969.sp008861

Cardaci, M., Di Gesù, V., Petrou, M., & Tabacchi, M. E. (2009). A fuzzy approach to the evaluation of image complexity. *Fuzzy Sets and Systems*, *160*(10), 1474–1484. https://doi.org/10.1016/j.fss.2008.11.017

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 660–675. https://doi.org/10.1037/0096-1523.34.3.660

Casti, J. L. (1979). *Connectivity, complexity and catastrophe in large-scale systems* (Vol. 7). Chichester, UK: John Wiley & Sons.

Cavalcante, A., Mansouri, A., Kacha, L., Barros, A. K., Takeuchi, Y., Matsumoto, N., & Ohnishi, N. (2014). Measuring streetscape complexity based on the statistics of local contrast and spatial frequency. *PloS One*, *9*(2), e87097. https://doi.org/10.1371/journal.pone.0087097

Chaitin, G. J. (1977). Algorithmic Information Theory. *IBM Journal of Research and Development*, *21*(4), 350–359. https://doi.org/10.1147/rd.214.0350

Chamorro-Premuzic, T., Burke, C., Hsu, A., & Swami, V. (2010). Personality predictors of artistic preferences as a function of the emotional valence and perceived complexity of paintings. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4), 196–204. https://doi.org/10.1037/a0019211

Chang, D., Dooley, L., & Tuovinen, J. E. (2002). Gestalt theory in visual screen design: a new look at an old subject. In *Selected Papers from the 7th World Conference on Computers in Education (WCCE’01), Copenhagen, Computers in Education 2001: Australian Topics, Volume 8* (pp. 5–12). Melbourne: Australian Computer Society, Inc.

Chassy, P., Lindell, T. A. E., Jones, J. A., & Paramei, G. V. (2015). A Relationship Between Visual Complexity and Aesthetic Appraisal of Car Front Images: An Eye-Tracker Study. *Perception*, *44*(8-9), 1085–1097. https://doi.org/10.1177/0301006615596882

Chater, N. (1997). Simplicity and the mind. *The Psychologist*, *10*(11), 495–498.

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22. https://doi.org/10.1016/S1364-6613(02)00005-0

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(8), 790–799. https://doi.org/10.1109/34.400568

Chikhman, V., Bondarko, V., Danilova, M., Goluzina, A., & Shelepin, Y. (2012). Complexity of images: Experimental and computational estimates compared. *Perception*, *41*(6), 631–647. https://doi.org/10.1068/p6987

Chipman, S. F. (1977). Complexity and structure in visual patterns. *Journal of Experimental Psychology: General*, *106*(3), 269–301. https://doi.org/10.1037/0096-3445.106.3.269

Chipman, S. F., & Mendelson, M. J. (1979). Influence of six types of visual structure on complexity judgments in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(2), 365–378. https://doi.org/10.1037/0096-1523.5.2.365

Christensen, R. H. B. (2018). *ordinal–Regression Models for Ordinal Data: R package version 2018.8-25.* Retrieved from http://www.cran.r-project.org/package=ordinal/*, last accessed 14.05.2019.

Cinaz, B., La Marca, R., Arnrich, B., & Tröster, G. (2010). Monitoring of mental workload levels. In M. Macedo (Ed.), *Proceedings of IADIS International Conference e-Health 2010 in Freiburg, Germany* (pp. 189–193). International Association for Development of the Information Society.

Ciocca, G., Corchs, S., Gasparini, F., Bricolo, E., & Tebano, R. (2015). Does Color Influence Image Complexity Perception? In A. Trémeau, R. Schettini, & S. Tominaga (Eds.), *Computational Color Imaging* (pp. 139–148). Cham: Springer International Publishing.

Clark, P. J., & Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, *35*(4), 445–453. https://doi.org/10.2307/1931034

Collins Dictionary (2019a). Complexity Definition und Bedeutung | Collins Wörterbuch. Retrieved from https://www.collinsdictionary.com/de/worterbuch/englisch/complexity*, last accessed 31.10.2019.

Collins Dictionary (2019b). Definition of 'user interface'. Retrieved from https://www.collinsdictionary.com/dictionary/english/user-interface*, last accessed 04.11.2019.

Colosi, R. (2005). Negatively worded questions cause respondent confusion. In *Proceedings of the Survey Research Methods Section, Joint Statistical Meeting 2005 in Minneapolis, Minnesota* (pp. 2896–2903). American Statistical Association.

Complex - Wiktionary (2019, February 24). Retrieved from https://en.wiktionary.org/wiki/complex*, last accessed 24.02.2019.

Corchs, S. E., Ciocca, G., Bricolo, E., & Gasparini, F. (2016). Predicting Complexity Perception of Real World Images. *PloS One*, *11*(6), 1-22. https://doi.org/10.1371/journal.pone.0157986

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cox, D. S., & Cox, A. D. (1988). What Does Familiarity Breed? Complexity as a Moderator of Repetition Effects in Advertisement Evaluation. *Journal of Consumer Research*, *15*(1), 111. https://doi.org/10.1086/209149

Cummings, M., Sasangohar, F., Thornburg, K. M., Xing, J., & D'Agostino, A. (2010). Human-system interface complexity and opacity part i: literature review. *Massachusetts: Institute of Technology Editors*.

Da Silva, M. P., Courboulay, V., & Estraillier, P. (2011). Image complexity measure based on visual attention. In *18th IEEE International Conference on Image Processing (ICIP), 2011: 11 - 14 Sept. 2011, Brussels, Belgium* (pp. 3281–3284). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICIP.2011.6116371

Davies, A., Vigo, M., Harper, S., & Jay, C. (2016). The visualisation of eye-tracking scanpaths: What can they tell us about how clinicians view electrocardiograms? In

M. Burch, L. L. Chuang, & A. T. Duchowski (Eds.), *Proceedings of the second Workshop on Eye Tracking and Visualization: Etvis 2016* (pp. 79–83). Piscataway, NJ: IEEE. https://doi.org/10.1109/ETVIS.2016.7851172

Davis, M., Sigal, R., & Weyuker, E. J. (1994). *Computability, complexity, and languages: Fundamentals of theoretical computer science* (2. ed.). *Computer science and scientific computing*. San Diego: Acad. Press.

Day, H. (1967). Evaluations of subjective complexity, pleasingness and interestingness for a series of random polygons varying in complexity. *Perception & Psychophysics*, *2*(7), 281–286. https://doi.org/10.3758/BF03211042

Day, H. (1968). The importance of symmetry and complexity in the evaluation of complexity, interest and pleasingness. *Psychonomic Science*, *10*(10), 339–340. https://doi.org/10.3758/BF03331550

Demberg, V., Sayeed, A., Mahr, A., & Müller, C. (2013). Measuring linguistically-induced cognitive load during driving using the ConTRe task. In J. Terken (Ed.), *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Eindhoven, 2013, [October 28 - 30, 2013]* (pp. 176–183). New York, NY: ACM. https://doi.org/10.1145/2516540.2516546

Deng, & Poole (2010). Affect in Web Interfaces: A Study of the Impacts of Web Page Visual Complexity and Order. *MIS Quarterly*, *34*(4), 711. https://doi.org/10.2307/25750702

Desnoyer, M., & Wettergreen, D. (2010). Aesthetic Image Classification for Autonomous Agents. In *20th International Conference on Pattern Recognition (ICPR), 2010: 23 - 26 Aug. 2010, Istanbul, Turkey ; proceedings* (pp. 3452–3455). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICPR.2010.843

Deutsch, J. A., & d. Deutsch (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90. https://doi.org/10.1037/h0039515

Deutsches Institut für Normung e.V. (2008). *DIN EN ISO 11064-5:2008-10, Ergonomische Gestaltung von Leitzentralen - Teil 5: Anzeigen und Stellteile (ISO 11064-5:2008)*. (DIN EN ISO, 11064-5). Berlin: Beuth Verlag GmbH.

Deutsches Institut für Normung e.V. (2009). *DIN EN ISO 14738:2009-07, Sicherheit von Maschinen - Anthropometrische Anforderungen an die Gestaltung von Maschinenarbeitsplätzen (ISO_14738:2002_+ Cor._1:2003_+ Cor._2:2005)*. (DIN EN ISO, 14738). Berlin: Beuth Verlag GmbH.

Deutsches Institut für Normung e.V. (2018). *Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts*. (ISO/DIS , 9241--11:2018). Berlin: Beuth Verlag GmbH.

Di Nocera, F., Mastrangelo, S., Colonna, S. P., Steinhage, A., Baldauf, M., & Kataria, A. (2015). Mental workload assessment using eye-tracking glasses in a simulated maritime scenario. *Proceedings of the Human Factors and Ergonomics Society Europe, Groningen, the Netherlands*, 14–16.

Di Nocera, F., Terenzi, M., & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbour as a measure of mental workload. In D. de Waard, K. A. Brookhuis, & A. Toffetti (Eds.), *Developments in Human Factors in Transportation, Design, and Evaluation* (pp. 295–303). Maastricht, the Netherlands: Shaker Publishing.

Dinges, D. F., & Grace, R. (1998). PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. *US Department of Transportation, Federal Highway Administration, Publication Number FHWA-MCRT-98-006*.

Djokic, J., Lorenz, B., & Fricke, H. (2010). Air traffic control complexity as workload driver. *Transportation Research Part C: Emerging Technologies*, *18*(6), 930–936. https://doi.org/10.1016/j.trc.2010.03.005

Dlugosch, C., Conti, A. S., & Bengler, K. (2013). Driver Distraction through Conversation Measured with Pupillometry. In *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: Driving Assessment 2013* (pp. 198–204). Iowa City, Iowa: University of Iowa. https://doi.org/10.17077/drivingassessment.1488

Donderi, D. C. (2006a). An information theory analysis of visual complexity and dissimilarity. *Perception*, *35*(6), 823–835. https://doi.org/10.1068/p5249

Donderi, D. C. (2006b). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73–97. https://doi.org/10.1037/0033-2909.132.1.73

Donderi, D. C., & McFadden, S. (2005). Compressed file length predicts search time and errors on visual displays. *Displays*, *26*(2), 71–78. https://doi.org/10.1016/j.displa.2005.02.002

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*(1), 53–78. https://doi.org/10.1348/000712601162103

Duchowski, A. T. (2017). *Eye Tracking Methodology: Theory and Practice*. Cham, s.l.: Springer International Publishing. Retrieved from http://dx.doi.org/10.1007/978-3-319-57883-5 https://doi.org/10.1007/978-3-319-57883-5

Duchowski, A. T., & Krejtz, K. (2017). Visualizing Dynamic Ambient/Focal Attention with Coefficient $$K$$. In M. Burch, L. Chuang, B. Fisher, & A. Schmidt (Eds.), *Mathematics and Visualization. Eye Tracking and Visualization: Foundations, Techniques, and Applications. ETVIS 2015* (Vol. 27, pp. 217–233). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47024-5_13

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., . . . Giannopoulos, I. (2018). The Index of Pupillary Activity. In R. Mandryk & M. Hancock (Eds.), *Engage with CHI: Chi 2018 : Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems : April 21 -26, 2018, Montréal, QC, Canada* (pp. 1–13). New York, New York: The Association for Computing Machinery. https://doi.org/10.1145/3173574.3173856

Durlauf, S. N. (2005). Complexity and Empirical Economics. *The Economic Journal*, *115*(504), F225-F243. https://doi.org/10.1111/j.1468-0297.2005.01003.x

Edmonds, B. (1999). What is Complexity?-The philosophy of complexity per se with application to some examples in evolution. In *The evolution of complexity*. Kluwer, Dordrecht.

Edquist, J., Rudin-Brown, C. M., & Lenné, M. G. (2012). The effects of on-street parking and road environment visual complexity on travel speed and reaction time. *Accident; Analysis and Prevention*, *45*, 759–765. https://doi.org/10.1016/j.aap.2011.10.001

Eilers, K., Nachreiner, F., & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift Für Arbeitswissenschaft*. (4), 214–224.

Eisenman, R. (1967). Birth-order and sex differences in aesthetic preference for complexity-simplicity. *The Journal of General Psychology*, *77*(1st Half), 121–126. https://doi.org/10.1080/00221309.1967.9710762

Eisenman, R. (1968). Personality and demography in complexity-simplicity. *Journal of Consulting and Clinical Psychology*, *32*(2), 140–143. https://doi.org/10.1037/h0025484

Elawady, M., Ducottet, C., Alata, O., Barat, C., & Colantoni, P. (2017). Wavelet-Based Reflection Symmetry Detection via Textural and Color Histograms: Algorithm and Results. In *Proceedings of the 2017 IEEE International Conference on Computer Vision workshops, 22-29 October 2017, Venice, Italy*. Piscataway, NJ: IEEE.

Endsley, M. R. (2016). *Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition* (2nd): CRC Press.

