BLANK PAGE

AMERICAN UNIVERSITY OF BEIRUT

EYE TRACKING TO ANALYZE THE EFFECTS OF CLUTTER AND STRESS ON PERFORMANCE IN REAL TIME

by
MALK KANAAN

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Engineering Management
to the Department of Industrial and Engineering Management
of the Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
December 2017

AMERICAN UNIVERSITY OF BEIRUT

EYE TRACKING TO ANALYZE THE EFFECTS OF CLUTTER AND STRESS ON PERFORMANCE IN REAL TIME

by
MALK KANAAN

Approved by:

[Signature]

Prof. Nadine Marie Moacdieh, Assistant Professor Industrial and Engineering Management Department	Advisor
	[Signature]
Prof. Hazem Hajj, Associate Professor Electrical and Computer Engineering Department	Member of Committee
	[Signature]
Prof. Saif El Qaisi, Assistant Professor Industrial and Engineering Management Department	Member of Committee

Date of thesis defense: December 12, 2017

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name:

Kanaan	Malk	Ali
Last	First	Middle

● Master's Thesis ○ Master's Project ○ Doctoral Dissertation

■ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes
after: **One --- year from the date of submission of my thesis, dissertation, or project.**
Two ---- years from the date of submission of my thesis, dissertation, or project.
Three ---- years from the date of submission of my thesis, dissertation, or project.

Malk 2/6/2018

Signature Date

AN ABSTRACT OF THE THESIS OF

Malk Kanaan for Master of Engineering Management
Major: Industrial and Engineering Management

Title: Eye Tracking to Analyze the Effects of Clutter and Stress on Performance In Real Time

E-commerce websites have come to form an integral part of people's web activity, with increasingly large numbers relying on such websites for their purchases. However, the design of e-commerce websites does not always support users in making their purchase quickly and effectively. A major design problem with such websites is clutter, or data overload. Clutter can be defined as the presence of a large amount of task-irrelevant data that leads to slower and less accurate task performance. However, it is not yet clear how best to assess clutter in websites and how to use that knowledge to develop the optimal display. In addition, clutter is not a fixed construct, and might vary among different users and in different situations. There is thus a need to develop techniques to evaluate clutter 1) in a way that reflects user factors and 2) in real time. In this way, display adjustments can be triggered immediately, before performance breakdowns occur. Eye tracking is one promising tool to do that, but it is not yet known which eye tracking metrics are best suited for real-time clutter detection in the context of e-commerce websites. The goals of this research were thus to identify what display features contribute the most to delays in e-commerce websites, determine whether and to what extent clutter and time pressure interact to bring about performance decrements, and identify the eye tracking metrics that best reflect clutter in e-commerce websites. To this end, an experiment was carried out with college students in which they were asked to search for certain targets as part of an online purchasing task. Performance and eye tracking data were collected from the experiment, and these were combined with clutter measures obtained from image processing algorithms in order to identify the eye tracking metrics that best reflect clutter. This research contributed to the literature on clutter and eye tracking and will have the potential to help improve the design of e-commerce websites.

CONTENTS

Appendix

FIGURES

Figure Page

TABLES

Table Page

CHAPTER I

INTRODUCTION

Online commercial websites, or e-commerce websites, have become an integral part of business operations for all large retail companies. As the number of people using e-commerce websites continues to increase, it is becoming increasingly important for companies to optimize the presentation of information on their websites if they want to retain the business of their customers. Studies show that more than 40% of current internet users, or around 1 billion people, have purchased a product/good online at some point in their lives, and this number is only expected to rise (Statista, 2016). Thus companies are now strongly motivated to make sure their websites provide users with the information that they need and makes the online shopping experience as smooth as possible.

However, despite the importance of e-commerce websites, there is little agreement as to the best way of displaying the information within these websites so as to maximize user satisfaction. Web users often complain about how poor website design negatively affects their experience and leads to delays in their tasks (Wauters, 2011). In particular, one major problem that is frequently encountered on e-commerce websites is clutter, or data overload. Clutter can be defined as the presence of a large amount of irrelevant data, which leads to slower and less accurate task performance (Moacdieh & Sarter, 2015). These negative performance effects are particularly prominent during search tasks for a particular target (Neider & Zelinsky, 2011), the type of task that is most commonly performed on e-commerce websites. Moreover, the effects of clutter may also be exacerbated by factors such as stress, fatigue, and time pressure (Naylor, 2010), although the interaction effects between clutter and these factors is not well known. Another complication with clutter is that it can very much depend on who is using the website: an experienced user may not feel that there is clutter in the website, whereas a novice user might. Additionally, one person may immediately find the target they are looking for, while another may struggle. So while in some cases the design of the website may be ideal for the user's needs, in other cases it may be a hindrance. However, this is something that can only be adjusted in real time, once the behavior of the user is observed. There is only very little that can be done in the initial design phase on the website. Real-time monitoring and adjustment techniques are needed as early as possible in the search process in order to adjust the display as the user is searching through it, but these techniques are still not well understood or well-developed.

There is thus a need to better understand 1) what display features lead to clutter, 2) how clutter is affected by factors such as time pressure, and 3) whether clutter can be detected early on in the search process so that display adjustments can be triggered in real time. Eye tracking is one useful tool for these purposes as it can trace the changes in attention allocation in real time, and can then form the basis for real-time clutter detection. Which eye tracking metrics are the best for detecting clutter, however, remains unclear, especially in the presence of time pressure. In particular, there is very little focus on which eye tracking metrics reflect clutter very early on in the search process and whether these can be used to develop real-time monitoring techniques for online display adjustment.

Thus the specific aims of this research were to:

1. Identify what display features contribute the most to delays in e-commerce websites

2. Determine whether and to what extent clutter and time pressure interact to bring about performance decrements
3. Identify the eye tracking metrics that best reflect clutter in e-commerce websites

This research contributes to the literature on clutter and eye tracking, and also helps inform the design of e-commerce displays. This research will help web users find what they need quickly and effectively while shopping online, and also increase sales and profits for companies. In addition, this research can generalize to other types of complex displays in different environments, helping to achieve adjustable and adaptive displays that cater to the needs to users and prevent them from making errors.

CHAPTER II

BACKGROUND

A. Display clutter

There is widespread agreement that display clutter is detrimental to the design of displays. Display clutter has been shown to degrade monitoring and signal/change detection (Schons & Wickens, 1993), delay visual search (Henderson, Chanceaux, & Smith, 2009; Neider & Zelinsky, 2011) increase memory load (Westerbeek & Maes, 2011), instill confidence in wrong judgments (Baldassi, Megna, & Burr, 2006), lead to confusion (Ewing, Woodruff, & Vickers, 2006), and negatively affect situation awareness (S.-H. Kim & Kaber, 2009), object recognition (Bravo & Farid, 2006), reading (Phillips & Noyes, 1982) and linguistic processing (Coco & Keller, 2009).

In particular, clutter has been known to negatively affect user's performance on websites. Clutter has been shown to increase the time to find targets on website pages (Grahame et al., 2004) and also to lower the total amount that users purchase online (e.g., Galletta et al. 2006). Clutter in e-commerce websites could stem from the density of the objects themselves or from the moving elements and advertisements that are typically added, which users often find distracting and disruptive (Spiekermann & Korunovska, 2014). Consumers' irritation by such advertisements can also lead consumers to avoid visiting the same website again (Edwards et al. 2002).

1. Defining display clutter

Despite the concern about the negative effects of clutter, there is no consensus on a definition of the phenomenon or the factors that cause it. This disagreement stems, in large part, from the fact that many different disciplines have studied the problem of clutter. Researchers in wide-ranging fields have tackled the issue, including aviation (Kaber et al., 2013), radar (e.g., McKenzie, Wong, & Gibbins, 2013), and healthcare (e.g., Hammond et al., 2012), marketing (Rotfeld, 2006), and, namely, website design (Grahame et al., 2004). This has led to a variety of definitions about what clutter is.

However, most of the present literature relate display clutter to the amount of data in a display. In this case, clutter is viewed as the presence of a large quantity of display items (e.g., Clay, 1993; Horrey & Wickens, 2004; Kroft & Wickens, 2002; Mack & Oliva, 2004; Tufte, 1983; Tullis, 1983; Ververs & Wickens, 1998). More specifically, clutter can be better defined as the density, not quantity, of display items – "too much data on too small an area" (Ellis & Dix, 2007; p.1216) – that results in clutter (Coco & Keller, 2009; van den Berg et al., 2009). The real nature of the "items" on any given display varies widely, depending on the application domain in question. For example, the objects may be words or graphics on a webpage (Grahame et al., 2004), symbols or marks on a map (e.g., Lohrenz, Trafton, Beck, & Gendron, 2009; Yeh & Wickens, 2001), sensor readings in an infrared display (Wang & Zhang, 2011), icons on an airplane cockpit display (S.-H. Kim & Kaber, 2009), words in an electronic health record (Hammond et al., 2012), or the point sources in a radar display (McKenzie, Wong, & Gibbins, 2013.

The view of clutter as display density is also somewhat oversimplified and incomplete; increasing the number or density of objects may increase clutter, but other factors – such as the good organization of these items – may mitigate the performance effects of density (Doyon-Poulin et al., 2012). The second display-based dimension of clutter – organization – emphasizes this point. Here, the poor organization of display entities, not merely their number or density, is highlighted (for a review of display organization, see Nielsen (1993), Shneiderman (1998), or Wickens, Hollands, Banbury, and Parasuraman (2013)). This perspective has been reflected in numerous clutter definitions (e.g., Bravo & Farid, 2008; Doyon-Poulin et al., 2012; Rosenholtz, Li, Mansfield, & Jin, 2005; Tufte, 1991; van den Berg et al., 2009). For example, Peng, Ward, and Rundensteiner (2004; p.89) refer to clutter as “a disordered collection of graphical entities” or “the opposite of structure”. Factors that fall under the broad category of display organization include the lack of logical or conceptual grouping of items (Wickens & Carswell, 1995), the absence of symmetry (Oliva, Mack, Shrestha, & Peeper, 2004), the degree to which the target is obscured or masked (Alexander, Stelzer, Kim, & Kaber, 2008; Bravo & Farid, 2004; Chu, Yang, & Li, 2012; Toet, 2010; Xing, 2007), the presence of high entropy (or lack of predictability) within a display (Rosenholtz, Li, & Nakano, 2007), and the variability in color, luminance, contrast, and orientation in a display (Lohrenz et al., 2009; Rosenholtz et al., 2005). All of these factors contribute to poor guidance to the target, which, in turn, leads to poor search performance.