Endsley, M. R. (2019). A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM. *Human Factors*, *63*(1), 124-150. https://doi.org/10.1177/0018720819875376

Endsley, M. R., & Jones, D. G. (2012). *Designing for situation awareness: An approach to user-centered design* (Second edition). Boca Raton, London, New York: CRC Press Taylor & Francis Group.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 97–120. https://doi.org/10.1016/j.trf.2005.04.012

Ergoneers D-Lab (Version 3.50) [Computer software] (2017): Ergoneers GmbH. Retrieved from https://www.ergoneers.com/mess-software-und-analyse-software/d-lab/*, last accessed 02.01.2021.

Eve, R. A., Horsfall, S., & Lee, M. E. (2003). *Chaos, complexity, and sociology: Myths, models, and theories* ([Nachdr.]). Thousand Oaks, Calif.: Sage Publ.

Eysenck, M. W., & Keane, M. T. (2000). *Cognitive psychology: A student's handbook* (4. edition). Hove, UK: Psychology Press.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied Ergonomics*, *52*, 95–103. https://doi.org/10.1016/j.apergo.2015.07.009

Farley, F. H., & Weinstock, C. A. (1980). Experimental aesthetics: Children's complexity preference in original art and photoreproductions. *Bulletin of the Psychonomic Society*, *15*(3), 194–196. https://doi.org/10.3758/BF03334506

Faure, V., Lobjois, R., & Benguigui, N. (2016). The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *40*, 78–90. https://doi.org/10.1016/j.trf.2016.04.007

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, *12*(6), 893–924. https://doi.org/10.1080/13506280444000571

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633. https://doi.org/10.1038/35036586

Feldman, J. (2004). How surprising is a simple pattern? Quantifying "Eureka!". *Cognition*, *93*(3), 199–224. https://doi.org/10.1016/j.cognition.2003.09.013

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

Forsythe, A. (2009). Visual Complexity: Is That All There Is? In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 158–166). Berlin, Heidelberg: Springer Berlin Heidelberg.

Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods*, *40*(1), 116–129.

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*(1), 49–70. https://doi.org/10.1348/000712610X498958

Forsythe, A., Sheehy, N., & Sawey, M. (2003). Measuring icon complexity: An automated analysis. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 334–342. https://doi.org/10.3758/BF03202562

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/*, last accessed 02.01.2021.

Fredrikson, M., & Ohman, A. (1979). Heart-rate and electrodermal orienting responses to visual stimuli differing in complexity. *Scandinavian Journal of Psychology*, *20*(1), 37–41.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. Retrieved from http://www.jstatsoft.org/v33/i01/

Funke, F. (2016). A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales. *Social Science Computer Review*, *34*(2), 244–254. https://doi.org/10.1177/0894439315575477

Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the Smart: Type of Rating Scale on the Web Interacts With Educational Level. *Social Science Computer Review*, *29*(2), 221–231. https://doi.org/10.1177/0894439310376896

Gale, A., Coles, M., & Boyd, E. (1971). Variation in visual input and the occipital EEG: II. *Psychonomic Science*, *23*(1), 99–100. https://doi.org/10.3758/BF03336026

Gale, A., Spratt, G., Christie, B., & Smallbone, A. (1975). Stimulus complexity, EEG abundance gradients, and detection efficiency in a visual recognition task. *British Journal of Psychology*, *66*(3), 289–298. https://doi.org/10.1111/j.2044-8295.1975.tb01464.x

García, M., Badre, A. N., & Stasko, J. T. (1994). Development and validation of icons varying in their abstractness. *Interacting with Computers*, *6*(2), 191–211. https://doi.org/10.1016/0953-5438(94)90024-8

Gartus, A., & Leder, H. (2017). Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception. *PloS One*, *12*(11), e0185276. https://doi.org/10.1371/journal.pone.0185276

Gartus, A., & Leder, H. (2018). *A two-factor model of the visual complexity of abstract patterns (and some preliminary data about the complexity of textures)*. Poster presentation. Tagung experimentell arbeitender Psychologen, 11.03.-14.03.2018, Marburg.

Gaudette, L., & Japkowicz, N. (2009). Evaluation Methods for Ordinal Classification. In Y. Gao & N. Japkowicz (Eds.), *Advances in Artificial Intelligence* (pp. 207–210). Berlin, Heidelberg: Springer Berlin Heidelberg.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Geissler, G. L., Zinkhan, G. M., & Watson, R. T. (2006). The Influence of Home Page Complexity on Consumer Attention, Attitudes, and Purchase Intent. *Journal of Advertising*, *35*(2), 69–80. https://doi.org/10.1080/00913367.2006.10639232

GetFireShot: FireShot [Computer software] (2021). Retrieved from https://getfire-shot.com/*, last accessed 02.01.2021.

Geva, R., Zivan, M., Warsha, A., & Olchik, D. (2013). Alerting, orienting or executive attention networks: Differential patters of pupil dilations. *Frontiers in Behavioral Neuroscience*, *7*, 145. https://doi.org/10.3389/fnbeh.2013.00145

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Oxford, England: Houghton Mifflin.

Giesbrecht, B., Woldorff, M.G., Song, A.W., & Mangun, G.R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *NeuroImage*, *19*(3), 496–512. https://doi.org/10.1016/S1053-8119(03)00162-9

Goldberg, J. H., & Helfman, J. I. (2010). Visual scanpath representation. In S. N. Spencer (Ed.), *Etra 2010: Symposium on Eye Tracking Research & Applications ; Austin, TX, USA, March 22 - 24, 2010* (p. 203). New York, NY: ACM SIGGRAPH. https://doi.org/10.1145/1743666.1743717

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7

Goldstein, E. B. (2010). *Sensation and perception* (8. ed.). Belmont, Calif.: Wadsworth Cengage Learning.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance* (pp. 1–49). Oxford, England: John Wiley & Sons.

Gordon, I. E. (2004). *Theories of visual perception* (3rd ed.). Hove, East Sussex, New York: Psychology Press.

Graef, P. de (2005). Semantic effects on object selection in real-world scene perception. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 189–211). Oxford University Press: Oxford.

Gregory, R. L. (1970). *The intelligent eye*. New York: McGraw-Hill.

Groll, A. (2018). *glmmLasso: Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation. R package version 1.6.0*. Retrieved from https://cran.r-project.org/web/packages/glmmLasso/index.html*, last accessed 02.01.2021.

Groll, A., & Tutz, G. (2014). Variable selection for generalized linear mixed models by L 1-penalized estimation. *Statistics and Computing*, *24*(2), 137–154. https://doi.org/10.1007/s11222-012-9359-z

Güçlütürk, Y., Jacobs, R. H. A. H., & van Lier, R. (2016). Liking versus Complexity: Decomposing the Inverted U-curve. *Frontiers in Human Neuroscience*, *10*, 112. https://doi.org/10.3389/fnhum.2016.00112

Hajjem, A., Bellavance, F., & Larocque, D. (2014). Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, *84*(6), 1313–1328. https://doi.org/10.1080/00949655.2012.741599

Hall, A. C. (1969). Measures of the complexity of random black and white and coloured stimuli. *Perceptual and Motor Skills*, *29*(3), 773–774. https://doi.org/10.2466/pms.1969.29.3.773

Halverson, T., Estepp, J., Christensen, J., & Monnin, J. (2012). Classifying Workload with Eye Movements in a Complex Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56th Meeting in Boston, Massachusetts*, *56*(1), 168–172. https://doi.org/10.1177/1071181312561012

Harper, S., Jay, C., Michailidou, E., & Quan, H. (2013). Analysing the visual complexity of web pages using document structure. *Behaviour & Information Technology*, *32*(5), 491–502. https://doi.org/10.1080/0144929X.2012.726647

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

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In N. Meshkati & P. A. Hancock (Eds.), *Advances in Psychology. Human Mental Workload* (Vol. 52, pp. 139–183). Oxford: Elsevier Science. https://doi.org/10.1016/S0166-4115(08)62386-9

Hasler, D., & Suesstrunk, S. E. (2003). Measuring colorfulness in natural images. In B. E. Rogowitz & T. N. Pappas (Eds.), *SPIE Proceedings, Human Vision and Electronic Imaging VIII, Santa Clara, CA* (p. 87). SPIE. https://doi.org/10.1117/12.477378

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behaviour & Information Technology*, *25*(2), 91–97. https://doi.org/10.1080/01449290500330331

Hastie, T., Tibshirani, R., & Friedman, J. H. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (Second edition, corrected at 12th printing 2017). *Springer series in statistics*. New York, NY: Springer.

Hatfield, G., & Epstein, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin*, *97*(2), 155–186. https://doi.org/10.1037/0033-2909.97.2.155

Heaps, C., & Handel, S. (1999). Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 299–320. https://doi.org/10.1037/0096-1523.25.2.299

Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., & Deml, B. (2017). Electrocardiographic features for the measurement of drivers' mental workload. *Applied Ergonomics*, *61*, 31–43. https://doi.org/10.1016/j.apergo.2016.12.015

Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*. Leipzig: Voss.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Chapter 25 - Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel (Ed.), *Eye movements: A window on mind and brain* (537-III). Amsterdam: Elsevier. https://doi.org/10.1016/B978-008044980-7/50027-6

Henderson, J. M., Weeks, P. A., JR., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. https://doi.org/10.1037/0096-1523.25.1.210

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349–350. https://doi.org/10.1126/science.132.3423.349

Heylighen, F. (1997). *The growth of structural and functional complexity during evolution*. Dordrecht: Kluwer Academic.

Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, *46*(5), 361. https://doi.org/10.1037/h0055809

Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2nd edition). Charleston, SC: CreateSpace.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident; Analysis and Prevention*, *38*(1), 185–191. https://doi.org/10.1016/j.aap.2005.09.007

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*(2), 79–102. https://doi.org/10.1016/j.ijhcs.2005.06.002

Howard, C., Troscianko, T., Gilchrist, I., Behera, A., & Hogg, D. (2009). Searching for threat: Factors determining performance during CCTV monitoring. In D. de Waard, J. Godthelp, F. L. Kooi, & K. A. Brookhuis (Eds.), *Human Factors, Security and Safety* (pp. 1–7). Maastricht, the Netherlands: Shaker Publishing.

Hübner, R., & Fillinger, M. G. (2016). Comparison of Objective Measures for Predicting Perceptual Balance and Visual Aesthetic Preference. *Frontiers in Psychology*, *7*, 1–15. https://doi.org/10.3389/fpsyg.2016.00335

Hugo, J., & Gertman, D. (2013). A qualitative method to estimate HSI display complexity. *Nuclear Engineering and Technology*, *45*(2), 141–150. https://doi.org/10.5516/NET.04.2013.701

Hutcheson, G., & Sofroniou, M. (1999). *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models.* Thousand Oaks, CA: SAGE Publication. https://doi.org/10.4135/9780857028075

Hutton, S. B., & Tegally, D. (2005). The effects of dividing attention on smooth pursuit eye tracking. *Experimental Brain Research*, *163*(3), 306–313. https://doi.org/10.1007/s00221-004-2171-z

Ichikawa, S. (1985). Quantitative and structural factors in the judgment of pattern complexity. *Perception & Psychophysics*, *38*(2), 101–109.

Imamoglu, Ç. (2000). Complexity, Liking and Familiarity: Architecture and non-architecture Turkish student's assessment of traditional and modern house facades. *Journal of Environmental Psychology*, *20*(1), 5–16. https://doi.org/10.1006/jevp.1999.0155

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task-evoked pupillary response to mental workload in human-computer interaction. In E. Dykstra-Erickson (Ed.), *Chi*

'04 Extended Abstracts on Human Factors in Computing Systems (p. 1477). New York, NY: ACM. https://doi.org/10.1145/985921.986094

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*(2), 146–154. https://doi.org/10.1016/j.learninstruc.2009.02.019

Johnson, N. F. (2009). *Simply complexity: A clear guide to complexity theory* (First published in trade paperback). *A Oneworld book*. Oxford: Oneworld.

Jones, J. M., Ma, R., Starkey, R. L., & Ma, Z. (2007). Information complexity and appropriate interface design in nuclear power plant control rooms. In *2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting in Monterey, CA* (pp. 45–49). IEEE. https://doi.org/10.1109/HFPP.2007.4413177

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Kaftan, J. N., Bell, A. A., & Aach, T. (2008). Mean Shift Segmentation - Evaluation of Optimization Techniques. In *Proceedings of the Third International Conference on Computer Vision Theory and Applications, VISAPP 2008* (pp. 365–374). Funchal, Madeira - Portugal: INSTICC - Institute for Systems and Technologies of Information, Control and Communication.

Kahneman, D. (1973). *Attention and effort*. *Prentice Hall series in experimental psychology*. Englewood Cliffs: Prentice Hall.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kayaert, G., Beeck, H. P. op de, & Wagemans, J. (2011). Dynamic prototypicality effects in visual search. *Journal of Experimental Psychology. General*, *140*(3), 506–519. https://doi.org/10.1037/a0023494

Ke, Y., Tang, X., & Jing, F. (2006). The Design of High-Level Features for Photo Quality Assessment. In C. Schmid, S. Soatto, & C. Tomasi (Eds.), *Conference on Computer Vision and Pattern Recognition workshop, 2006: Cvpr 2006 ; New York City, New York, 17 - 22 June, 2006* (pp. 419–426). Piscataway, NJ: IEEE. https://doi.org/10.1109/CVPR.2006.303

Kersten, J. (2013). The Enjoyment of Complexity: A New Political Anthropology for the Anthropocene? *RCC Perspectives*. (3), 39–56. Retrieved from http://www.jstor.org/stable/26240508*, last accessed 03.01.2021.

Kimchi, R., & Hadad, B.-S. (2002). Influence of past experience on perceptual grouping. *Psychological Science*, *13*(1), 41–47. https://doi.org/10.1111/1467-9280.00407

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt, Brace and Company.

Köhler, W. (1947). *Gestalt psychology; an introduction to new concepts in modern psychology, Rev. ed*. *Gestalt psychology: an introduction to new concepts in modern psychology, Rev. ed*. Oxford, England: Liveright.

Kootstra, G., Boer, B. de, & Schomaker, L. R. B. (2011). Predicting Eye Fixations on Complex Visual Stimuli Using Local Symmetry. *Cognitive Computation*, *3*(1), 223–240. https://doi.org/10.1007/s12559-010-9089-5

Kotval, X. P., & Goldberg, J. H. (1998). Eye Movements and Interface Component Grouping: An Evaluation Method. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(5), 486–490. https://doi.org/10.1177/154193129804200509

Kovesi, P. (2000). Phase congruency: A low-level image invariant. *Psychologische Forschung*, *64*(2), 136–148. https://doi.org/10.1007/s004260000024

Krejtz, K., Coltekin, A., Duchowski, A., & Niedzielska, A. (2017). Using coefficient K to distinguish ambient/focal visual attention during map viewing. *Journal of Eye Movement Research*, *10*(2), 1–13. https://doi.org/10.16910/jemr.10.2.3

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning Ambient/Focal Attention with Coefficient K. *ACM Transactions on Applied Perception*, *13*(3), 1–20. https://doi.org/10.1145/2896452

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., . . . Villalobos, N. (2015). Gaze Transition Entropy. *ACM Transactions on Applied Perception*, *13*(1), 1–20. https://doi.org/10.1145/2834121

Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. In S. N. Spencer, P. Qvarfordt, & D. W. Hansen (Eds.), *ETRA 2014: Proceedings : Safety Harbor, Florida, March 26 - 28, 2014* (pp. 159–166). New York: Association for Computing Machinery, Inc. https://doi.org/10.1145/2578153.2578176

Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta – A System for Feature Selection. *Fundamenta Informaticae*, *101*, 271–285. https://doi.org/10.3233/FI-2010-288

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*(25-26), 3559–3565. https://doi.org/10.1016/S0042-6989(01)00102-X

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report a-8. University of Florida, Gainesville, Fl.*

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.), *Lecture Notes in Computer Science: Vol. 5298. HCI and usability for education and work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20 - 21, 2008 ; proceedings* (Vol. 5298, pp. 63–76). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-89350-9_6

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197.

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In A. T. Duchowski & R. Vertegaal (Eds.), *Proceedings of the 2004 symposium on Eye tracking research & applications* (pp. 41–48). New York, NY: ACM. https://doi.org/10.1145/968363.968370

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P.O.S., & Kort, J. (2009). Understanding, scoping and defining user experience. In D. R. Olsen, R. B. Arthur, K. Hinckley, M. R. Morris, S. Hudson, & S. Greenberg (Eds.), *Chi 2009 - digital life, new world: Conference proceedings and extended abstracts ; the 27th Annual CHI Conference on Human Factors in Computing Systems, April 4 - 9, 2009 in Boston, USA* (p. 719). New York, NY: ACM. https://doi.org/10.1145/1518701.1518813

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489–508. https://doi.org/10.1348/0007126042369811

Lee, S. C., Kim, Y. W., & Ji, Y. G. (2019). Effects of visual complexity of in-vehicle information display: Age-related differences in visual search task in the driving context. *Applied Ergonomics*, *81*, 1–8. https://doi.org/10.1016/j.apergo.2019.102888

Leeuwenberg, E. L. (1968). *Structural information of visual patterns: an efficient coding system in perception*. Berlin, Boston: De Gruyter.

Leeuwenberg, E. L. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, *76*(2), 216–220. https://doi.org/10.1037/h0027285

Leigh, J. H., & Kinnear, T. C. (1980). On Interaction Classification. *Educational and Psychological Measurement*, *40*(4), 841–843. https://doi.org/10.1177/001316448004000405

Leiner, D. (2013). *Too Fast, Too Straight, Too Weird: Post Hoc Identification of Meaningless Data in Internet Surveys*. Working Paper. Munich. Retrieved from https://www.researchgate.net/publication/258997762*, last accessed 06.01.2021.

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Retrieved from https://CRAN.R-project.org/package=emmeans*, last accessed 20.11.2020.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

Li, M., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications*. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-49820-1

Lin, Y., & Cai, H. (2009). A method for building a real-time cluster-based continuous mental workload scale. *Theoretical Issues in Ergonomics Science*, *10*(6), 531–543. https://doi.org/10.1080/14639220902836547

Liu, Y., Hsueh, P.-Y., Lai, J., Sangin, M., Nussli, M.-A., & Dillenbourg, P. (2009). Who is the expert? Analyzing gaze data to predict expertise level in collaborative applications. In *2009 IEEE International Conference on Multimedia and Expo: ICME 2009 : proceedings : June 28-July 3, 2009, Waldorf-Astoria Hotel, New York, New York, U.S.A* (pp. 898–901). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICME.2009.5202640

Locher, P., & Nodine, C. (1989). The perceptual value of symmetry. *Computers & Mathematics with Applications*, *17*(4-6), 475–484. https://doi.org/10.1016/0898-1221(89)90246-0

Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A Significance Test for the Lasso. *Annals of Statistics*, *42*(2), 413–468. https://doi.org/10.1214/13-AOS1175

Lok, S., Feiner, S., & Ngai, G. (2004). Evaluation of visual balance for automated layout. In J. Vanderdonckt, N. J. Nunes, & C. Rich (Eds.), *Proceedings of the 9th international conference on Intelligent user interface - IUI '04* (p. 101). New York, New York, USA: ACM Press. https://doi.org/10.1145/964442.964462

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception & Psychophysics*, *73*(6), 1780–1789. https://doi.org/10.3758/s13414-011-0144-4

Mace, D. J., & Pollack, L. (1983). Visual complexity and sign brightness in detection and recognition of traffic signs. *Transportation Research Record*, *20*(904), 33–41.

Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, *160*, 43–57. https://doi.org/10.1016/j.actpsy.2015.06.005

Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, *9*(12), 11.1-11. https://doi.org/10.1167/9.12.11

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics*, *2*(11), 547–552. https://doi.org/10.3758/BF03210264

Madan, C. R., Bayer, J., Gamer, M., Lonsdorf, T. B., & Sommer, T. (2017). Visual Complexity and Affect: Ratings Reflect More Than Meets the Eye. *Frontiers in Psychology*, *8*, 1–19. https://doi.org/10.3389/fpsyg.2017.02368

Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research*, *13*(7), 1255–1267. https://doi.org/10.1016/0042-6989(73)90201-0

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, *9*(3), 363–386. https://doi.org/10.1163/156856895x00052

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165–188. https://doi.org/10.1163/156856896x00123

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Rösler (Ed.), *Ergebnisse und Anwendungen der Psychophysiologie. Enzyklopädie der Psychologie* (pp. 799–864). Göttingen: Hogrefe.

Marin, M. M., & Leder, H. (2013). Examining complexity across domains: Relating subjective and objective measures of affective environmental scenes, paintings and music. *PloS One*, *8*(8), e72412. https://doi.org/10.1371/journal.pone.0072412

Marquart, G., Cabrall, C., & Winter, J. de (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing*, *3*, 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783

Marr, D. (1980). Visual information processing: The structure and creation of visual representations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *290*(1038), 199–218. https://doi.org/10.1098/rstb.1980.0091

Marshall, S. P. (2002). The Index of Cognitive Activity: Measuring cognitive workload. In J. J. Persensky (Ed.), *New century, new trends: Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants ; September 15 - 19, 2002, Scottsdale, Arizona* ((7) 5–9). New York, NY: IEEE. https://doi.org/10.1109/HFPP.2002.1042860

Martin, B. A. S., Sherrard, M. J., & Wentzel, D. (2005). The role of sensation seeking and need for cognition on Web-site evaluations: A resource-matching perspective. *Psychology and Marketing*, *22*(2), 109–126. https://doi.org/10.1002/mar.20050

Martinez-Conde, S., & Macknik, S. L. (2015). From Exploration to Fixation: An Integrative View of Yarbus's Vision. *Perception*, *44*(8-9), 884–899. https://doi.org/10.1177/0301006615594963

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

MathWorks (2019). Correct Nonuniform Illumination and Analyze Foreground Objects. Retrieved from https://de.mathworks.com/help/images/correcting-nonuniform-illumination.html*, last accessed 07.11.2019.

MathWorks (2020). Quadtree Decomposition - MATLAB & Simulink. Retrieved from https://de.mathworks.com/help/images/quadtree-decomposition.html*, last accessed 26.01.2020.

The Mathworks, Inc (2018). *MATLAB version 9.5.0.944444 (R2018b)*. Natick, Massachusetts.

McDougall, S. J. P., Curry, M. B., & Bruijn, O. de (1999). Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 487–519. https://doi.org/10.3758/BF03200730

Mehler, B., Reimer, B., & Wang, Y. (2011). A Comparison of Heart Rate and Heart Rate Variability Indices in Distinguishing Single-Task Driving and Driving Under Secondary Cognitive Workload. In *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : driving assessment 2011* (pp. 590–597). Iowa City, Iowa: University of Iowa. https://doi.org/10.17077/drivingassessment.1451

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. *An approach to environmental psychology*. Cambridge, MA, US: The MIT Press.

Michailidou, E., Harper, S., & Bechhofer, S. (2008). Visual complexity and aesthetic perception of web pages. In H. O'Neill (Ed.), *Proceedings of the 26th annual ACM international conference on Design of communication* (p. 215). New York, NY: ACM. https://doi.org/10.1145/1456536.1456581

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miniukovich, A., & Angeli, A. de (2014). Quantification of interface visual complexity. In P. Paolini & F. Garzotto (Eds.), *AVI 2014 : proceedings of the 2014 International Working Conference on Advanced Visual Interfaces : May 27-30, 2014, Como, Italy* (pp. 153–160). New York, New York, USA: ACM Press. https://doi.org/10.1145/2598153.2598173

Miniukovich, A., Sulpizio, S., & Angeli, A. de (2018). Visual complexity of graphical user interfaces. In T. Catarci, M. Mecella, F. Leotta, A. Marrella, & K. Norman (Eds.), *ICPS, AVI 2018: Proceedings of the 2018 International Conference on Advanced Visual Interfaces : May 29-June 1, 2018, Riva del Sole, Castiglione della Pescaia, Grosseto (Italy)* (pp. 1–9). New York, New York: The Association for Computing Machinery. https://doi.org/10.1145/3206505.3206549

Moffitt, K. (1980). Evaluation of the fixation duration in visual search. *Perception & Psychophysics*, *27*(4), 370–372. https://doi.org/10.3758/bf03206127

Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). *The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature*. U.S. Department of Transportation, Federal Aviation Administration (FAA) Report Number DOT/FAA/CT-TN95/22. Atlantic City, NJ. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a297433.pdf*, last accessed 06.01.2021.

Moray, N. (2013). *Mental workload: Its theory and measurement* (Vol. 8). New York: Springer Science & Business Media.

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PloS One*, *7*(5), e37527. https://doi.org/10.1371/journal.pone.0037527

Moshagen, M., & Thielsch, M. (2013). A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology*, *32*(12), 1305–1311. https://doi.org/10.1080/0144929X.2012.694910

Mulder, G. (1979). Mental Load, Mental Effort and Attention. In N. Moray (Ed.), *NATO Conference Series: Vol. 8. Mental Workload: Its Theory and Measurement* (Vol. 37, pp. 299–325). Boston, MA: Springer US; Imprint; Springer. https://doi.org/10.1007/978-1-4757-0884-4_18

Mulder, L. J. M., Dijksterhuis, C., Stuiver, A., & Waard, D. de (2009). Cardiovascular state changes during performance of a simulated ambulance dispatchers' task: Potential use for adaptive support. *Applied Ergonomics*, *40*(6), 965–977. https://doi.org/10.1016/j.apergo.2009.01.009

Musatti, C. (1930). I fattori empirici della percezione e la teoria della forma. *Rivista Di Psicologia*. (26), 259–264.

Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual Complexity and Beauty Appreciation: Explaining the Divergence of Results. *Empirical Studies of the Arts*, *28*(2), 173–191. https://doi.org/10.2190/EM.28.2.d

Nadkarni, & Gupta (2007). A Task-Based Model of Perceived Website Complexity. *MIS Quarterly*, *31*(3), 501. https://doi.org/10.2307/25148805

Näsänen, R., Kukkonen, H., & Rovamo, J. (1993). Spatial integration of band-pass filtered patterns in noise. *Vision Research*, *33*(7), 903–911. https://doi.org/10.1016/0042-6989(93)90073-6

Norman, D. A. (2016). *Living with complexity* (first paperback edition). Cambridge, Mass. [u.a.]: MIT Press.

Noton, D., & Stark, L. (1971a). Eye Movements and Visual Perception. *Scientific American*, *224*(6), 34–43.

Noton, D., & Stark, L. (1971b). Scanpaths in Eye Movements during Pattern Perception. *Science*, *171*(3968), 308–311. https://doi.org/10.1126/science.171.3968.308

Noton, D., & Stark, L. (1971c). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*(9), 929-IN8. https://doi.org/10.1016/0042-6989(71)90213-6

Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, *24*(2), 370–392. https://doi.org/10.3758/s13423-016-1124-4

Oliva, A. (2005). Gist of the Scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 251–256). Amsterdam: Elsevier Academic Press. https://doi.org/10.1016/B978-012375731-9/50045-8

Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity in scenes. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society, August 4-7, 2004, Chicago, Illinois* (pp. 1041–1046). New Jersey: Cognitive Science Society.

Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, *26*(9), 1277–1294. https://doi.org/10.1016/0031-3203(93)90135-J

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In S. N. Spencer (Ed.), *Etra 2010: Symposium on Eye Tracking Research & Applications ; Austin, TX, USA, March 22 - 24, 2010* (p. 141). New York, NY: ACM SIGGRAPH. https://doi.org/10.1145/1743666.1743701

Palumbo, L., Ogden, R., Makin, A. D. J., & Bertamini, M. (2014). Examining visual complexity and its influence on perceived duration. *Journal of Vision*, *14*(14), 3. https://doi.org/10.1167/14.14.3

Pandir, M., & Knight, J. (2006). Homepage aesthetics: The search for preference factors and the challenges of subjectivity. *Interacting with Computers*, *18*(6), 1351–1370. https://doi.org/10.1016/j.intcom.2006.03.007

Paré, D. E., & Cree, G. S. (2009). Web-based image norming: How do object familiarity and visual complexity ratings compare when collected in-lab versus online? *Behavior Research Methods*, *41*(3), 699–704. https://doi.org/10.3758/BRM.41.3.699

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123. https://doi.org/10.1016/S0042-6989(01)00250-4

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*(2), 125–154. https://doi.org/10.1163/15685680360511645

Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. Z. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors*, *33*(4), 389–400. https://doi.org/10.1177/001872089103300402

Peysakhovich, V., Causse, M., Scannella, S., & Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *97*(1), 30–37. https://doi.org/10.1016/j.ijpsycho.2015.04.019

Phillips, M. H., & Edelman, J. A. (2008). The dependence of visual scanning performance on saccade, fixation, and perceptual metrics. *Vision Research*, *48*(7), 926–936. https://doi.org/10.1016/j.visres.2007.12.020

Pierno, A. C., Caria, A., Glover, S., & Castiello, U. (2005). Effects of increasing visual load on aurally and visually guided target acquisition in a virtual environment. *Applied Ergonomics*, *36*(3), 335–343. https://doi.org/10.1016/j.apergo.2004.11.002

Pikaar, R., Lenior, D., Schreibers, K., & Bruijn, D. de (2015). Human Factors Guidelines for CCTV system design. In G. Lindgaard & D. Moore (Eds.), *Proceedings 19th Triennial Congress of the IEA, Melbourne 9-14 August 2015* (Vol. 9, pp. 1–8). Melbourne: IEA.

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A. F., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, *13*(3), 16. https://doi.org/10.1167/13.3.16

Poitschke, T. M. (2011). *Blickbasierte Mensch-Maschine-Interaktion im Automobil* (München, Technische Universität München, Diss., 2011). Universitätsbibliothek der TU München, München.

Pollen, D. A., & Ronner, S. F. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*(5), 907–916. https://doi.org/10.1109/TSMC.1983.6313086

Pomplun, M., & Sunkara, S. (2003). Pupil Dilation as an Indicator of Cognitive Workload in Human-Computer Interaction. In V. D. D. Harris, M. Smith, & C. Stephanidis (Chairs), *Proceedings of the 10th International Conference on Human-Computer Interaction, June 22-27, 2003,* Crete, Greece.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, *129*(5), 643–673. https://doi.org/10.1037/0033-2909.129.5.643

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Core Team. Retrieved from https://www.R-project.org/*, last accessed 10.01.2021.

R. Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/*, last accessed 10.01.2021.

Rao, A. R., & Lohse, G. L. (1993). Identifying High Level Features of Texture Perception. *CVGIP: Graphical Models and Image Processing*, *55*(3), 218–233. https://doi.org/10.1006/cgip.1993.1016

Rauterberg, M. (1996). How to measure cognitive complexity in human-computer interaction. In R. Trappl (Ed.), *Cybernetics and systems '96: proceedings of the Thirteenth European Meeting on Cybernetics and Systems Research, Vienna, Austria, 9-12 April 1996* (pp. 815–820). Austrian Society for Cybernetic Studies.

Ray, S., & Turi, R. H. (1999). Determination of number of clusters in K-means clustering and application in colour segmentation. In *4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99), Calcutta, India, 27 - 29 December 1999* (pp. 137–143). New Delhi India: Narosa Publishing House.

Recarte, M. Á., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental Workload and Visual Impairment: Differences between Pupil, Blink, and Subjective Rating. *The Spanish Journal of Psychology*, *11*(2), 374–385. https://doi.org/10.1017/S1138741600004406

Reddy, L., Moradi, F., & Koch, C. (2007). Top-down biases win against focal attention in the fusiform face area. *NeuroImage*, *38*(4), 730–739. https://doi.org/10.1016/j.neuroimage.2007.08.006

Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, *4*(2), 106–117. https://doi.org/10.1167/4.2.4

Reinecke, K., & Gajos, K. Z. (2014). Quantifying visual preferences around the world. In M. Jones, P. Palanque, A. Schmidt, & T. Grossman (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 11–20). New York, NY: ACM. https://doi.org/10.1145/2556288.2557052

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., & Gajos, K. Z. (2013). Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In W. E. Mackay, S. Brewster, & S. Bødker (Eds.), *Chi 2013: Changing perspectives ; the 31st Annual CHI Conference on Human Factors in Computing Systems, conference proceedings, 27 April - 2 May 2013, Paris, France* (p. 2049). New York, NY: ACM. https://doi.org/10.1145/2470654.2481281

Renshaw, J. A., Finlay, J. E., Tyfa, D., & Ward, R. D. (2003). Designing for visual influence: An eye tracking study of the usability of graphical management information. In M. Rauterberg, M. Menozzi, & J. Wesson (Eds.), *Human-Computer Interaction INTERACT '03: IFIP TC13 International Conference on Human-Computer Interaction, 1st – 5th September 2003, Zurich, Switzerland* (Vol. 1, pp. 144–151). IOS Press.

Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois. Retrieved from https://CRAN.R-project.org/package=psych*, last accessed 10.01.2021.

Ries, F., & Deml, B. (2019). *Videoaufnahmen in Leitwarten: Belastungs- und Beanspruchungsanalyse*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. https://doi.org/10.21934/BAUA:BERICHT20190111

Ries, F., Wolf, E., Olschowski, P., Döllken, M., & Deml, B. (2018). *Disentangling effects of mental workload and visual stimulus properties on pupil size*. Retrieved from https://www.researchgate.net/publication/323943103_Disentangling_effects_of_mental_workload_and_visual_stimulus_properties_on_pupil_size*, last accessed 10.01.2021. https://doi.org/10.13140/RG.2.2.35819.85287

Riglis, E. (1998). *Modeling Complexity in Visual Images*. First Year PhD Report. Image Systems Engineering Laboratory. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.22.8892&rep=rep1&type=pdf*, last accessed 10.01.2021.

Rohmert, W. (1984). Das Belastungs-Beanspruchungs-Konzept. *Zeitschrift Für Arbeitswissenschaft*, *38*(4), 193–200.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, *34*(2), 259–287. https://doi.org/10.1016/0301-0511(92)90018-P

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17.1-22. https://doi.org/10.1167/7.2.17

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236. https://doi.org/10.1068/p5117

Roth, S. P., Schmutz, P., Pauwels, S. L., Bargas-Avila, J. A., & Opwis, K. (2010). Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages? *Interacting with Computers*, *22*(2), 140–152. https://doi.org/10.1016/j.intcom.2009.10.004

Rötting, M. (2001). *Parametersystematik der Augen-und Blickbewegungen für arbeitswissenschaftliche Untersuchungen*. Aachen: Shaker.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 5975–5979. https://doi.org/10.1073/pnas.0711295105

Saegert, S. C., & Jellison, J. M. (1970). Effects of initial level of response competition and frequency of exposure on liking and exploratory behavior. *Journal of Personality and Social Psychology*, *16*(3), 553–558. https://doi.org/10.1037/h0029952

Saklofske, D. H. (1975). Visual aesthetic complexity, attractiveness and diversive exploration. *Perceptual and Motor Skills*, *41*(3), 813–814. https://doi.org/10.2466/pms.1975.41.3.813

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. T. Duchowski (Ed.), *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71–78). New York, NY: ACM. https://doi.org/10.1145/355017.355028

Savage, S. W., Potter, D. D., & Tatler, B. W. (2013). Does preoccupation impair hazard perception? A simultaneous EEG and Eye Tracking study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *17*, 52–62. https://doi.org/10.1016/j.trf.2012.10.002

Schauerte, B., & Stiefelhagen, R. (2012). Quaternion-Based Spectral Saliency Detection for Eye Fixation Prediction. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *Computer Vision - ECCV 2012* (pp. 116–129). Berlin, Heidelberg: Springer Berlin Heidelberg.

Schlick, C. M., Winkelholz, C., Motz, F., Duckwitz, S., & Grandt, M. (2010). Complexity assessment of human–computer interaction. *Theoretical Issues in Ergonomics Science*, *11*(3), 151–173. https://doi.org/10.1080/14639220802609903

Schlochtermeier, L. H., Kuchinke, L., Pehrs, C., Urton, K., Kappelhoff, H., & Jacobs, A. M. (2013). Emotional picture and word processing: An FMRI study on effects of stimulus complexity. *PloS One*, *8*(2), e55619. https://doi.org/10.1371/journal.pone.0055619

Schneider, M., & Deml, B. (2016). An Integrated Approach of Mental Workload Assessment. In B. Deml, P. Stock, R. Bruder, & C. M. Schlick (Eds.), *Advances in Ergonomic Design of Systems, Products and Processes* (pp. 191–208). Berlin, Heidelberg: Springer Berlin Heidelberg.

Schwalm, M. (2009). *Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext*. Dissertation. Retrieved from Universität des Saarlandes website: https://publikationen.sulb.uni-saarland.de/handle/20.500.11880/23353*, last accessed 10.01.2021. https://doi.org/10.22028/D291-23297

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Seckler, M., Opwis, K., & Tuch, A. N. (2015). Linking objective design factors with subjective aesthetics: An experimental study on how structure and color of websites affect the facets of users' visual aesthetic perception. *Computers in Human Behavior*, *49*, 375–389. https://doi.org/10.1016/j.chb.2015.02.056

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465. https://doi.org/10.3758/s13428-017-0913-7

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1951). Prediction and Entropy of Printed English. *Bell System Technical Journal*, *30*(1), 50–64. https://doi.org/10.1002/j.1538-7305.1951.tb01366.x

Shao, J., Loy, C. C., & Wang, X. (2017). Learning Scene-Independent Group Descriptors for Crowd Understanding. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(6), 1290–1303. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7428859*, last accessed 10.01.2021.