A third factor when it comes to clutter is crowding, or the close spacing between a target and surrounding distractors. Crowding is related to local, as opposed to global, clutter (e.g., Beck, Lohrenz, & Trafton, 2010; Ewing et al., 2006). While global clutter refers to the amount or organization of information across an entire display, local clutter focuses on the amount and organization of information surrounding an important small area which includes the target.

Target-background or target-distractor similarity is the fourth major display-based dimension of clutter. Proponents of this perspective view clutter as the degree of similarity between characteristics of the background, such as color, and those of the target (Christ, 1975; Duncan & Humphreys, 1989; Liao, Wu, & Sheu, 2013; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). This perspective of clutter features prominently in the radar and Automatic Target Recognition (ATR) domains, where the main concern is to the detection of a target against background noise. In these domains, clutter has been defined as "objects or features in the scene that are similar to the desired target" (Chu et al., 2012; p. 067003-1), the "quantity of background signatures similar to the target" (H. Camp, Moyer, & Moore, 2010; p. 76620A-1), or the "signal caused by the background objects resembling a target. The more target-like objects or attributes in the background, the higher the clutter level" (He, Zhang, Liu, & Chang, 2008; p.5534).

The above views of clutter are all based on features of the display, independent of the user and task. For example, from a display density perspective, an additional item, whether task relevant or not, will still contribute to clutter in the same way. However, the fifth aspect of clutter is different in that it focuses on the task relevance of display objects, which can have considerable bearing on whether they negatively affect performance (Alexander et al., 2008; Barbu, Lohrenz, & Layne, 2006; Doyon-Poulin et al., 2012; Horrey & Wickens, 2004; Lohrenz, Layne, Edwards, Gendron, & Bradley, 2006; Rosenholtz et al., 2007). User-based factors such as workload (Naylor, 2010) may also influence the perception and effects of clutter. Consider the example of an operator using a complex display in a power plant. At any given time, a number of task irrelevant items may be present on this display but, during routine operations, this does not necessarily result in adverse performance effects for an experienced pilot who imposes structure on the display in a top-down fashion. However, during high stress and time pressure, these irrelevant elements may distract or slow down search for critical information. These situations of high time pressure, where a user wants to find something as quickly as possible, was investigated in this research.

These five aspects of clutter are all important when it comes to defining what clutter is; nevertheless, the most prominent and most important aspect remains the data density aspect. This aspect was the one emphasized in this research. Moreover, from a human factors standpoint, what is important is that these factors lead to performance and attentional decrements. Otherwise, one cannot assume that a display is “cluttered”. Highlighting this point, clutter in this study was defined as “the presence of performance and attentional costs that result from the interaction between high data density, poor display organization, and an abundance of irrelevant information” (Moacdieh & Sarter, 2014; p. 65). Thus the focus is on the performance and attentional effects that result from certain display features, and not the display features alone.

2. Measuring display clutter

Having defined clutter, the next step is to determine a means of reliably measuring the phenomenon in websites. Several different approaches can be used to this end, with the resulting wide range of techniques reflecting the widespread disagreement about what clutter is. The most commonly used techniques are image processing, performance evaluation, and subjective evaluation.

In the case of image processing algorithms, clutter is calculated based on pixel-based display characteristics, such as luminosity or contrast (e.g., Bravo & Farid, 2008; Chang, Zhang, Liu, Yang, & Li, 2010; Lohrenz, Trafton, Beck, & Gendron, 2009; van den Berg et al., 2009). This makes it possible to determine clutter from a display-based perspective only, and thus image processing was used in this research as a way of assessing data density. However, image processing techniques do not give any consideration to the performance effects that may result from display features, an aspect that we are highly interested in. For example, Rosenholtz, Li, and Nakano (2007) developed a measure of clutter based on using three different metrics: feature congestion, edge density, and subband entropy. Feature congestion refers to the idea that the more cluttered a display is, the more difficult it would be to add an item that would be able to capture attention. First, clutter maps are developed separately for color, texture, and orientation, and then they are combined into a single clutter map and provide a scalar measure of clutter. Edge density is based on the idea by Mack and Oliva (2004) that clutter is related to the number of edges in a display. In this case, a filter calculates the density of edge pixels as a percentage of the total number of pixels to give a measure of clutter. Finally, the subband entropy method creates a clutter map and a scalar measure of clutter based on the assumption that clutter is inversely proportional to the amount of redundancy in an image. This measure was later found to be highly related to clutter around a target or local clutter (Asher, Tolhurst, Troscianko, & Gilchrist, 2013). Together, the three techniques provide an overall estimate of display clutter. Other image processing techniques for clutter include the threshold and cluster methods of Olsson, Pippig, Harrie, & Stigmar (2011). In the threshold method, the authors divided maps into grid cells and then computed, for each cell, parameters such as object line length, number of vertices, and number of object types. In the cluster method, on the other hand, they used a density-based algorithm to identify clusters of points (which could be polygon objects or line vertices). The clusters that contained more than a certain number of points were considered to be higher in display clutter, and this method proved to be more computationally efficient. Another measure that has been used for website clutter evaluation consists of dividing the number of pixels used to display information in webpages by the total available pixels (Grahame et al., 2004). In this research, several image processing techniques were used to obtain an initial and completely objective impression of the amount of clutter in a given website.

A second approach to measuring clutter is performance evaluation. In other words, this approach relies on comparing people's performance using displays that have (supposedly) different levels of clutter (e.g., Beck et al., 2010; Wickens, Nunes, Alexander & Steelman, 2005; Yamani & McCarley, 2011; Yeh, Brandenburg, Wickens, & Merlo, 2000). Performance measures that have been calculated include search time and error rate (Beck et al., 2012; Grahame et al., 2004), as well as a number of domain-specific measures. For example, in aviation research, Kaber et al. (2013) measured clutter in a flight simulator based on pilots' tracking error. In the medical domain, Zeng, Cimino, and Zou (2002) asked physicians to retrieve specific patient information and calculated the search time and accuracy of the responses to determine the level of clutter in medical displays. This approach does not give very detailed insight into what is wrong with a display, but is more in keeping with the second part of our definition of clutter as related to performance decrements. In this research study, participants' performance on tasks involving ecommerce websites were collected.

Finally, subjective assessment requires users to rate or rank the amount of clutter perceived in displays, thus relying on people's judgment to determine whether a display is cluttered (e.g., Alexander, Stelzer, Kim, & Kaber, 2008; Kaufmann & Kaber, 2010; Kim & Sundar, 2010). This approach has been used extensively in the literature as well in domains such as aviation (Alexander et al., 2012; Kaufmann & Kaber, 2010; S.-H.

Kim et al., 2011; McCrobie, 2000) as well as web-pages (Ling, Lopez, & Shehab, 2011; Michailidou, Harper, & Bechhofer, 2008). In general, the approaches to eliciting clutter ratings can be divided into clutter rating and clutter ranking. In clutter rating, participants are asked to provide a subjective estimate of the amount of perceived clutter in a display, such as on a scale from 1 to 10 (Lohrenz et al., 2009) or a 7-point Likert scale (Arthur et al., 2005; Bailey, Kramer, & Prinzel, 2006).

However, none of these techniques enable us to achieve the goal of this study, which is to detect website clutter in real time. Image processing techniques can be used in real time but do not really provide any input about the state of the user or of performance effects. Performance and subjective evaluation are more useful in that regard; however, they cannot be used in real time. Another, different measure of clutter is needed.

B. Eye Tracking

Recording eye movements, or the location of gaze, using an eye tracker appears to be a very promising approach for real-time detection of clutter. Eye movements can be obtained using an eye tracker, a device that uses infrared light to trace where people are looking at on a display (see Duchowski (2007) or Poole and Ball (2006) for a detailed review). Eye trackers can be desktop-mounted, meaning that they are placed by the display to be tracked, or head-mounted, in which case they are worn by the person either as a headpiece or as a device similar to glasses. In both cases, eye trackers output raw eye location data known as points of regard (POR) that indicate where a person is looking in the display. PORs are typically recorded at 50 Hz or higher and are expressed in x and y coordinates (Munn, Stefano, & Pelz, 2008). In turn, the PORs can be used to determine the two basic components of eye tracking research, fixations and saccades. Fixations, which are characterized by location and duration, are formed from spatially stable PORs and it is during this time that visual processing takes place (Findlay, 2004; see Figure 1). The rapid eye movements between successive fixations are called saccades, during which time visual processing is usually suppressed (Yarbus, Haigh, & Riggs, 1967). A scanpath is the path of a sequence of fixations and saccades, and it provides a means to visualize eye movements (Noton & Stark, 1971). Finally, an area of interest (AOI) is an experimenter-defined region of the display on which analysis of eye tracking data is performed. Together, these components have been used as the building blocks for eye tracking research.

There are several techniques that can be used to identify fixations and saccades from the raw gaze points (Salvucci & Goldberg, 2000). In this research, a set of consecutive gaze points constituted a fixation if there were at least six gaze points and they were within a two-degree visual angle radius (Goldberg & Kotbal, 1999). With an eye tracker that sampled at 60 Hz, this meant that the minimum duration for fixations was 100 ms. This method is classified as a dispersion algorithm by Salvucci and Goldberg (2000). Any other gaze points that were not classified as fixations were grouped together as saccades, meaning that it was not possible to detect any smaller microsaccades that may have been present. However, for the purposes of this research, these smaller saccades can be considered negligible and do not have any significant influence on results.