Shepard, S., & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 3–11. https://doi.org/10.1037/0096-1523.14.1.3

Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajaratnam, S. M. W., Berlowitz, D. J., . . . Howard, M. E. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Scientific Reports*, *8*(1), 2220. https://doi.org/10.1038/s41598-018-20588-7

Shigeto, H., Ishiguro, J., & Nittono, H. (2011). Effects of visual stimulus complexity on event-related brain potentials and viewing duration in a free-viewing task. *Neuroscience Letters*, *497*(2), 85–89. https://doi.org/10.1016/j.neulet.2011.04.035

Simonin, J., Kieffer, S., & Carbonell, N. (2005). Effects of Display Layout on Gaze Activity During Visual Search. In M. F. Costabile & F. Paternò (Eds.), *Lecture Notes in Computer Science: Vol. 3585. Human-computer interaction - INTERACT 2005: IFIP TC13 international conference, Rome, Italy, September 12 - 16, 2005 ; proceedings* (Vol. 3585, pp. 1054–1057). Berlin: Springer. https://doi.org/10.1007/11555261_103

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews. Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Smith, G. F., & Dorfman, D. D. (1975). The effect of stimulus uncertainty on the relationship between frequency of exposure and liking. *Journal of Personality and Social Psychology*, *31*(1), 150–155. https://doi.org/10.1037/h0076174

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal* *of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

SR Research EyeLink Data Viewer [Computer software] (2019). Mississauga, Canada: SR Research.

Standish, R. K. (2008). Concept and Definition of Complexity. In A. Yang & Y. Shan (Eds.), *Intelligent Complex Adaptive Systems* (pp. 105–124). Hershey, Pennsylvania: IGI Global. https://doi.org/10.4018/978-1-59904-717-1.ch004

Stickel, C., Ebner, M., & Holzinger, A. (2010). The XAOS Metric – Understanding Visual Complexity as Measure of Usability. In G. Leitner, M. Hitz, & A. Holzinger (Eds.), *Lecture Notes in Computer Science: Vol. 6389. HCI in work and learning, life and leisure: 6th symposium of the workgroup human-computer interaction and usability engineering, USAB 2010, Klagenfurt, Austria, November 4-5, 2010 : proceedings* (Vol. 6389, pp. 278–290). Berlin: Springer. https://doi.org/10.1007/978-3-642-16607-5_18

Svensson, E., Angelborg-Thanderz, M., Sjoberg, L., & Olsson, S. (1997). Information complexity--mental workload and performance in combat aircraft. *Ergonomics*, *40*(3), 362–380. https://doi.org/10.1080/001401397188206

Tan, G. W., & Wei, K. K. (2006). An empirical study of Web browsing behaviour: Towards an effective Website design. *Electronic Commerce Research and Applications*, *5*(4), 261–271. https://doi.org/10.1016/j.elerap.2006.04.007

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Thomas, H. (1968). Spatial Models and Multidimensional Scaling of Random Shapes. *The American Journal of Psychology*, *81*(4), 551. https://doi.org/10.2307/1421059

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. Retrieved from www.jstor.org/stable/2346178*, last accessed 14.01.2021.

Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., . . . Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied Ergonomics*, *59*, 326–332. https://doi.org/10.1016/j.apergo.2016.09.013

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

Traub, J. F. (2003). Information-based Complexity. In A. Ralston, E. D. Reilly, & D. Hemmendinger (Eds.), *Encyclopedia of Computer Science* (pp. 850–854). Chichester, UK: John Wiley and Sons Ltd.

Treder, M. S. (2010). Behind the Looking-Glass: A Review on Human Symmetry Perception. *Symmetry*, *2*(3), 1510–1543. https://doi.org/10.3390/sym2031510

Treisman, A. M. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, *3*(6), 449–459. https://doi.org/10.1016/S0022-5371(64)80015-3

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tuch, A. N., Bargas-Avila, J. A., & Opwis, K. (2010). Symmetry and aesthetics in website design: It's a man's business. *Computers in Human Behavior*, *26*(6), 1831–1837. https://doi.org/10.1016/j.chb.2010.07.016

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

Tuch, A. N., Kreibig, S., Roth, S. P., Bargas-Avila, J., Opwis, K., & Wilhelm, F. (2011). The Role of Visual Complexity in Affective Reactions to Webpages: Subjective, Eye Movement, and Cardiovascular Responses. *IEEE Transactions on Affective Computing*, *2*(4), 230–236. https://doi.org/10.1109/T-AFFC.2011.18

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

Tullis, T. S. (1983). The formatting of alphanumeric displays: A review and analysis. *Human Factors*, *25*(6), 657–682. https://doi.org/10.1177/001872088302500604

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, *43*(3), 333–346. https://doi.org/10.1016/S0042-6989(02)00498-4

Ungerleider, L. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, *4*(2), 157–165. https://doi.org/10.1016/0959-4388(94)90066-3

Vala, H. J., & Baxi, A. (2013). A review on Otsu image segmentation algorithm. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *2*(2), 387–389.

Van der Helm, P. A., & Leeuwenberg, E. L. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review*, *103*(3), 429–456. https://doi.org/10.1037/0033-295x.103.3.429

Van Gompel, R. P. G. (Ed.) (2007). *Eye movements: A window on mind and brain*. Amsterdam: Elsevier.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). *Statistics and computing*. New York: Springer.

Verwey, W. B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics*, *43*(2), 187–209. https://doi.org/10.1080/001401300184558

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, *2*(3), 270–285. https://doi.org/10.1037//1076-898X.2.3.270

Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral Science*, *11*(2), 105–114. https://doi.org/10.1002/bs.3830110204

Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision*, *9*(1), 9–32. https://doi.org/10.1163/156856895X00098

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Heydt, R. von der (2012). A Century of Gestalt Psychology in Visual Perception I. Perceptual Grouping and Figure-Ground Organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: Ii. Conceptual and theoretical foundations. *Psychological Bulletin*, *138*(6), 1218–1252. https://doi.org/10.1037/a0029334

Wang, H.-F., Wang, P.-Y., Liao, C.-C., & Lin, Y.-Y. (2014). The Impact of Perceived Visual Complexity, Gender, and Cognitive Style on Children's Aesthetic Preferences for Learning Web Pages. In P. Isaias & K. Blashki (Eds.), *Advances in Human and Social Aspects of Technology. Human-computer interfaces and interactivity: Emergent research and applications* (pp. 248–265). Hershey, Pennsylvania (701 E. Chocolate Avenue, Hershey, Pa., 17033, USA): IGI Global. https://doi.org/10.4018/978-1-4666-6228-5.ch014

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10. https://doi.org/10.1016/j.dss.2014.02.007

Wasserman, L. (2013). Assumption-Free High-Dimensional Inference. Retrieved from https://normaldeviate.wordpress.com/2013/10/03/assumption-free-high-dimensional-inference/*, last accessed 25.5.2020.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, *4*(1), 301–350. https://doi.org/10.1007/BF00410640

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). Academic Press New York.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York. Retrieved from https://ggplot2-book.org/*, last accessed 14.01.2021.

Wierwille, W., Ellsworth, L. A., Wreggit, S., Fairbanks, J., & Kirn, L. (1994). *Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness.: National Highway Traffic Safety Administration Final Report: DOT HS 808 247.*

Wilson, A., & Chatterjee, A. (2005). The Assessment of Preference for Balance: Introducing a New Test. *Empirical Studies of the Arts*, *23*(2), 165–180. https://doi.org/10.2190/B1LR-MVF3-F36X-XR64

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, *12*(1), 3–18. https://doi.org/10.1207/S15327108IJAP1201_2

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*(1), 60–82. https://doi.org/10.1016/0749-5978(86)90044-0

Wright, M. N., & Ziegler, A. (2017). ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*(1). https://doi.org/10.18637/jss.v077.i01

Wundt, W. M. (1874). *Grundzüge der physiologischen Psychologie* (Vol. 1). Leipzig: W. Engelman.

Xing, J. (2007). Information Complexity in Air Traffic Control Displays. In J. A. Jacko (Ed.), *Human-Computer Interaction. HCI Applications and Services* (pp. 797–806). Berlin, Heidelberg: Springer Berlin Heidelberg.

Xing, J., & Manning, C. A. (2005). *Complexity and automation displays of air traffic control: Literature review and analysis*. U.S. Department of Transportation, Federal Aviation Administration (FAA) Report Number DOT/FAA/AM-05/4. Washington, D.C. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a460107.pdf*, last accessed 14.01.2021.

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In A. L. Yarbus (Ed.), *Eye Movements and Vision* (pp. 171–211). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-5379-7_8

Yendrikhovskij, S. N., Ridder, H. de, Fedorovskaya, E. A., & Blommaert, F.J.J. (1997). Colourfulness judgments of natural scenes. *Acta Psychologica*, *97*(1), 79–94. https://doi.org/10.1016/S0001-6918(97)00025-5

Yoon, S. H., Lim, J., & Ji, Y. G. (2015). Assessment model for perceived visual complexity of automotive instrument cluster. *Applied Ergonomics*, *46 Pt A*, 76–83. https://doi.org/10.1016/j.apergo.2014.07.005

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zajonc, R. B. (2001). Mere Exposure: A Gateway to the Subliminal. *Current Directions in Psychological Science*, *10*(6), 224–228. https://doi.org/10.1111/1467-8721.00154

Zheng, X. S., Chakraborty, I., Lin, J. J.-W., & Rauschenberger, R. (2009). Correlating low-level image statistics with users - rapid aesthetic and affective judgments of web pages. In D. R. Olsen, R. B. Arthur, K. Hinckley, M. R. Morris, S. Hudson, & S. Greenberg (Eds.), *Chi 2009 - digital life, new world: Conference proceedings and extended abstracts ; the 27th Annual CHI Conference on Human Factors in Computing Systems, April 4 - 9, 2009 in Boston, USA* (p. 1). New York, NY: ACM. https://doi.org/10.1145/1518701.1518703

Zijlstra, F. R. H., & van Doorn, L. (1985). *The construction of a scale to measure subjective effort*. Delft, The Netherlands: Delft University of Technology, Department of Philosophy and Social Sciences.

Supervised Bachelor- and Master Theses

Ding, Y. (2019). *Visuelle Komplexität bei Webseiten - eine Eyetracking-Studie mit realistischen Stimuli* (Master Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Feng, X. (2017). *Cardiovascular parameters of visual complexity in videos* (Bachelor Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Ma, A. (2017). *Perspective 3D-depictions and mental workload: an investigation of the relationship based on eyetracking parameters* (Bachelor Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Olschowski, P. (2017). *Okulare Parameter und deren Zusammenhang mit visueller Komplexität und mentaler Beanspruchung* (Master Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Pennamen, A. (2019). *Visual complexity in websites - an eye-tracking investigation* (Master Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Vandière, S. de (2018). *An eye-tracking study on visual complexity* (Master Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Waibel, J. (2019). *Visuelle Komplexität von Webseiten - eine Online-Studie zu den Einflussfaktoren* (Bachelor Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Waldvogel, J. (2017). *Augmented Reality und mentale Beanspruchung: Untersuchung der Zusammenhänge anhand von elektrodermalen Parametern* (Bachelor Thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

Xie, Y. (2018). *What affects visual complexity? An online investigation of the influencing factors* (Master thesis). Karlsruhe Institute of Technology, Institute for Human and Industrial Engineering, Karlsruhe.

9. Appendix

9.1 Linear regression model for NASA-TLX in study 1

	NASA Mean		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	14.22	13.23 – 15.21	<.001
Frequency [1/6s]	-0.45	-1.13 – 0.24	.200
Frequency [1/9s]	-1.31	-1.99 – -0.62	<.001
Complexity [low]	-0.29	-0.97 – 0.40	.409
Frequency [1/6s]:Complexity [low]	-0.46	-1.43 – 0.50	.346
Frequency [1/9s]:Complexity [low]	-0.43	-1.40 – 0.54	.384
Random Effects			
σ^2	2.07		
$\tau_{00\ Subject}$	6.63		
ICC	0.76		
$N_{Subject}$	34		
Observations	204		
Marginal R^2 / Conditional R^2	.053 / .775		

9.2 Linear regression model for percentage of correct reactions in study 1

	Hit Ratio		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.86	0.81 – 0.90	<.001
Frequency [1/6s]	-0.01	-0.06 – 0.03	.482
Frequency [1/9s]	-0.02	-0.06 – 0.02	.291
Complexity [low]	0.05	0.01 – 0.09	.019
Frequency [1/6s]:Complexity [low]	0.01	-0.05 – 0.07	.804
Frequency [1/9s]:Complexity [low]	0.01	-0.04 – 0.07	.619
Random Effects			
σ^2	0.01		
τ_{00} Subject	0.01		
ICC	0.51		
N Subject	34		
Observations	204		
Marginal R^2 / Conditional R^2	.053 / .533		

9.3 Linear regression model for reaction times in study 1

	Reaction Time		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1376.94	1304.71 – 1449.18	<.001
Frequency [1/6s]	41.75	-32.52 – 116.02	.271
Frequency [1/9s]	-26.33	-100.59 – 47.94	.487
Complexity [low]	-99.97	-174.24 – -25.71	.008
Frequency [1/6s]:Complexity [low]	-160.39	-265.42 – -55.36	.003
Frequency [1/9s]:Complexity [low]	-16.41	-121.44 – 88.62	.759
Random Effects			
σ^2	24409.01		
τ_{00} Subject	21770.14		
ICC	0.47		
N Subject	34		
Observations	204		
Marginal R^2 / Conditional R^2	.147 / .549		

9.4 Linear regression model for RMSSD in study 1

	RMSSD		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	2.54	2.40 – 2.68	<0.001
Frequency [1/6s]	0.02	-0.06 – 0.09	0.676
Frequency [1/9s]	0.07	-0.00 – 0.14	0.061
Complexity [low]	0.02	-0.05 – 0.09	0.522
Frequency [1/6s]:Complexity [low]	-0.00	-0.10 – 0.10	0.969
Frequency [1/9s]:Complexity [low]	-0.01	-0.10 – 0.09	0.871
N Subject	32		
Observations	359		

9.5 Linear regression model for LF-Power in study 1

	ECG: LF Power		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	337.76	257.49 – 418.04	<0.001
Frequency [1/6s]	22.52	-51.38 – 96.74	0.551
Frequency [1/9s]	61.50	-12.64 – 135.70	0.104
Complexity [low]	86.11	13.84 – 158.48	0.019
Frequency [1/6s]:Complexity [low]	-64.37	-167.82 – 39.09	0.222
Frequency [1/9s]:Complexity [low]	-53.97	-156.45 – 48.51	0.301
Random Effects			
σ^2	36464.85		
τ_{00} Subject	29807.75		
ICC	0.45		
N Subject	32		
Observations	322		
Marginal R^2 / Conditional R^2	.017 / .459		

9.6 Linear regression model for PERCLOS in study 1

	PERCLOS		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.31	2.42 – 4.19	<0.001
Frequency [1/6s]	0.42	-0.03 – 0.88	0.066
Frequency [1/9s]	-0.09	-0.74 – 0.56	0.783
Complexity [low]	-0.23	-0.87 – 0.41	0.473
Frequency [1/6s]:Complexity [low]	0.73	0.27 – 1.18	0.002
Frequency [1/9s]:Complexity [low]	0.86	0.40 – 1.31	<0.001
Random Effects			
σ^2	1.57		
τ_{00} Subject	5.83		
ICC	0.79		
N Subject	33		
Observations	353		
Marginal R^2 / Conditional R^2	.027 / .794		

9.7 Regression model for visual complexity rating in study 2a)

	Visual complexity rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Threshold coefficients				
1|2	-1.07	0.54	-1.98	
2|3	1.15	0.54	2.14	
3|4	2.41	0.54	4.47	
4|5	3.59	0.54	6.65	
5|6	5.18	0.55	9.48	
6|7	7.41	0.56	13.22	
Coefficients				
Number of elements [2]	0.13	0.21	0.61	.54
Number of elements [3]	-0.07	0.26	-0.27	.79
Number of elements [4]	0.21	0.29	0.73	.47
Number of elements [5]	0.33	0.28	1.15	.25
Number of elements [6]	0.62	0.31	1.96	<.05
Number of elements [7]	2.10	0.39	5.44	<.0001
Variety of elements [2]	0.32	0.21	1.50	.13
Variety of elements [3]	1.19	0.26	4.56	<.0001
Variety of elements [4]	1.36	0.29	4.74	<.0001
Variety of elements [5]	1.47	0.29	5.07	<.0001
Variety of elements [6]	1.80	0.32	5.57	<.0001
Variety of elements [7]	1.66	0.43	3.82	<.001
Density of elements [2]	0.23	0.20	1.11	.27
Density of elements [3]	0.47	0.23	2.05	<.05
Density of elements [4]	0.81	0.25	3.29	<.01
Density of elements [5]	1.20	0.26	4.58	<.0001

Density of elements [6]	1.44	0.28	5.11	<.0001
Density of elements [7]	1.56	0.37	4.17	<.0001
Variety of Colours [2]	0.64	0.22	2.95	<.01
Variety of Colours [3]	0.92	0.24	3.83	<.001
Variety of Colours [4]	0.94	0.26	3.58	<.001
Variety of Colours [5]	1.13	0.28	4.11	<.0001
Variety of Colours [6]	0.93	0.31	3.03	<.01
Variety of Colours [7]	1.57	0.40	3.93	<.001
Colour contrast [2]	0.27	0.29	0.91	.36
Colour contrast [3]	0.47	0.30	1.58	.12
Colour contrast [4]	0.37	0.31	1.20	.23
Colour contrast [5]	-0.04	0.31	-0.12	.91
Colour contrast [6]	0.29	0.32	0.92	.36
Colour contrast [7]	0.26	0.37	0.70	.49
Organization [2]	0.03	0.28	0.10	.92
Organization [3]	0.19	0.29	0.66	.51
Organization [4]	-0.09	0.30	-0.29	.78
Organization [5]	-0.02	0.31	-0.06	.95
Organization [6]	-0.18	0.32	-0.58	.56
Organization [7]	-0.60	0.36	-1.70	.09
Symmetry [2]	0.12	0.23	0.54	.59
Symmetry [3]	-0.16	0.24	-0.67	.50
Symmetry [4]	-0.50	0.25	-1.97	<.05
Symmetry [5]	-0.14	0.26	-0.54	.59
Symmetry [6]	-0.55	0.28	-1.98	<.05
Symmetry [7]	-0.56	0.33	-1.72	.09
Visual Balance [2]	0.18	0.43	0.43	.67
Visual Balance [3]	-0.16	0.41	-0.40	.69

Visual Balance [4]	-0.07	0.42	-0.16	.87
Visual Balance [5]	-0.01	0.43	-0.03	.98
Visual Balance [6]	0.18	0.44	0.41	.68
Visual Balance [7]	-0.01	0.47	-0.01	.99
Random Effects				
σ^2	3.29			
τ_{00} Subject	0.74			
τ_{00} Stimulus	0.52			
ICC	0.28			
N Subject	94			
N Stimulus	18			
Observations	1692			
Marginal R^2 / Conditional R^2	.451 / .603			

9.8 Factor loadings and regression with factor scores for study 2a)

	Factor 1	Factor 2	Factor 3	Communality
Number of elements	0.93	-0.18	0.10	1.1
Variety of elements	0.80	-0.23	0.28	1.4
Density of elements	0.85	-0.15	0.11	1.1
Variety of colours	0.42	-0.02	0.71	1.6
Colour contrast	0.08	0.29	0.51	1.7
Organization	0.37	0.74	-0.05	1.5
Symmetry	-0.26	0.76	-0.06	1.2
Visual balance	-0.08	0.75	0.21	1.2
SS loadings	2.62	1.89	0.91	
Proportion Variance	0.33	0.24	0.11	
Cumulative Variance	0.33	0.56	0.68	
Proportion Explained	0.48	0.35	0.17	
Cumulative Proportion	0.48	0.83	1.00	

	Visual complexity rating		
Predictors	*Odds Ratios*	*CI*	*p*
Threshold 1|2	0.03	0.02 – 0.04	<.001
Threshold 2|3	0.24	0.16 – 0.35	<.001

Threshold 3|4	0.80	0.55 – 1.18	.262
Threshold 4|5	2.53	1.72 – 3.72	<.001
Threshold 5|6	11.63	7.79 – 17.37	<.001
Threshold 6|7	95.46	60.64 – 150.26	<.001
Factor 1 score	5.49	4.60 – 6.55	<.001
Factor 2 score	0.53	0.46 – 0.61	<.001
Factor 3 score	1.71	1.46 – 2.00	<.001
Random Effects			
σ^2	3.29		
τ_{00} Subject	0.71		
τ_{00} Stimulus	0.47		
ICC	0.26		
N Subject	94		
N Stimulus	18		
Observations	1692		
Marginal R^2 / Conditional R^2	.431 / .581		

9.9 Stimuli table of Study 2b)

Category	Complexity	Website
News	high	#01 Triggerfish
News	high	#02 ProPhysik
News	high	#03 Die Presse
News	high	#04 Parfümerie
News	medium	#05 Der Standard
News	medium	#06 ArtInWords
News	medium	#07 ArtScene
News	medium	#08 Vol.at
News	low	#09 DasNeueste
News	low	#10 Bilanz
News	low	#11 LBV
News	low	#12 ArchiNews
Online-Shop	high	#13 Lubera

Online-Shop	high	#14 Baldur
Online-Shop	high	#15 Easy
Online-Shop	high	#16 Puzzle
Online-Shop	medium	#17 Bakker
Online-Shop	medium	#18 Kunstsupermarkt
Online-Shop	medium	#19 Gourvita
Online-Shop	medium	#20 Ikarus
Online-Shop	low	#21 Wogg
Online-Shop	low	#22 Kaffeeonline
Online-Shop	low	#23 Zeitwunder
Online-Shop	low	#24 Wagner
Company	high	#25 StammMetall
Company	high	#26 Reichenbacher
Company	high	#27 Medienversicherung
Company	high	#28 ETHMess
Company	medium	#29 Tucher
Company	medium	#30 MaxBoegl
Company	medium	#31 LeckerProdukte
Company	medium	#32 Heidenhain
Company	low	#33 Ospa
Company	low	#34 BrotamHaken
Company	low	#35 Lotto
Company	low	#36 SHK

9.10 Regression model for visual complexity rating in study 2b)

	Visual complexity rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Threshold coefficients				
1|2	-1.28	0.53	-2.42	
2|3	0.62	0.53	1.18	
3|4	1.77	0.53	3.34	
4|5	3.17	0.53	5.96	
5|6	4.80	0.54	8.96	
6|7	6.75	0.54	12.47	
Coefficients				
Number of elements [2]	1.62	0.39	4.19	<.0001
Number of elements [3]	2.36	0.41	5.78	<.0001
Number of elements [4]	2.82	0.43	6.64	<.0001
Number of elements [5]	3.28	0.43	7.6	<.0001
Number of elements [6]	3.73	0.45	8.29	<.0001
Number of elements [7]	5.15	0.49	10.5	<.0001
Variety of elements [2]	-0.49	0.35	-1.43	.15
Variety of elements [3]	-0.3	0.36	-0.83	.41
Variety of elements [4]	-0.11	0.37	-0.3	.76
Variety of elements [5]	-0.04	0.37	-0.12	.91
Variety of elements [6]	0.21	0.39	0.54	.59
Variety of elements [7]	0.65	0.44	1.5	.13
Density of elements [2]	0.06	0.28	0.23	.82
Density of elements [3]	0.44	0.3	1.44	.15
Density of elements [4]	0.62	0.33	1.92	.06

Density of elements [5]	0.83	0.33	2.5	<.05
Density of elements [6]	1.57	0.35	4.54	<.0001
Density of elements [7]	1.71	0.39	4.36	<.0001
Variety of Colours [2]	-0.02	0.24	-0.1	.92
Variety of Colours [3]	-0.07	0.25	-0.27	.79
Variety of Colours [4]	0.23	0.26	0.9	.37
Variety of Colours [5]	0.08	0.27	0.31	.76
Variety of Colours [6]	0.41	0.28	1.47	.14
Variety of Colours [7]	0.01	0.31	0.03	.97
Colour contrast [2]	-0.06	0.29	-0.21	.83
Colour contrast [3]	0.04	0.29	0.14	.89
Colour contrast [4]	-0.17	0.29	-0.57	.57
Colour contrast [5]	-0.01	0.29	-0.04	.97
Colour contrast [6]	-0.03	0.3	-0.08	.93
Colour contrast [7]	0.31	0.33	0.95	.34
Organization [2]	-0.46	0.36	-1.29	.20
Organization [3]	-0.37	0.36	-1.02	.30
Organization [4]	-0.54	0.36	-1.47	.14
Organization [5]	-0.66	0.37	-1.77	.09
Organization [6]	-1.15	0.38	-3.01	<.01
Organization [7]	-1.31	0.43	-3.05	<.01
Symmetry [2]	-0.4	0.36	-1.12	.26
Symmetry [3]	-0.08	0.36	-0.21	.83
Symmetry [4]	0	0.36	-0.01	.99
Symmetry [5]	0.29	0.36	0.79	.43
Symmetry [6]	0.23	0.37	0.63	.53
Symmetry [7]	0.67	0.42	1.6	.11
Visual Balance [2]	0.69	0.4	1.72	.09

Visual Balance [3]	0.55	0.4	1.36	.17
Visual Balance [4]	0.41	0.41	0.99	.32
Visual Balance [5]	0.2	0.42	0.47	.64
Visual Balance [6]	0.33	0.43	0.77	.44
Visual Balance [7]	0.7	0.46	1.5	.13
Random Effects				
σ^2	3.29			
τ_{00} Subject	0.89			
τ_{00} Stimulus	0.07			
ICC	0.23			
N Subject	60			
N Stimulus	36			
Observations	2160			
Marginal R^2 / Conditional R^2	.486 / .602			