As for challenges related to using eye tracking, these include the high cost and setup/analysis time (Jacob & Karn, 2003) and the correct selection of parameters for fixation calculation (Inhoff & Radach, 1998). However, the benefits of using eye tracking to avoid delays and misses in data-rich, safety-critical domains outweigh these issues, and hardware developments in the future may alleviate many of these problems (Pavlas, Lum, & Salas, 2012).

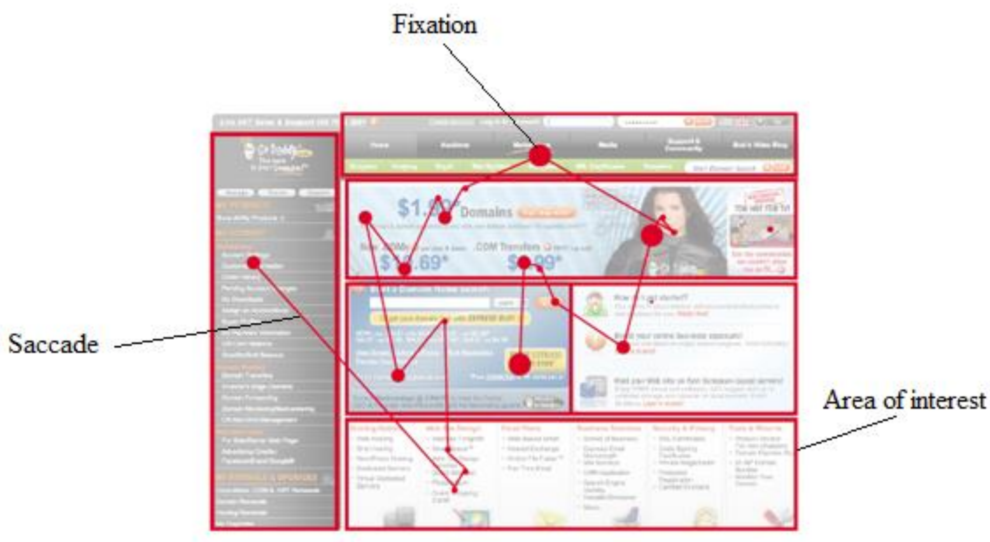


Figure 1. Fixations are usually depicted as a circle whose diameter is proportional to fixation duration. Saccades are represented as lines between two successive fixations, while areas of interest are typically drawn as rectangles. All of the fixations and saccades together create the scanpath (adapted from Bonigala (2009))

Eye tracking has been used extensively in human factors research, including studies in aviation (e.g., Alexander & Wickens, 2005; Ellis, Kramer, Shelton, Arthur, & Prinzel, 2011; Schnell, Kwon, Merchant, & Etherington, 2004), driving (Di Stasi, Contreras, Candido, & Cantena, 2011; Liang , Reyes, & Lee, 2007), website design (Katsanos, Tselios, & Avouris, 2010; DeWitt, 2010), and medicine (Chetwood et al., 2012; Marquard, Jo, Henneman, Fisher, & Henneman, 2012). There are several benefits to using eye tracking that have encouraged its use in research, as described below.

1. Benefits of using eye tracking for clutter research

The use of eye tracking is based on the assumption that the location of a person's gaze can be considered a stand-in for the locus of attention, a theory known as the eye-mind hypothesis (Just & Carpenter, 1978). Some researchers (e.g., Anderson, Bothell, & Douglass, 2004) claim that there could be a dissociation or lag between attention and the location of gaze. Others propose that attention can sample different areas of the display at a higher rate than eye movements (e.g., Horowitz & Wolfe, 2003). However, for most human factors applications, eye tracking can still provide a relatively good estimate of the location of attention (Goldberg & Wichansky, 2003; Rayner, 1998; Zelinsky, 2008).

In particular, the use of eye tracking has proven invaluable in the study of attention and visual search (e.g., Findlay & Gilchrist, 2005; Trukenbrod & Engbert, 2007; Williams & Pollatsek, 2007). At the most basic level, the correlation between the number of saccades and search time has been used to prove that eye movements are necessary for visual search (e.g., Zelinsky & Sheinberg, 1997). In addition, eye tracking has been used to test and validate models of eye movements during visual search (e.g., Zelinksy, 2008) and to create salience maps (maps that indicate the regions of the screen most likely to attract attention; e.g., Chukoskie, Snider, Mozer, Krauzlis, & Sejnowski, 2013; Henderson, Brockmole, Castelhano, & Mack, 2007; Judd, Ehinger, Durand, & Torralba, 2009; Parkhurst, Law, & Niebur, 2002). The salience of abrupt visual onsets has also been verified using eye tracking, with saccades to suddenly appearing distractors observed during visual search for a target (e.g., Theeuwes et al., 1998). Other eye tracking experiments have examined the concept of inhibition of return (Klein, 2000) or the tendency to avoid already visited areas of a display (e.g., Beck, Peterson, Boot, Vomela, & Kramer, 2006; Geyer, von Muhlenen, & Muller, 2007). The above studies show that, in contrast to performance outcome measures such as response time, an eye tracker is a process-oriented tool that can trace the shifts and degradations of attention at a fine-grained level of analysis. In addition, eye tracking data can reflect user- and task-based influences on attention control, such as experience (Beck et al., 2012; Konstantopoulos, Chapman, & Crundall, 2010).

To a limited extent, researchers have used eye tracking to explore the effects of clutter on attention and performance (e.g., Beck et al., 2010; Beck et al., 2012; Grahame, Laberge, & Scialfa, 2004; Neider & Zelinsky, 2011; Zhu & Sun, 2012). For example, research on large set size has led to the conclusion that the higher or more dense the number of objects on a display, the slower the search as the number of fixations and fixation duration increase, while the mean saccade amplitude decreases (e.g., Bertera & Rayner, 2000; Greene & Rayner, 2001; Vlaskamp & Hooge, 2006).

These types of experiments all used very simple, artificial stimuli, such as letters and shapes. Researchers in human factors and other applied disciplines have attempted to build on such studies to determine whether they could be generalized to more complex displays. For example, some studies showed a significant increase in the number of fixations in more cluttered aeronautical charts and websites (e.g., Beck et al., 2010, 2012; Grahame et al., 2004). Calculating the number of fixations on added objects or distractors (Hegarty, De Leeuw, & Bonura, 2008; Beck et al., 2010) as well as the amount of time spent on task-relevant items (Fabrikant, Hespana, & Hegarty, 2010) has also been used to determine the effects of clutter. For example, longer fixation times and thus lower fixation rates have been linked to higher levels of clutter (Beck et al., 2010; Henderson et al., 2009; Zhu & Sun, 2012), as have the latency of the first saccade (Henderson et al., 2009; Zhu & Sun, 2012; Zelinsky, 2001), scanpath length (sum of saccade amplitudes; Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002), scanpath ratio (scanpath length divided by the length of the shortest path from the starting point to the target (Neider & Zelinsky, 2011)), and final saccade amplitude (length of the last saccade before target detection can give an indication of how easily noticeable the target is in peripheral vision (e.g., Henderson, Weeks, & Hollingsworth, 1999)). Eye tracking has also been used to study users' search strategies, although there is no firm agreement on the matter. In one case, a "coarse-to-fine" search strategy was observed where participants tried to quickly extract as much as they could from the display before resorting to slower, more deliberate search (Beck et al., 2012). Henderson et al. (2009), on the other hand, found that users tend to search in areas of high clutter first, which may be because these were regions of high salience as well.

Within the context of e-commerce websites, eye tracking has been used to analyze the relationship between target location and efficiency in finding the target in online shops, online newspapers, and company webpages (Roth, Tuch, Mekler, & Bargas-Avila, 2013). Participants were asked to spot a target web object as fast as possible on a webpage. The results showed that usual web object placement led to less fixations and faster target detection. More specifically, Cowen, Ball, and Delin (2002) found that average fixation duration, number of fixations, spatial density of fixations, and total fixation duration best reflected the effects of different pages and tasks. Bruneau, Sasse, and McCarthy (2002) studied the use of eye tracking metrics on different types of websites and emphasized the need to tailor the metrics used to the type of question under study. In other studies, eye tracking helped show which elements of websites were most looked at by users. For instance, recommender systems tended to grab people's attention as much as description boxes (Castagnos & Pu, 2010) and complexity of the website also influenced how people scan (Pan et al., 2004). Goldberg (2010) also linked eye tracking metrics to webpage complexity. Finally, Ehmke and Wilson (2007) made a list of possible metrics that can be used for webpage usability, dividing them into those that are fixation-related, saccade-related, and scanpath-related. While comprehensive lists of this kind are available, as are numerous studies linking eye tracking, website design, and website usability, to our knowledge nobody has attempted to combine these different eye tracking metrics in real time to study the specific problem of clutter in e-commerce websites.

Eye tracking also provides a number of advantages as a sensing mechanism in a real-time adaptive display. First and most importantly, eye tracking data can be obtained in real time. Such data has been used to detect driver distraction (e.g., Liang et al., 2007) and sleepiness (Jin et al., 2013), evaluate user learning (Kardan & Conati, 2012), measure workload (Durkee, Geyer, Pappada, Ortiz, & Galster, 2013), select areas of a display (Kumar, Paepcke, & Winograd, 2007), and provide support to non-native English speakers by displaying the meanings of difficult words (Hyrskykari, 2006). Desktop-mounted eye trackers are also completely non-invasive, in contrast to many other sensing mechanisms, such as EEG.