9.11 Factor loadings and regression with factor scores for study 2b)

	Factor 1	**Factor 2**	**Factor 3**	**Communality**
Number of elements	0.92	-0.14	0.17	1.1
Variety of elements	0.78	-0.03	0.26	1.2
Density of elements	0.84	-0.19	0.16	1.2
Variety of colours	0.26	0.03	0.95	1.2
Colour contrast	0.17	0.21	0.55	1.5
Organization	-0.21	0.90	0.04	1.1
Symmetry	-0.06	0.77	0.01	1.0
Visual balance	-0.08	0.81	0.08	1.0
SS loadings	2.31	2.17	1.33	
Proportion Variance	0.29	0.27	0.17	
Cumulative Variance	0.29	0.56	0.73	
Proportion Explained	0.40	0.37	0.23	
Cumulative Proportion	0.40	0.77	1.00	

	Visual complexity rating		
Predictors	*Odds Ratios*	*CI*	*p*
Threshold 1|2	0.01	0.01 – 0.01	<.001
Threshold 2|3	0.05	0.04 – 0.07	<.001

Threshold 3|4	0.17	0.13 – 0.23	<.001
Threshold 4|5	0.70	0.53 – 0.93	.013
Threshold 5|6	3.50	2.63 – 4.66	<.001
Threshold 6|7	21.63	15.81 – 29.58	<.001
Factor 1 score	6.25	5.45 – 7.18	<.001
Factor 2 score	0.68	0.61 – 0.77	<.001
Factor 3 score	1.48	1.33 – 1.63	<.001
Random Effects			
σ^2	3.29		
τ_{00} Subject	0.89		
τ_{00} Website	0.09		
ICC	0.23		
N Subject	60		
N Website	36		
Observations	2160		
Marginal R^2 / Conditional R^2	.445 / .573		

9.12 Ordinal Regression Model for Visual Complexity Ratings in Study 3

	Visual complexity ratings		
Predictors	*Odds Ratios*	*CI*	*p*
Threshold	0.02	0.01 – 0.04	<.001
spacing	5.53	5.13 – 5.96	<.001
Number of elements [9]	7.58	4.83 – 11.90	<.001
Number of elements [13]	32.72	20.46 – 52.32	<.001
Symmetry [0.5]	0.75	0.48 – 1.16	.194
Symmetry [1]	0.17	0.11 – 0.26	<.001
Number [9] * Symmetry [0.5]	0.54	0.29 – 1.00	.049
Number [13] * Symmetry [0.5]	0.53	0.28 – 0.99	.045
Number [9] * Symmetry [1]	0.37	0.20 – 0.70	.002
Number [13] * Symmetry [1]	0.54	0.29 – 1.02	.058
Random Effects			
σ^2	3.29		
τ_{00} Stimulus	0.05		
τ_{00} Subject	1.66		
τ_{00} Object	0.00		
ICC	0.34		
N Subject	33		
N Stimulus	54		
N Object	3		
Observations	1782		
Marginal R^2 / Conditional R^2	.338 / .564		

9.13 Linear Regression Model for Number of Fixations in Study 3

	Number of Fixations		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	16.56	15.06 – 18.05	<.001
Number of elements [9]	1.18	0.10 – 2.26	.033
Number of elements [13]	0.31	-0.77 – 1.39	.576
Symmetry [0.5]	-1.78	-2.54 – -1.01	<.001
Symmetry [1]	1.83	1.07 – 2.60	<.001
Number [9] * Symmetry [0.5]	-0.11	-1.19 – 0.97	.840
Number [13] * Symmetry [0.5]	-0.64	-1.40 – 0.12	.099
Number [9] * Symmetry [1]	0.91	-0.16 – 1.99	.097
Number [13] * Symmetry [1]	1.23	0.46 – 1.99	.002
Random Effects			
σ^2	12.00		
τ_{00} Stimulus	0.09		
τ_{00} Subject	13.62		
τ_{00} Object	0.29		
ICC	0.54		
N Subject	33		
N Stimulus	54		
N Object	3		
Observations	1782		
Marginal R^2 / Conditional R^2	.054 / .563		

9.14 Linear Regression model for Scanpath Length in Study 3

	Scanpath Length (in pixels)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	4326.05	3884.00 – 4768.11	<.001
Number of elements [9]	2577.26	1997.14 – 3157.38	<.001
Number of elements [13]	737.39	157.65 – 1317.12	.013
Symmetry [0.5]	-1880.19	-2290.13 – -1470.26	<.001
Symmetry [1]	-516.92	-927.02 – -106.82	.013
Number [9] * Symmetry [0.5]	872.53	292.91 – 1452.14	.003
Number [13] * Symmetry [0.5]	-1604.12	-2014.06 – -1194.19	<.001
Number [9] * Symmetry [1]	2098.15	1518.28 – 2678.01	<.001
Number [13] * Symmetry [1]	-273.38	-683.81 – 137.05	.192
Random Effects			
σ^2	1320774.04		
τ_{00} Stimulus	90786.57		
τ_{00} Subject	954542.65		
ICC	0.44		
N Subject	33		
N Stimulus	54		
Observations	1763		
Marginal R^2 / Conditional R^2	.190 / .548		

9.15 Regression model for Spatial Density in Study 3

	Spatial Density		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	10.30	9.26 – 11.35	<.001
Number of elements [9]	2.54	1.30 – 3.77	<.001
Number of elements [13]	0.42	-0.82 – 1.66	.507
Symmetry [0.5]	-1.54	-2.42 – -0.67	.001
Symmetry [1]	3.11	2.23 – 3.98	<.001
Number [9] * Symmetry [0.5]	1.32	0.09 – 2.56	.036
Number [13] * Symmetry [0.5]	-1.26	-2.13 – -0.38	.005
Number [9] * Symmetry [1]	1.02	-0.22 – 2.26	.106
Number [13] * Symmetry [1]	1.95	1.07 – 2.82	<.001
Random Effects			
σ^2	8.22		
τ_{00} Stimulus	0.35		
τ_{00} Subject	5.16		
τ_{00} Object	0.09		
ICC	0.40		
N Subject	33		
N Stimulus	54		
N Object	3		
Observations	1782		
Marginal R^2 / Conditional R^2	.204 / .526		

9.16 Ordinal Regression Model for single visual complexity ratings in study 3

	Visual Complexity Rating		
Predictors	*Estimate*	*SE*	*z*
Category-specific fixed effects			
1|2:(Intercept)	-3.74	0.18	-20.79
2|3:(Intercept)	-1.05	0.15	-7.16
3|4:(Intercept)	0.49	0.17	2.85
4|5:(Intercept)	1.73	0.11	15.39
5|6:(Intercept)	3.43	0.21	16.38
6|7:(Intercept)	5.39	0.35	15.46
Global fixed effects			
Number of Blinks	0.15	0.11	1.36
Number of Saccades	-0.24	0.13	-1.81
Average Drift	0.40	0.12	0.01
Coefficient K	-0.00	0.11	-0.01
ICA drop	0.03	0.09	0.34
TIFF filesize	0.92	0.74	1.24
Mean Edge Canny GIF	2.96	1.92	1.54
Mean Edge Perimeter GIF	-0.08	7.34	-0.01
SD Edge Canny GIF	-1.77	2.25	-0.79
Mean x SD Edge Perimeter GIF	-3.17	7.72	-0.41
Symmetry diagonal top left bottom right	1.00	0.23	4.36
Quads 1x1	0.07	0.27	0.27
Quads 32x32	0.22	0.25	0.89
Quads 128x128	-0.40	0.28	-1.41
APB Horizontal Inner Outer	-0.30	0.23	-1.30
APB Vertical Inner Outer	0.11	0.19	0.60

Homogeneity	-1.43	0.43	-3.37
Random Effects			
SD $_{Subject}$	1.10		

9.17 glmmLASSO Model for single visual complexity ratings with random effects Subject and stimulus in study 3

	Visual Complexity Rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Category-specific fixed effects				
Theta 1	-3.53	0.15	-23.45	<.0001
Theta 2	-0.87	0.09	-10.00	<.0001
Theta 3	0.65	0.09	7.52	<.0001
Theta 4	1.88	0.09	21.61	<.0001
Theta 5	3.52	0.09	38.94	<.0001
Theta 6	5.39	0.10	53.78	<.0001
Global fixed effects				
Homogeneity	-2.92	0.13	-22.62	<.0001
Random effects				
$SD_{Subject}$	0.98			
$SD_{Stimulus}$	0.83			

9.18 glmmLASSO Model for single visual complexity ratings with random effect Subject in study 3

	Visual Complexity Rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Category-specific fixed effects				
Theta 1	-3.18	0.13	-24.40	<.0001
Theta 2	-0.96	0.08	-12.44	<.0001
Theta 3	0.49	0.08	6.44	<.0001
Theta 4	1.70	0.08	21.99	<.0001
Theta 5	3.30	0.08	40.80	<.0001
Theta 6	5.08	0.09	56.03	<.0001
Global fixed effects				
TIFF filesize	-0.78	0.11	-7.25	<.0001
Homogeneity	-1.95	0.12	-15.95	<.0001
Random effects				
$SD_{Subject}$	0.99			

9.19 glmmLASSO Model for single visual complexity ratings without random effects in study 3

	Visual Complexity Rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Category-specific fixed effects				
Theta 1	-2.84	0.07	-42.44	<.0001
Theta 2	-0.69	0.03	-24.55	<.0001
Theta 3	0.56	0.03	21.21	<.0001
Theta 4	1.54	0.03	54.56	<.0001
Theta 5	2.80	0.03	83.31	<.0001
Theta 6	4.31	0.05	87.51	<.0001
Global fixed effects				
Homogeneity	-2.35	0.07	-33.09	<.0001

9.20 Regression Visual Complexity Rating for Study 4

	Visual Complexity Rating		
Predictors	*Odds Ratios*	*CI*	*p*
threshold.1	0.24	0.10 – 0.53	.001
spacing	5.79	5.43 – 6.16	<.001
Number of elements [6]	19.43	9.37 – 40.32	<.001
Number of elements [9]	51.52	24.59 – 107.93	<.001
Symmetry [B]	2.31	1.11 – 4.82	.025
Colourfulness [B]	1.14	0.55 – 2.35	.727
Prototypicality [B]	2.55	1.23 – 5.28	.012
Number [6] * Symmetry [B]	0.98	0.35 – 2.73	.962
Number [9] * Symmetry [B]	0.78	0.28 – 2.19	.641
Number [6] * Colour [B]	1.03	0.37 – 2.85	.955
Number [9] * Colour [B]	0.81	0.29 – 2.25	.679
Symmetry [B] * Colour [B]	1.15	0.41 – 3.23	.791
Number [6] * Prototyp. [B]	1.22	0.44 – 3.37	.708
Number [9] * Prototyp. [B]	0.95	0.34 – 2.64	.919
Symmetry [B] * Prototyp. [B]	0.42	0.15 – 1.20	.106
Colour [B] * Prototyp. [B]	0.94	0.34 – 2.63	.909
(Number [6] * Symmetry [B]) * Colour [B]	0.74	0.17 – 3.15	.683
(Number [9] * Symmetry [B]) * Colour [B]	0.75	0.18 – 3.22	.703
(Number [6] * Symmetry [B]) * Prototyp. [B]	1.77	0.41 – 7.56	.440
(Number [9] * Symmetry [B]) * Prototyp. [B]	1.73	0.40 – 7.38	.460
(Number [6] * Colour [B]) * Prototyp. [B]	0.66	0.16 – 2.78	.568
(Number [9] * Colour [B]) * Prototyp. [B]	1.25	0.29 – 5.31	.766
(Symmetry [B] * Colour [B]) * Prototyp. [B]	0.78	0.18 – 3.35	.735
(Number [B] * Symmetry [B] * Colour [B]) * Prototyp. [B]	1.55	0.20 – 12.04	.674

(Number [C] * Symmetriy [B] * Colour [B]) * Prototyp. [B]	1.30	0.17 – 10.15	.799
Random Effects			
σ^2	3.29		
τ_{00} Stimulus	0.11		
τ_{00} Subject	2.62		
τ_{00} Type	0.11		
ICC	0.46		
N Subject	40		
N Stimulus	72		
N Type	3		
Observations	2688		
Marginal R^2 / Conditional R^2	.326 / .639		