2. Eye tracking metrics for detection of clutter

Despite the promise of eye tracking and the studies conducted to date, there are numerous gaps in the literature on clutter and eye tracking that need to be resolved before an eye-tracking based adaptive display can be developed. It is not clear how best to trace and detect clutter in real time. In particular, this has never been done for websites before, and there is thus a need to determine 1) which metrics can best reflect the effects of clutter in real time, and 2) how a model of these eye metrics should be developed to quickly and reliably detect the presence of clutter.

The metrics shown in Table 1 have been proposed as metrics for clutter detection by Moacdieh and Sarter (2015) and have been shown to be promising in the context of electronic medical records. Spread metrics depend only on fixation coordinates; they show whether clutter causes a dispersion of eye movements across the display, thus preventing the user from focusing on important information. Increased spread suggests increased coverage of sampling of different areas of the display, which could occur with a large amount of irrelevant data and poor guidance to the target. Directness measures differ in that the sequence of fixations is taken into account; these measures can indicate whether clutter made search less ordered or systematic. Directness measures help show how efficiently users reached the target destination, which, in turn, can provide insight into whether there was strong guidance to the target or whether elements of the display were distracting. Finally, duration measures indicate how long a person looked at a particular area and relate clutter primarily to the difficulty extracting information from the display or the perceived importance of the information. In addition, these metrics have been found to be able to detect clutter in real time in the context of simple simulated graphics programs (Moacdieh & Sarter, under review). This research tested whether and which of these metrics can be used to detect clutter in real time in the context of websites.

Table 1. A list of the Eye Tracking metrics used in this study

Name	*Explanation*
Spread Metrics	
Convex Hull area (pixels^2) (Moacdieh & Sarter, 2015)	*Minimum convex which contains the fixation points*
Spatial density (Moacdieh & Sarter, 2015)	*Number of Grid cells containing gaze points divided by the total number of cells*

Nearest neighbor index (Di Nocera et al., 2007)	*The ratio between (1) the average of the observed minimum distance between points and (2) the mean random distance expected if the distribution were random*
Directness Metrics	
Scanpath length (pixels) (Moacdieh & Sarter, 2015)	*The sum of the lengths of all saccades in a defined period*
Scanpath length per second (pixles/sec) (Moacdieh & Sarter, 2015)	*The scanpath length divided by the time*
Backtrack rate (/sec) (Moacdieh & Sarter, 2015)	*The backtrack is defined as an angle between two saccades that is greater than 90^o*
Rate of transitions (/sec) (Moacdieh & Sarter, 2015)	*Rate of transitions between equal grid cells*
Duration Metrics	
Mean fixation duration (sec) (Moacdieh & Sarter, 2012)	*Mean duration of all fixations within a defined period*

3. Real-time detection of clutter

In order to detect clutter in real time, a model must be adopted that will use the eye tracking metrics as input and output a result that indicates whether or not clutter is detected. This can then lead to the creation of an adaptive display of information, where the system is responsible for deciding when and how to change the display (Inagaki, 2003; Kaber, Wright, Prinzel, & Clamann, 2005; Keeble & Macredie, 2000). For example, in the study by Cummings, Brzezinski, and Lee (2007) on the design of an unmanned aerial vehicle control display, an intelligent algorithm identified upcoming high-workload situations and highlighted these conditions for operators to alert them when several critical tasks were predicted to occur at the same. Adaptive systems have been shown to be less acceptable to users who perceive a loss of control (Shneiderman, 1997; Wickens, 1994). However, adaptive systems present a major advantage in that they do not add to a user's workload during critical times (Hou, Kobierski, & Brown, 2007; Parasuraman et al., 1998). For example, Bailey et al. (2006) found that mental workload decreased with an adaptive, as opposed to an adaptable approach to a failure detection task. This advantage of adaptive systems seems particularly desirable in the time-pressured situations that are the focus of this research, where operators' workload levels are usually already high. Support for this notion is provided by a flight simulator study by Olson and Sarter (2000) which showed that pilots generally preferred to have control over flight deck automation (i.e., an adaptable design); however, in high-stress cases, involving time pressure and high workload, they preferred to have the automation make critical decisions while still retaining the right to make adjustments afterwards (an adaptive approach).

The adaptive approach to context-sensitive design was thus adopted in this study. In order to implement an adaptive display system, three components or mechanisms are needed:

1. A sensing mechanism that measures or traces user and/or environmental conditions in real time

2. A control algorithm that decides when intervention or change is needed in the display and what that change should be

The sensing mechanism in this study was eye tracking, for the numerous reasons detailed earlier. The second important component of any adaptive system is a means of determining when adaptations or changes are needed. This subsystem analyzes the data collected by the sensing mechanism (in this case, eye tracking data) and determines whether the system and/or user is currently in a desirable (i.e., no change needed) or non-desirable (i.e., changes needed) state. At the simplest level, researchers can compare the different measures collected in real time and evaluate them or compare them to some threshold (e.g., Prinzel et al., 2000). For example, Pope et al. (1995) determined whether manual or automatic mode was needed based on whether individual EEG values that indicated task engagement were increasing or decreasing. Alternatively, a modeling or machine learning approach can be used in order to determine the state of the system based on combinations of the input features. Techniques that have been used include artificial neural networks (e.g., Russell, 2005), genetic programming (Bergstrom et al., 2000), Naïve Bayes (Mokhtar, Abdullah, & Zin, 2011), support vector machines (SVM; e.g., Liang et al., 2004), and logistic regression (Barr et al., 2008; Ratwani, McCurry, & Trafton, 2008; Steichen, Garenini, & Conati, 2013).

Which one of these techniques is preferable to use is still open to debate. Evidence from studies that have compared multiple techniques is inconclusive. For example, Liang et al., (2004) compared SVM and logistic regression for detecting driver distraction based on eye movements, and found that SVM outperformed logistic regression. On the other hand, Steichen et al. (2013) tested SVM, decision trees, multilayer perception, and logistic regression and found that logistic regression had the highest accuracy among all. Several researchers have pointed out that logistic regression is a good option for real-time analysis given that it is light-weight and efficient for online processing, as well as the fact that it provides insight into the importance of different features (Kozma, Klami, & Kaski, 2009). In this case, it would help identify the best eye movement metrics to use. Other advantages of logistic regression include its robustness, which would be particularly useful in this research where unequal sample sizes and a small data set size are expected (Blom, Paradis, & Duncan, 2012).

Another possible approach would be the use of deep neural networks, a type of artificial neural network that automatically analyze and extract features from raw data (Huang, Shen, Boix, & Zhao, 2015). Deep neural networks help to group unlabeled data according to similarities among the example inputs, and then classify data based on a labeled data set. This approach has been successfully applied to speech recognition (Seide et al., 2011) and image retrieval (Krizhevsky & Hinton, 2011). Moreover, the ability of deep neural networks to automatically learn complex patterns from data in a hierarchical fashion makes them applicable to a wide range of problems with different modalities of data (Tang, Lu, Wang, Huang, & Li, 2015). In particular, deep neural networks have been used successfully in the context of eye tracking in order to determine the saliency of display items, for example (Kruthiventi, Ayush, & Babu, 2015). Logistic regression, SVM, and deep neural networks were investigated in this research to determine which is the best approach to use in the context of websites.

CHAPTER III

METHODS

A. Participants

The participants were 50 students from the American University of Beirut who were between the ages of 18 and 30. All participants were expected to have self-reported normal or corrected to normal vision and were not color blind. The rationale for the selection of this population as participants was that they had the necessary computer skills for and were familiar with e-commerce websites. Participants were recruited using flyers posted around campus and emails sent to a random sample of students.

B. Stimuli

The research stimuli consisted of 50 screenshots of existing e-commerce websites (see Figure 2.a. as an example). These websites deliberately included a range of applications, such as purchasing electronic devices, clothes, and home appliances. For each website, there was a corresponding search target. For example, Figure 2.b. shows an image search target.

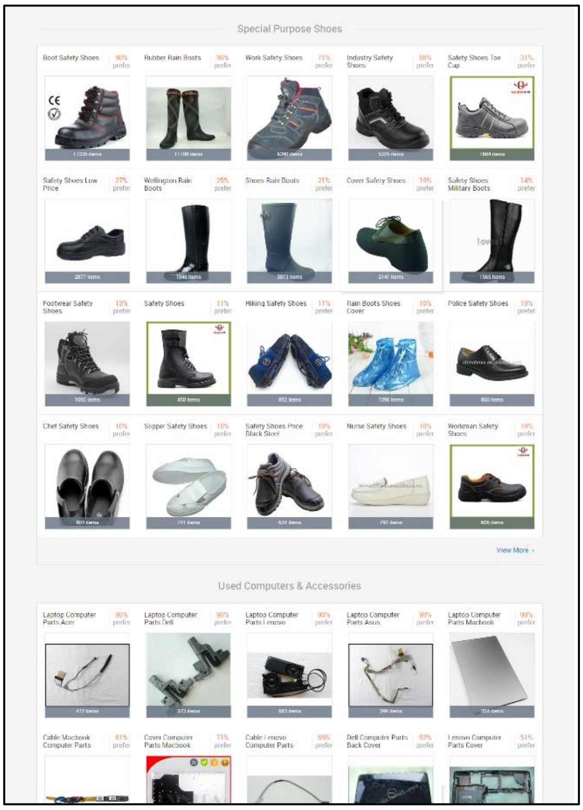


Figure 2.a. An example of a website screenshot

Figure 2.b. An example of an image search target

The search targets, whether images or texts, were all segments from (i.e., parts of) the website screenshot itself. The image search targets were all part of the image, whereas the word search targets were words that were found in the website (not in an image, though). The word search target always consisted of a single word. In addition, the search target font size (for words) and search target image size (for images) were also all the same.

C. Experiment setup

The location of this experiment was conducted in the Scientific Research Building, room 407, at the American University of Beirut. A Tobii X3-120 eye tracker was used for the experiment (see Figure 3). This eye tracker was infrared-based and was attached to the bottom of the monitor. There was nothing that was in contact with the participant at any time; this eye tracker was completely non-invasive. The sampling rate for this eye tracker was 120 Hz.