9.21 Regression Number of Fixations for Study 4

	Number of Fixations		
Predictors	*Estimate*	*CI*	*p*
(Intercept)	22.88	22.14 – 23.62	<.001
Number of elements [6]	-0.51	-0.64 – -0.37	<.001
Number of elements [9]	0.02	-0.11 – 0.16	.730
Symmetry [B]	0.17	0.08 – 0.26	<.001
Colourfulness [B]	-0.13	-0.26 – 0.00	.054
Prototypicality [B]	0.03	-0.11 – 0.16	.693
Number [6] * Symmetry [B]	-0.03	-0.16 – 0.11	.707
Number [9] * Symmetry [B]	0.04	-0.06 – 0.13	.460
Number [6] * Colour [B]	0.02	-0.07 – 0.11	.682
Number [9] * Colour [B]	0.08	-0.06 – 0.21	.258
Symmetry [B] * Colour [B]	0.24	0.10 – 0.37	.001
Number [6] * Prototyp. [B]	0.16	0.06 – 0.25	.001
Number [9] * Prototyp. [B]	0.07	-0.06 – 0.20	.307
Symmetry [B] * Prototyp. [B]	0.02	-0.12 – 0.15	.822
Colour [B] * Prototyp. [B]	-0.03	-0.16 – 0.11	.678
(Number [6] * Symmetry [B]) * Colour [B]	0.09	-0.00 – 0.19	.050
(Number [9] * Symmetry [B]) * Colour [B]	0.02	-0.08 – 0.11	.733
(Number [6] * Symmetry [B]) * Prototyp. [B]	-0.02	-0.16 – 0.11	.734
(Number [9] * Symmetry [B]) * Prototyp. [B]	0.01	-0.12 – 0.15	.834
(Number [6] * Colour [B]) * Prototyp. [B]	0.03	-0.11 – 0.16	.701
(Number [9] * Colour [B]) * Prototyp. [B]	0.07	-0.07 – 0.20	.318
(Symmetry [B] * Colour [B]) * Prototyp. [B]	-0.08	-0.21 – 0.05	.235
(Number [B] * Symmetry [B] * Colour [B]) * Prototyp. [B]	-0.00	-0.10 – 0.09	.930
(Number [C] * Symmetriy [B] * Colour [B]) * Prototyp. [B]	-0.01	-0.14 – 0.13	.929

Random Effects	
σ^2	5.74
τ_{00} Stimulus	0.01
τ_{00} Subject	4.54
τ_{00} Type	0.08
ICC	0.45
N Stimulus	72
N Subject	40
N Type	3
Observations	2684
Marginal R^2 / Conditional R^2	.020 / .458

9.22 Regression Scanpath Length for Study 4

	Scanpath Length		
Predictors	*Estimate*	*CI*	*p*
(Intercept)	4857.65	4551.40 – 5163.89	<.001
Number of elements [6]	-11.65	-108.46 – 85.17	.814
Number of elements [9]	32.90	-63.92 – 129.71	.505
Symmetry [B]	-70.25	-138.70 – -1.79	.044
Colourfulness [B]	-51.79	-148.60 – 45.01	.294
Prototypicality [B]	-11.52	-108.33 – 85.29	.816
Number [6] * Symmetry [B]	14.77	-82.04 – 111.58	.765
Number [9] * Symmetry [B]	-11.96	-80.42 – 56.49	.732
Number [6] * Colour [B]	-1.02	-69.47 – 67.44	.977
Number [9] * Colour [B]	5.48	-91.34 – 102.29	.912
Symmetry [B] * Colour [B]	43.50	-53.31 – 140.30	.379
Number [6] * Prototyp. [B]	87.69	19.24 – 156.15	.012
Number [9] * Prototyp. [B]	17.60	-79.21 – 114.40	.722
Symmetry [B] * Prototyp. [B]	0.73	-96.09 – 97.54	.988
Colour [B] * Prototyp. [B]	-41.38	-138.18 – 55.43	.402
(Number [6] * Symmetry [B]) * Colour [B]	-7.44	-75.90 – 61.01	.831
(Number [9] * Symmetry [B]) * Colour [B]	3.13	-65.32 – 71.59	.929
(Number [6] * Symmetry [B]) * Prototyp. [B]	-75.08	-171.89 – 21.74	.129
(Number [9] * Symmetry [B]) * Prototyp. [B]	6.02	-90.78 – 102.83	.903
(Number [6] * Colour [B]) * Prototyp. [B]	-59.42	-156.23 – 37.40	.229
(Number [9] * Colour [B]) * Prototyp. [B]	30.00	-66.82 – 126.81	.544
(Symmetry [B] * Colour [B]) * Prototyp. [B]	0.30	-96.50 – 97.11	.995
(Number [B] * Symmetry [B] * Colour [B]) * Prototyp. [B]	-13.34	-81.79 – 55.12	.703
(Number [C] * Symmetriy [B] * Colour [B]) * Prototyp. [B]	107.14	10.33 – 203.95	.030

Random Effects

σ^2	923120.86
τ_{00} Stimulus	63046.86
τ_{00} Subject	560388.28
τ_{00} Type	27526.71
ICC	0.41
N Subject	40
N Stimulus	72
N Type	3
Observations	2684
Marginal R^2 / Conditional R^2	.016 / .423

9.23 Regression Spatial Density for Study 4

	Spatial Density		
Predictors	*Estimate*	*CI*	*p*
(Intercept)	16.05	15.37 – 16.73	<.001
Number of elements [6]	-1.29	-1.47 – -1.11	<.001
Number of elements [9]	-0.00	-0.18 – 0.18	.963
Symmetry [B]	0.29	0.16 – 0.42	<.001
Colourfulness [B]	-0.14	-0.32 – 0.04	.119
Prototypicality [B]	-0.00	-0.19 – 0.18	.959
Number [6] * Symmetry [B]	0.04	-0.14 – 0.22	.683
Number [9] * Symmetry [B]	0.01	-0.12 – 0.14	.863
Number [6] * Colour [B]	0.04	-0.08 – 0.17	.507
Number [9] * Colour [B]	0.03	-0.15 – 0.21	.740
Symmetry [B] * Colour [B]	0.39	0.21 – 0.57	<.001
Number [6] * Prototyp. [B]	0.15	0.02 – 0.28	.020
Number [9] * Prototyp. [B]	-0.02	-0.20 – 0.16	.817
Symmetry [B] * Prototyp. [B]	-0.06	-0.24 – 0.12	.498
Colour [B] * Prototyp. [B]	-0.05	-0.23 – 0.13	.571
(Number [6] * Symmetry [B]) * Colour [B]	0.13	0.01 – 0.26	.041
(Number [9] * Symmetry [B]) * Colour [B]	0.06	-0.07 – 0.18	.378
(Number [6] * Symmetry [B]) * Prototyp. [B]	0.23	0.05 – 0.41	.012
(Number [9] * Symmetry [B]) * Prototyp. [B]	0.04	-0.14 – 0.22	.688
(Number [6] * Colour [B]) * Prototyp. [B]	-0.02	-0.21 – 0.16	.788
(Number [9] * Colour [B]) * Prototyp. [B]	0.04	-0.14 – 0.22	.687
(Symmetry [B] * Colour [B]) * Prototyp. [B]	0.03	-0.15 – 0.21	.734
(Number [B] * Symmetry [B] * Colour [B]) * Prototyp. [B]	-0.07	-0.20 – 0.06	.296
(Number [C] * Symmetriy [B] * Colour [B]) * Prototyp. [B]	0.03	-0.15 – 0.21	.767

Random Effects	
σ^2	7.27
$\tau_{00\ Stimulus}$	0.11
$\tau_{00\ Subject}$	1.82
$\tau_{00\ Type}$	0.21
ICC	0.23
$N_{Subject}$	40
$N_{Stimulus}$	72
N_{Type}	3
Observations	2684
Marginal R^2 / Conditional R^2	.101 / .306

9.24 Ordinal Regression model for single visual complexity ratings in study 4

	Visual Complexity Rating		
Predictors	*Estimate*	*SE*	*z*
Category-specific fixed effects			
1|2:(Intercept)	-4.74	0.18	-26.39
2|3:(Intercept)	-2.02	0.11	-17.95
3|4:(Intercept)	-0.43	0.12	-3.48
4|5:(Intercept)	0.72	0.10	7.48
5|6:(Intercept)	2.77	0.09	31.05
6|7:(Intercept)	4.71	0.11	42.22
Global fixed effects			
Mean Velocity	0.14	0.07	2.06
Mean Drift	-0.05	0.14	-0.36
Stationary Entropy	0.03	0.09	0.33
SD Nr. of Nodes	0.07	0.08	0.95
PERCLOS	0.11	0.07	1.44
SD Edge Phase Congruency GIF	-0.02	0.18	-1.13
Mean x SD Edge Perimeter GIF	-0.16	0.25	-0.64
Vertical Symmetry	0.42	0.21	2.05
Quads 4x4	-0.70	0.18	-3.92
RMS SD	-0.32	0.29	-1.13
APB Horizontal Inner Outer	-0.13	0.09	-1.39
APB Vertical	0.03	0.10	0.32
APB Vertical Inner Outer	0.08	0.06	1.29
Nr. of Segments	-0.19	0.16	-1.16
Average of Elawady et al.'s (2017) five largest symmetries	0.26	0.11	2.33

SD for Colour Congestion clutter map	0.12	0.08	1.43
SD for Orientation Congestion clutter map	0.41	0.22	1.85
Feature Congestion	-0.38	0.17	-2.28
Random Effects			
$SD_{Subject}$ 1.23			

9.25 glmmLASSO Model for single visual complexity ratings with random effects Subject and stimulus in study 4

	Visual Complexity Rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Category-specific fixed effects				
Theta 1	-5.04	0.19	-26.30	<.0001
Theta 2	-2.29	0.11	-20.46	<.0001
Theta 3	-0.69	0.11	-6.24	<.0001
Theta 4	0.47	0.11	4.25	<.0001
Theta 5	2.52	0.11	22.50	<.0001
Theta 6	4.49	0.12	37.95	<.0001
Global fixed effects				
Scanpath Length (by Fixations)	0.15	0.05	2.84	<.01
Vertical Symmetry	0.71	0.14	5.06	<.0001
Quads 4x4	-0.75	0.23	-3.27	<.01
Quads 8x8	-0.13	0.19	-0.70	.48
RMS SD	-0.73	0.19	-3.83	<.001
RMS Mean x SD	1.56	0.47	3.30	<.001
RMS JPEG Size	-0.60	0.66	-0.91	.36
Nr. of Segments	-0.48	0.20	-2.38	<.05
Näsänen Complexity	0.04	0.38	0.10	.92
Orientation	0.54	0.66	0.82	.41
Feature Congestion map Filesize	0.07	1.13	0.06	.95
Orientation map Filesize	-1.11	1.24	-0.90	.37
Random effects				
SD $_{Subject}$	1.63			

$SD_{Stimulus}$	0.44

9.26 glmmLASSO Model for single visual complexity ratings with random effect Subject in study 4

	Visual Complexity Rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Category-specific fixed effects				
Theta 1	-4.75	0.18	-26.53	<.0001
Theta 2	-2.16	0.10	-20.68	<.0001
Theta 3	-0.67	0.10	-6.49	<.0001
Theta 4	0.41	0.10	4.02	<.0001
Theta 5	2.37	0.10	22.62	<.0001
Theta 6	4.27	0.11	38.50	<.0001
Global fixed effects				
Quads 4x4	-0.93	0.10	-9.70	<.0001
Quads 8x8	-0.66	0.09	-7.46	<.0001
RMS JPEG Size	1.27	0.24	5.29	<.0001
Näsänen Complexity	-0.85	0.14	-6.27	<.0001
Orientation	-0.62	0.13	-4.97	<.0001
Random effects				
$SD_{Subject}$	1.55			

9.27 glmmLASSO Model for single visual complexity ratings without random effects in study 4

	Visual Complexity Rating			
Predictors	*Estimate*	*SE*	*z*	*p*
Category-specific fixed effects				
Theta 1	-3.63	0.07	-52.82	<.0001
Theta 2	-1.61	0.03	-64.30	<.0001
Theta 3	-0.44	0.02	-21.41	<.0001
Theta 4	0.38	0.02	18.71	<.0001
Theta 5	1.85	0.03	73.62	<.0001
Theta 6	3.45	0.04	83.26	<.0001
Global fixed effects				
Nr. of Fixations	0.27	0.04	6.74	<.0001
Vertical Symmetry	0.45	0.09	4.91	<.0001
Quads 4x4	-0.67	0.13	-5.34	<.0001
Quads 8x8	-0.10	0.12	-0.82	.41
RMS SD	-0.29	0.11	-2.61	<.01
RMS JPEG Size	-0.07	0.32	-0.23	.82
Nr. of Segments	-0.14	0.10	-1.34	.18
Spatial Frequency	0.24	0.21	1.14	.25
Näsänen Complexity	-0.14	0.22	-0.64	.52
Orientation	0.84	0.28	3.02	<.01
Feature Congestion	-1.08	0.19	-5.65	<.0001