The eye tracker tracked the participants' eye movements on a monitor that of size 24 inches. The participants were seated at a distance of around 60 cm from the screen and were instructed to keep their head steady so as to maximize the quality of the eye tracking results. The calibration procedure was accomplished using a 9-point grid. This calibration enabled the eye tracker to "learn" the characteristics of the subject's eyes so that the direction of his/her gaze on the surface of the screen or object was accurately calculated.

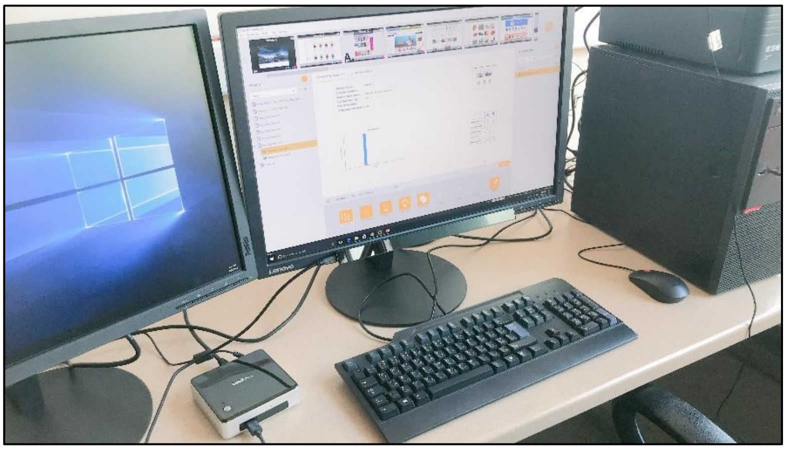


Figure 3. Tobii Pro X3-120

D. Experiment Design

The independent variables in this research study were the level of clutter (low, high) and the presence or absence of a time limit (time limit, no time limit). The latter simulated the presence or absence of stress. In addition, there was a blocking variable, which was the type of search target. As explained earlier, search targets were either images or words, which represented the two main types of targets that users search for in an e-commerce website. The reason this was treated as a blocking variable was because we were not interested in differences between images and words; rather, we were interested in analyzing, for each of these types of targets, what the effects of clutter and stress were.

The level of clutter for each website screenshot was assigned according to the results of a well-known image processing algorithms, those of Rosenholtz et al. (2007), as described in the background section. These algorithms produced a numerical measure of clutter that was based on three metrics combined: feature congestion, subband entropy, and edge density. The algorithms were applied to each website image using Matlab. Following the application of all the image processing algorithms, the scores obtained were normalized then aggregated and used to classify the images into 20 low-clutter and 20 high-clutter images. The clutter results are in table 2, in which no statistical difference between the two sets was found.

Table 2. Mean clutter values for websites used in this study

One-Way ANOVA	**Words**	**Images**
Clutter measure	(0.41 ± 0.27) p = 0.316	(0.34 ± 0.16) p =0.316

As for the time limit, the presence of a time limit meant that participants had to complete their given task within this amount of time. In the absence of a time limit, participants had as much time as they need to complete their task. The actual time limit used in the experiment was selected to be 10 seconds based on initial tests.

Participants were asked to perform search tasks using the website screenshots. Thus each experiment trial consisted of one search task on one given website: a search for an image target or a search for a word target. Both of these search tasks were representative of the types of search that could be performed on an e-commerce website. In the case of an image, a user might be looking for a certain feature of a product. On the other hand, there may be a particular word, such as the price or brand of a product that the user may be looking for as well. Each trial was associated with one of the website screenshots with its corresponding target. Thus participants went through each website, but they never looked for the same word or image twice. Since these websites were only screenshots, participants searched through them but couldn't click on any links. Participants had to click on the location of the target to indicate that they had found it.

The screen eccentricity of all the search targets (i.e., the distance, in visual angles, of the target from the center of the screen) across all images were all the same so as to ensure that search complexity was equivalent across scenarios. In order to prevent participants from noticing the fact that all targets were at the same eccentricity, there was a dummy search target (and corresponding dummy trial) associated with a random website. This dummy target had a different eccentricity. Thus for each set of five website screenshots, there was one dummy target: 4 that were associated with actual experiment trials, and one that was associated with a dummy trial (see Table 2). These were randomly assigned to be images or words. The data from the dummy trials were not considered in the data analysis.

Table 3. Example of the search targets for timed websites

Website Screenshots	Search targets associated with this screenshot	Clutter level (low/high)	Type (image/word)	Time pressure (presence/absence)
Website 1	Search target 1	Low	Image	Present
Website 2	Search target 2	High	Image	Present
Website 3	Search target 3	Low	Word	Present
Website 4	Search target 4	High	Word	Present
Website 5	Dummy target	Low or High	Image or word	Present

In total, each participant had 50 trials to complete (10x4=40 experiment trials and 10x1=10 dummy trials). The order of presentation of the variables was counterbalanced by dividing participants into four groups. Each group performed the tasks in a fixed order such that they alternated between screenshots that were high and low in clutter. The presence or absence of a time limit was varied after twenty-five tasks; in other words, participants did the first twenty-five trials all with or without a time limit, and then the second set with the inverse time pressure condition. Within each group of 25 trials, there were 5 dummy trials and 20 experiment trials.

The dependent variables that were recorded were response time, error rate, the eye tracking metrics of Table 1, and subjective data. Response time was calculated from when participants first saw a website screenshot until they clicked on the location of the target. In the case of a time limit trial, if the trial timed out before the participant found the target, the response time was not considered. In addition, if the participant gave up and decided not to continue a trial, the response time was not recorded.

An error trial was considered in three cases: if the participant pressed on the wrong target, if the trial timed out (in the case of a time limit trial), or if the participant gave up. For the eye tracking metrics, all of the ones mentioned in Table 1 were calculated for each participant. These were calculated for each trial from the start of the search task until the participant clicked on the location of the target or the trial timed out.

As for the subjective data, this was collected by means of two separate online questionnaires. The first, which can be seen in Appendix A, was a modified NASA Task Load Index (TLX) questionnaire (Hart & Staveland, 1988). This questionnaire assessed time pressure and performance and was administered at the end of each set of 25 trials. The questionnaire was completed online by the participant. The second questionnaire, the post-experiment questionnaire (see Appendix B), was only administered once, at the end of the experiment. This questionnaire gathered data about the participant, such as his/her age, major, year at AUB, and experience with e-commerce websites. This questionnaire also asked participants to rate their impression of clutter in all of the screenshots that they went through. This questionnaire was filled out online while the screenshots were displayed on the screen.

E. Experiment Procedure

When participants first came into the lab, they were asked to read and sign a consent form. They were then briefed about the main purpose of the experiment and what they will have to do during the experiment. Next, participants were asked to complete a set of training tasks to make sure that they understand the experiment procedure. Participants needed to complete four out of four training tasks correctly to be able to proceed to the actual experiment. They could repeat the tasks as often as necessary in order to achieve compliance. Two of the tasks had a time limit and two didn't, and two of the tasks involved word searches and the other two involved image searches. The training phase took around 5 minutes.

After completing the training phase, the eye tracker was calibrated. Participants whose eyes couldn't be calibrated were not allowed to continue with the experiment. Calibration was expected to take around 2 minutes. Participants then proceeded with the actual experiment. They did 25 of the 50 trials, with a 5-minute break in between. Each set of 25 trials were either timed trials or non-timed trials, and were followed by the participant filling out a NASA-TLX questionnaire. The order of the timed or not timed trials were counterbalanced across participants. The experimenter was seated to the side of the participant at all times to make sure that their eyes remain within range.

Before each website image was presented, the screen showed the target; either word or image. Participants looked at it as long as necessary and then pressed the right arrow key to proceed with the search trial. Depending on whether it was a trial with a time limit or without, participants had either a set amount of time or unlimited time to complete the search task. They needed to use the mouse to click on the location of the target to indicate that they had found it. The experiment trials, together with the NASA-TLX questionnaire, were expected to take around 30 minutes. After completing all trials, the post-experiment questionnaire was administered, which took approximately 5 minutes. The full experiment was thus expected to take around 40-50 minutes.

CHAPTER IV

RESULTS

The analysis of the results of this experiment took two forms. First, the dependent measures were analyzed using a two-way repeated measures analysis of variance (ANOVA) in order to study the main and interaction effects of clutter and time pressure (represented by a time limit). All of the performance, eye tracking, and subjective data were averaged over the trials of the same experiment condition and then analyzed across the levels of the independent variables. This was done separately for images and for words. The analysis for the performance and eye tracking data was done using the IBM SPSS package. The assumptions for the repeated measures ANOVA procedure were tested before running the ANOVA procedure, i.e., normality was checked using Shapiro-Wilk test for normality. If the results were not normal, transformation of data was applied using either logarithmic or inverse transformation before running ANOVA again. Outliers were also checked for using the studentized residuals. If outliers were present, the analysis was done with and without them. They were removed if they had significant effects on the results. The Wilcoxon Test was used for the ordinal data obtained from the subjective questionnaire. In addition, response time was correlated to the image processing algorithm results.

Unless otherwise specified, results were analyzed using a 2 × 2 repeated-measures analysis of variance (ANOVA), with Bonferroni corrections applied for multiple statistical tests. Significance was set at $p < .05$, and partial eta-squared (η_p^2) was used as a measure of effect size. The ANOVA results are reported for statistically significant results only, with some descriptive values highlighting notable trends. Error bars on graphs indicate the standard error of the mean (SEM).

A. Subjective Results

The subjective clutter ratings were analyzed using Spearman Correlation, in which response time was correlated with the subjective clutter rating for each website (timed cases were excluded in this analysis). A monotonic relationship was present as assessed by visual inspection of a scatterplot. For **Words**, there was a strong positive correlation between Subjective Clutter Rating and RT, $r_s = .662$, $p = 0.001$. For **Images**, there was also a positive correlation between Subjective Clutter Rating and RT, $r_s = .445$, $p = 0.049$.

The subjective clutter ratings were also correlated using Spearman's Correlation, with the calculated clutter measurements of each website (average of all FC, SE, and ED measures). A monotonic relationship was present as assessed by visual inspection of a scatterplot. There was a strong correlation between the clutter algorithms and the subjective clutter rating, $r_s=0.813$, $p<0.0005$.

Participants also rated their perceived mental workload in the low time pressure and high time pressure conditions using NASA-TLX scales (scale: 0 to 10 [largest effect]). These rankings were analyzed using a Wilcoxon test. As can be seen in Table 4, there were significant effects of time pressure on all six scales, with time pressure notably resulting in increased impressions of temporal demand and performance degradation.

Table 4. Results of the NASA Task Load Index (NASA-TLX) ratings along the different dimensions

Nasa TLX Scale (1-10)	Non-Timed Median (SD)	Timed Cases Median (SD)	Median Increase/Decrease	Wilconxon Ranked Sign Test
Mental Demand	6 (1.76)	7 (1.91)	1	z=-2.932, p=0.03

Temporal Demand	4 (1.99)	8 (1.16)	4	z=-6.055, p<0.0005
Performance	8 (1.42)	6 (1.52)	-2	z=4.689, p<0.0005
Effort	7 (2.02)	8 (1.617)	1	z=-3.124, p=0.002
Frustration	4 (2.74)	7 (2.58)	1	Z=-3.069, p=0.002

B. Performance Results

1. *Response time*

Response time was calculated for trials with correct answers within the time limit only. Note that only non-timed-out cases were considered for response time (i.e., neglected cases where participants were not able to find the target before the 10-second limit in timed cases); this is to avoid including the cutoff time of 10 seconds in the results as response time. For **Words**, clutter caused an overall increase in response time from 10.17 seconds (SEM=0.78) in the low clutter condition to 25.68 seconds (SEM=2.56) in the high clutter condition (see Figure 1.a), with a significant interaction effect between clutter and time: $F(1, 31) = 48.068$, $p < 0.0005$, $\eta_p^2 = 0.608$. Moreover, there was a simple main effect of clutter in the non-timed condition $F(1, 31) = 55.077$, $p < 0.0005$, $\eta_p^2 = .640$ and also a significant simple main effect of clutter in the timed condition $F(1,31) = 111.10$, $p<0.0005$, $\eta_p^2 = .782$. Note that 19 participants' data were removed from this analysis as they represented the 10-second cut-off limit. For **Images**, logarithmic data transformation was applied to obtain data normality. Results showed that clutter caused a slight overall increase in response time from 3.13 seconds (SEM=0.19) in the low clutter condition to 4.15 seconds (SEM=0.26) in the high clutter condition (see Figure 1.b). There was no interaction effect between clutter and time pressure. However, there was a main effect of clutter on response time $F(1, 48) = 51.493$, $p < 0.0005$. $\eta_p^2 = 0.518$, and a main effect of time pressure on response time

F(1, 48) = 13.275 , p = 0.001. η_p^2 = 0.217. Note that one outlier was removed in the case of images.

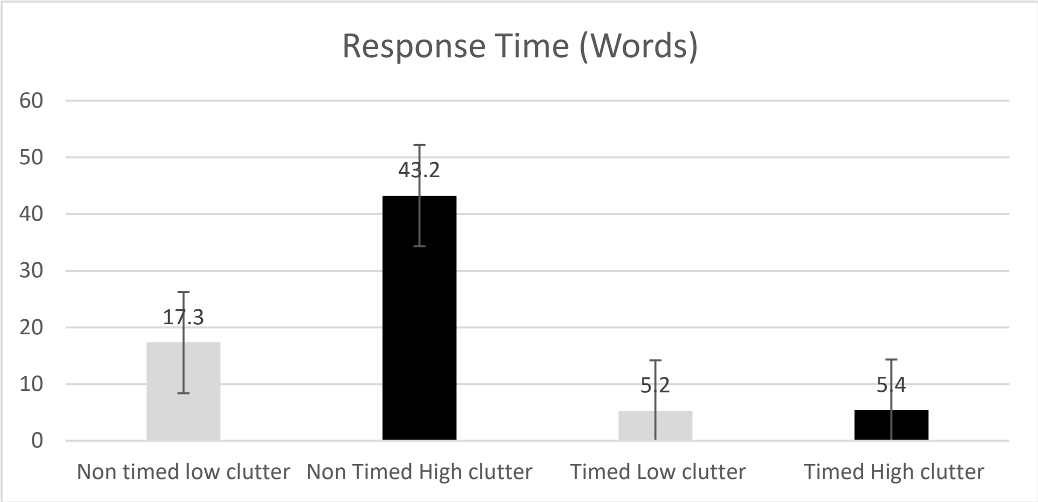


Figure 4.a. Response time in seconds in words

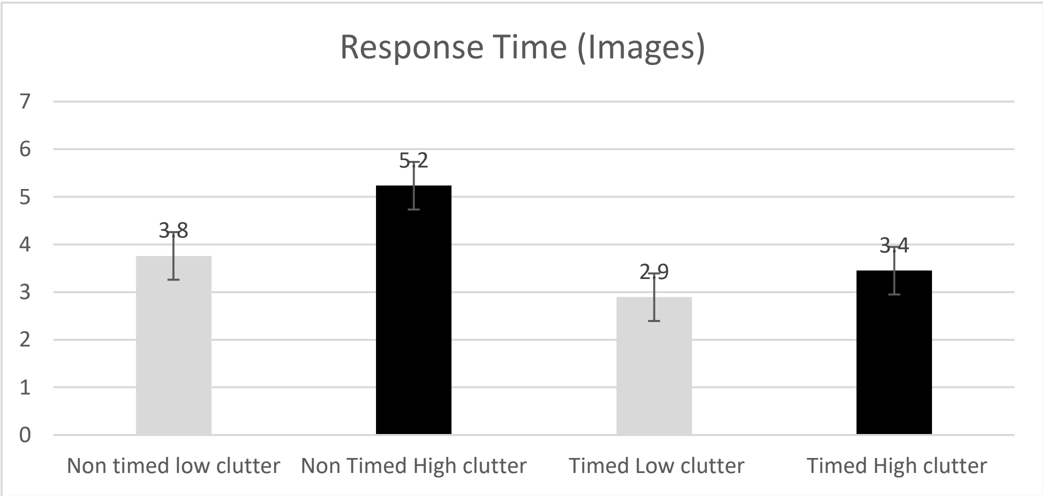


Figure 4.b. Response time in seconds in images

*2. Display Features and **Response** Time*

Rosenholtz's measures were also correlated with response time for both words and images targets. The results in table 5 show positive correlations for all three metrics in both cases.

Table 5. Rosenholtz's measures correlated with response time

Rosenholtz's measures correlated with response time	Words	Images
Feature Congestion	Positive Correlation, $r_s = .603$, $p = 0.005$	Positive Correlation, $r_s = .827$, $p < 0.0005$
Subband Entropy	Positive Correlation, $r_s = .669$, $p = 0.001$	Positive Correlation, $r_s = .967$, $p < 0.0005$
Edge Density	Positive Correlation, $r_s = .597$, $p = 0.005$	Positive Correlation, $r_s = .848$, $p < 0.005$

3. Search Error Rate

To determine the effects of clutter, and time pressure on the likelihood of not finding a target, a Wilcoxon signed-rank test for each website was performed. Three types of errors were identified: 1) time-out errors, meaning that the allocated time ran out before participants could find the target, 2) giving-up errors, meaning that the participants gave up on the target, and 3) target misses, meaning that participants incorrectly identified the target. The first type of errors, time-out errors are only relevant in the timed cases. For the **giving-up errors**, in the *non-timed cases*, there was a statistically significant change ($Z = -3.501$, $p < 0.0005$) between low and high clutter. In the *timed cases*, there was no statistically significant change ($Z = -1.342$, $p = 0.180$). For the **target misses errors**, in the *non-timed cases*, there was a statistically significant change ($Z = -2.456$, $p = 0.014$). In the *timed cases*, there was no statistically significant change ($Z = -1.732$, $p = 0.083$). For the **time-out errors**, there was a significant change between timed low clutter case and timed high clutter case (Z=-4.992, p<0.0005) (see table 6).

Table 6. Error percentages in different conditions

Error types	**Non-timed Low**	**Non-timed High**	**Timed Low**	**Timed High**
Time-out	-	-	**20.7%***	**39.2%***
Target misses	**1.5%***	**3.3%***	0.9%	1.5%
Giving-up	**1.3%***	**9.6%***	1.3%	0.2%

*Significant change is present between low and high clutter according to Wilcoxon test

C. Eye Tracking Results

After eliminating the periods corresponding to eye blinks, the raw eye tracking data (gaze points) were separated into fixations (minimum duration for fixations was 100 ms) and saccades (gaze points were within a radius of 1° visual angle). We examined both first examined the search eye tracking metrics (i.e., from the instance that the search begins until search target detection) as well as the early eye tracking metrics using data from only the first 3 seconds of search. The latter was done in order to identify eye tracking metrics that can capture the effects of clutter early on. We chose this time window as it was shorter than almost all the response times but still long enough to calculate the metrics in addition to the fact that it had already been tested in a previous study done by Moacdieh and Sarter (2017). Some responses' eye-tracking metrics were not able to be calculated within the first 3 seconds and had to be discarded. The overall eye tracking quality for the 50 participants was %90.4, an excellent quality as an average eye tracking quality is about %80.

1. *Search eye tracking metrics*

For **Words**, results showed a significant interaction effect of clutter with time pressure for four spread metrics: total fixation number, convex hull area, spatial density, and NNI (see Table 6). For the directness metrics, there was a significant interaction effect of clutter with time pressure on scanpath length per second, transition rate, mean saccade length, and scanpath ratio. There were also significant main effects of both clutter and time pressure on all metrics except for the backtrack rate (See Table 6). Note that for NNI analysis, four outliers were removed and logarithmic data transformation was applied. For mean fixation duration, logarithmic data transformation was applied, and for transition rate, three outliers were removed.

Table 7. Mean values of the search metrics (words) in the different clutter conditions.

Eye Tracking Metrics *(For Words)*	**Low Clutter (SEM)**	**High Clutter (SEM)**	**Simple Main effects of Clutter in Non-Timed Cases**	**Simple Main Effects of Clutter in Timed Cases**	**Main effects of Clutter**	**Interaction Effects**
SPREAD METRICS						
Convex Hull Area	720654.1 (26337.5)	**1066900.2 (33329.7)**	$F(1, 49) = 126.45$, $p < 0.0005$, $\eta_p^2 = 0.721$	$F(1, 49) = 26.09$, $p < 0.0005$, $\eta_p^2 = 0.347$	$F(1, 49) = 222.730$, $p < 0.0005$, $\eta_p^2 = 0.820$	Clutter*Time: $F(1,49)=12.914$, $p=0.001$, $\eta_p^2 = 0.209$
Spatial Density	0.0651 (0.003)	**0.1272 (0.008)**	$F(1, 49) = 109.24$, $p < 0.0005$, $\eta_p^2 = 0.690$	$F(1, 49) = 23.99$, $p < 0.0005$, $\eta_p^2 = 0.329$	$F(1, 49) = 143.949$, $p < 0.0005$, $\eta_p^2 = 0.746$	Clutter*Time: $F(1, 49) = 63.799$, $p < .0005$, $\eta_p^2 = 0.566$
Nearest Neighbor Index	0.552 (0.0117)	**0.659 (0.0136)**	$F(1, 46) = 84.8$, $p < 0.0005$, $\eta_p^2 = 0.648$	$F(1, 46) = 33.9$, $p < 0.0005$, $\eta_p^2 = 0.432$	$F(1, 46) = 6.299$, $p = .016$, $\eta_p^2 = 0.118$	Clutter*Time: $F(1, 46) = 4.190$, $p = 0.046$, $\eta_p^2 = 0.083$
Total Fixation Number	30.99 (2.351)	**67.38 (5.795)**	$F(1, 49) = 73.88$, $p < 0.0005$. $\eta_p^2 = 0.601$	$F(1, 49) = 73.88$, $p < 0.0005$, $\eta_p^2 = 0.601$	$F(1, 49) = 113.733$, $p < 0.0005$, $\eta_p^2 = 0.699$	Clutter*Time: $F(1, 49) = 48.53$, $p < .0005$, $\eta_p^2 = 0.498$
DURATION METRICS						
Mean Fixation Duration	**198.80 (2.43)**	192.51 (2.18)	_	_	$F(1, 49) = 5.631$, $p = 0.022$, $\eta_p^2 = 0.103$	_
DIRECTNESS METRICS						
Scan path Length per Second	**0.711 (0.016)**	0.701 (0.016)	$F(1,49) = 12.007$, $p < 0.0005$. $\eta_p^2 = 0.965$	_	$F(1, 49) = 62.703$, $p = 0 < 0.0005$, $\eta_p^2 = 0.561$	Clutter*Time: $F(1, 49) = 11.564$, $p = 0.001$, $\eta_p^2 = 0.191$
Scanpath Ratio	15.57 (0.82)	**30.47 (2.71)**	$F(1, 49) = 44.538$, $p < 0.005$, $\eta_p^2 = 0.805$	$F(1,49) = 11.955$, $p = 0.001$, $\eta_p^2 = 0.937$	$F(1, 49) = 57.276$, $p = 0 < 0.0005$, $\eta_p^2 = 0.539$	Clutter*Time: $F(1, 49) = 32.366$, $p < 0.0005$, $\eta_p^2 = 0.398$
Mean Saccade Length	**252.94 (4.03)**	242.28 (4.46)	_	_	$F(1, 49) = 48.285$, $p < 0.0005$, $\eta_p^2 = 0.164$	_
Backtrack Rate	0.00129 ($3.67*10^{-5}$)	**0.00132 ($3.82*10^{-5}$)**	_	_	_	_

Transition Rate	0.0020 ($4.7*10^{-5}$)	**0.0021 ($5.2*10^{-5}$)**	–	$F(1, 48) = 17.429$, $p < 0.0005$, $\eta_p^2 = 0.266$	$F(1, 48) = 16.973$, $p = 0 < 0.0005$, $\eta_p^2 = 0.261$	Clutter*Time: $F(1, 48) = 8.144$, $p=0.006$, $\eta_p^2 = 0.145$

[a]Higher values are in bold.
[b]Only significant results are reported here.

For **Images**, results showed a significant interaction effect of clutter with time pressure only for the backtrack rate metric. Simple main effects of clutter were only significant in timed cases as backtrack rate increases slightly from low to high clutter conditions (see Table 7). There was also significant main effects of clutter on all spread metrics. Spread metrics (total fixation number, convex hull area, spatial density, and NNI) all increase with clutter as well as the directness metrics (mean saccade length and scanpath ratio) increase with clutter. However, the rest of the directness metrics (scanpath length per second, and transition rate) stayed the same. The duration metric (mean fixation duration) has no significant interaction effect of clutter with time pressure and no significant main effects of both clutter and time pressure (see Table 7). Note that for convex hull area analysis, one outlier was removed. Moreover, logarithmic data transformation was applied for convex hull area, total fixation number, and scanpath ratio. Inverse data transformation was applied for spatial density.

Table 8. Mean Values of the Search Metrics (Images) in the Different Clutter Conditions

Eye Tracking Metrics *(For Images)*	**Low Clutter (SEM)**	**High Clutter (SEM)**	**Simple Main effects of Clutter in Non-Timed Cases**	**Simple Main Effects of Clutter in Timed Cases**	**Main effects of Clutter**	**Interaction Effects**
SPREAD METRICS						
Convex Hull Area	243285.9 (17959.2)	**352519.9 (30164.5)**	–	–	$F (1, 48) = 17.757, p < 0.0005, \eta_p^2 = 0.270$	–
Spatial Density	0.023 (0.001)	**0.033 (0.004)**	–	–	$F (1, 49) = 32.669, p < 0.0005, \eta_p^2 = 0.400$	–
Nearest Neighbor Index	0.47 (0.01)	**0.51 (0.01)**	–	–	$F (1, 49) = 9.210, p = 0.004, \eta_p^2 = 0.158$	–
Total Fixation Number	9.02 (0.58)	**13.66 (2.04)**	–	–	$F(1, 49) = 36.050, p < 0.0005, \eta_p^2 = 0.424$	–
DURATION METRICS						
Mean Fixation Duration	**233.1 (4.58)**	227.6 (4.33)	–	–	–	–
DIRECTNESS METRICS						
Scan path Length per Second	0.74 (0.018)	0.74 (0.018)	–	–	–	–
Scanpath Ratio	4.28 (0.25)	**5.63 (0.38)**	–	–	$F(1, 49) = 19.538, p < 0.0005, \eta_p^2 = 0.285$	–
Mean Saccade Length	259.95 (4.67)	**265.82 (7.00)**	–	–	–	–
BackTrack Rate	0.0011 ($3.88*10^{-5}$)	**0.0012 ($3.88*10^{-5}$)**	–	$F(1, 49) = 10.909, p = 0.002, \eta_p^2 = 0.182$	–	Clutter*Time: $F (1, 49) = 4.515, p=0.039, \eta_p^2 = 0.084$
Transition Rate	0.0021 ($6.21*10^{-5}$)	0.0021 ($5.77*10^{-5}$)	–	–	–	–

[a]Higher values are in bold.
[b]Only significant results are reported .

2. Early eye tracking metrics

Finally, we calculated once again all the metrics in the search, but this time over a period of 3 seconds.

For **Words**, none of the metrics showed significant interaction effects between clutter and time pressure; however, convex hull area, spatial density, NNI, total fixation number, and transition rate showed main effects of clutter and also showed a similar trend to what we observed during the search over the full response time (see Table 8).

For **Images**, none of the metrics showed significant interaction effects between clutter and time pressure; however, convex hull area, spatial density, NNI, total fixation number, scanpath ratio and backtrack rate showed main effects of clutter and also showed a similar trend to what we observed during the search over the full response time (see Table 9). Note that for convex hull area analysis, two outliers were removed.

Table 9. Mean Values of the Early Eye Tracking Measures (Words) in the Different Clutter Conditions

First 3 Seconds Eye Tracking Metrics *(For Words)*	**Low Clutter (SEM)**	**High Clutter (SEM)**	**Simple Main effects of Clutter in Non-Timed Cases**	**Simple Main Effects of Clutter in Timed Cases**	**Main effects of Clutter**	**Interaction Effects**
SPREAD METRICS						
Convex Hull Area	261585.9 (10366.64)	**285801.4 (9684.79)**	–	–	F (1, 49) =6.668, p = 0.013, η_p^2 = 0.120	–
Spatial Density	0.025 (0.0005)	**0.027 (0.0005)**	–	–	F(1, 49) = 11.982, p =0.001, η_p^2 = 0.196	–
Nearest Neighbor Index	0.529 (0.01)	**0.553 (0.01)**	–	–	F (1, 49) = 4.652, p = 0.036, η_p^2 = 0.087	–
Total Fixation Number	9.042 (0.22)	**9.448 (0.20)**	–	–	F(1, 49) = 6.617 , p = 0.013, η_p^2 = 0.119	–
DURATION METRICS						
Mean Fixation Duration	**188.48 (2.31)**	185.11 (2.06)	–	–	–	–
DIRECTNESS METRICS						
Scan path Length per Second	0.84 (0.018)	**0.86 (0.015)**	–	–	–	–
Scanpath Ratio	**4.86 (0.16)**	4.66 (0.14)	–	–	–	–
Mean Saccade Length	274.2 (5.14)	**280.4 (5.24)**	–	–	–	–
BackTrack Rate	**0.0013 ($3.7*10^{-5}$)**	0.0012 ($3.6*10^{-5}$)	–	–	–	–
Transition Rate	0.0025 ($4.8*10^{-5}$)	**0.0026 ($4.7*10^{-5}$)**	–	–	F (1, 49) = 0.308, p = 0.025, η_p^2 = 0.099	–

[a]Higher values are in bold.
[b]Only significant results are reported.

Table 10. Mean Values of the Early Eye Tracking Measures (Images) in the Different Clutter Conditions

First 3 Seconds Eye Tracking Metrics *(For Images)*	**Low Clutter (SEM)**	**High Clutter (SEM)**	**Simple Main effects of Clutter in Non-Timed Cases**	**Simple Main Effects of Clutter in Timed Cases**	**Main effects of Clutter**	**Interaction Effects**
SPREAD METRICS						
Convex Hull Area	159920 (6748.56)	**203939.4 (10603.17)**	–	–	$F(1, 49) = 5.742$, $p < 0.020$, $\eta_p^2 = 0.105$	–
Spatial Density	0.018 (0.0004)	**0.02 (0.0004)**	–	–	$F(1, 49) = 16.365$, $p < 0.0005$, $\eta_p^2 = 0.250$	–
Nearest Neighbor Index	0.47 (0.0102)	**0.51 (0.0121)**	–	–	$F(1, 49) = 6.250$, $p = 0.016$, $\eta_p^2 = 0.113$	–
Total Fixation Number	6.45 (0.17)	**7.21 (0.16)**	–	–	$F(1, 49) = 20.859$, $p < 0.0005$, $\eta_p^2 = 0.000$	–
DURATION METRICS						
Mean Fixation Duration	**231.512 (4.35)**	223.75 (3.85)	–	–	–	–
DIRECTNESS METRICS						
Scan path Length per Second	0.776 (0.018)	**0.814 (0.02)**	–	–	–	–
Scanpath Ratio	3.138 (0.113)	**3.456 (0.128)**	–	–	$F(1, 49) = 5.356$, $p = 0.025$, $\eta_p^2 = 0.099$	–
Mean Saccade Length	264.3 (5.10)	**277.3 (7.47)**	–	–	–	–
BackTrack Rate	0.0011 ($3.73*10^{-5}$)	**0.0012 ($4.09*10^{-5}$)**	–	–	$F(1, 49) = 5.022$, $p = 0.030$, $\eta_p^2 = 0.093$	–
Transition Rate	0.00224 ($5.65*10^{-5}$)	**0.00227 ($4.97*10^{-5}$)**	–	–	–	–

[a]Higher values are in bold.

[b]Only significant results are reported.

CHAPTER V

DISCUSSION

The overall goals of this research were to 1) identify what display features contribute the most to delays in e-commerce websites, 2) determine whether and to what extent clutter and time pressure interact to bring about performance decrements, and 3) identify the eye tracking metrics that best reflect clutter in e-commerce websites.

In general, the subjective results validated the manipulation of both clutter and time pressure. The participants' subjective ratings of clutter strongly correlated with the algorithm ratings, and also positively correlated with response time. It would thus seem that the metrics used to classify low and high clutter are valid. Similarly, for the manipulation of time pressure, participants' NASA-TLX ratings indicated that they felt under significant time pressure in the conditions labeled as high pressure, which was what was intended.

Data reliability was tested using Cronbach's alpha. Results showed that the subjective clutter ratings had a high level of internal consistency, with a high Cronbach's alpha of 0.966.(Tavakol & Dennick, 2011). Kendall's W was run to determine the interrater reliability between the 50 participants' clutter ratings on the 40 e-commerce websites. The 50 participants' statistically significantly agreed in their assessments ($W = .739$, $p < .0005$). The agreement between the 50 participants can explain 73.9% of all possible variability that would come with perfect agreement, which suggests good agreement between the participants.

A. Goal 1: Display features that contribute to delays in e-commerce websites

Based on the Spearman correlation values between the algorithm ratings and response time, it seems that the website feature that contributes most to the effects of clutter is the poor organization of items in the websites. This was the case for both sets of images – those that featured word targets as well as those that featured image targets. In both cases, the correlation coefficients between subband entropy and response time was the highest among all algorithms. This further confirms the importance of good organization when it comes to overcoming clutter (Doyon-Poulin et al., 2012). Color variation and the number of items in the display were not as strongly linked to the long response time as the organization and symmetry of the display. This held true regardless of whether the targets were images or words. This finding would suggest that website designers can add as much color variation and items as they like; what will mainly determine the target search time is how well-organized, logically grouped, and symmetrically designed their website is.

In comparison to previous studies using Rosenholtz' algorithms, this study found similar, and in some cases, better results (see Table 11). Comparisons between the different algorithms is difficult as many studies only used one metric. However, looking at the correlation values, apart from Rosenholtz et al.'s (2005, 2007) own validation, where maps and simple Gabor targets were used, other studies have found lower correlation values than in this study, or found no significant correlation. Studies such as Asher et al. (2013), used real-world images but did not see any significant results. No previous studies on website clutter have used these algorithms; the fact that this study showed significant correlation for website screenshots suggests that these algorithms can be used in this domain.

Table 11. RT correlated with Rosenholtz's measures

Rosenholtz's measures correlated with response time	Words	Images	Previous Literature
Feature Congestion	r_s= 0.603, p = 0.005	r_s= .827, p <0.0005	**Maps as stimuli:** r = 0.74 (p < .001) (Rosenholtz et al, 2005)
			Real-life images as stimuli: r = 0.53 (Henderson, Chanceaux, & Smith, 2009)
Subband Entropy	r_s= .669, p = 0.001	r_s= .967, p <0.0005	**Maps as stimuli:** r = .75 (p < .001) (Rosenholtz et al, 2005)
			Real-life images as stimuli: r = 0.42 (Henderson, Chanceaux, & Smith, 2009)
			Real-life images as stimuli: Not significant (Asher, Tolhurst, Troscianko, & Gilchrist, 2013)
Edge Density	r_s= .597, p = 0.005	r_s= .848, p <0.005	**Maps as stimuli:** r = .83 (p < .001) (Rosenholtz et al, 2005)
			Real-life images as stimuli: r = 0.53 (Henderson, Chanceaux, & Smith, 2009)

			City maps as stimuli: Not significant (Neider & Zelinksy, 2011)

It is also interesting to note that the correlations of the algorithms to response time were much higher in the case of image targets (up to 0.96) than the word targets (up to 0.66). Future research will need to further look into why this was the case.

Finally, it is worth noting that the relatively low correlation values across the literature provide support for the importance of user-based factors, such as previous experience or the knowledge of a given task (Fabrikant, Hespanha, & Hegarty, 2010; Kim & Kaber, 2009; Naylor, 2010; Neisser, 1976) when it comes to the effects of clutter. Display features alone cannot explain why one response time is long or short.

B. Goal 2: Performance effects of clutter and time pressure

The fact that increased clutter degrades response time and error rate has been well-established in the literature e.g., Moacdieh, & Sarter, 2015; Moacdieh, Ganji, & Sarter 2014; Neider, & Zelinsky, 2011; Bravo & Farid, 2006; Henderson, 2007; Levi, 2008; Beck, Trenchard, van Lamsweerde, Goldstein, & Lohrenz, 2012; Henderson, Chanceaux, & Smith, 2009). In other words, users tend to take longer to find their target under high clutter, and also tend to miss more of their targets due to clutter. This was what was expected in this study. It was also expected that the presence of time pressure would exacerbate the effects clutter, as was the case for Moacdieh and Sarter, 2015 high stress and task difficulty did exacerbate the performance effects of clutter.

In this study, performance results conformed to expectations overall, with both response time and error rate increasing in the case of high clutter. In the timed cases, the error rate refers specifically to the time-out error rate, with the time limit expiring significantly more often in the high-clutter condition, as opposed to low clutter.

However, in terms of response time, time pressure had no visible performance effects of clutter in this study. Nonetheless, in the case of word targets, interaction effects did occur between clutter and time pressure, but such interaction effects were absent in the case of image targets. This interaction effect between clutter and time pressure, in the case of words, suggests that time pressure worsened the effects of clutter when participants felt pressured to complete a task. In terms of the speed–accuracy trade-off, participants seemed to compromise accuracy in favor of speed. Response times were lower overall in case of high time pressure but this can be explained by the 10 second cut-off limit; 20% to 40% in timed-low clutter to timed-high clutter cases reached the cut-off limit and no further time was recorded. Search accuracy, on the other hand, was not significantly affected by time pressure. This could be due to the fact that search tasks used in this study were not necessarily difficult.

Moreover, it is interesting to note that response times were significantly shorter for image targets than those for word targets.. This could be attributed to the higher salience of image targets, although all targets when displayed to the participant, whether images or words, were of the same spatial area. A follow-up study could look further into this issue to identify exactly why it was faster to search for images, as this was outside the scope of this study.

C. Goal 3: Eye tracking metrics that best reflect clutter

After checking the eye-tracking metrics results, eye tracking seems to be a promising tool in detecting clutter, especially when using spread metrics. These seem to be the metrics that best reflect the effects of clutter.

In the case of word targets, the spread metrics all increased with clutter, as did the scanpath ratio and transition rate, two directness metrics. At the same time, another directness metric, mean saccade length, was significantly lower. The duration metric (mean fixation duration) significantly decreased with clutter. Taken together, these findings suggest that, in situations of high clutter, users tend to sample wide areas of the screen, while at the same time scanning less efficiently. In other words, users tend to go back and forth multiple times in an unsystematic fashion, an indication of confusion and lack of clear guidance to the target. At the same time, users were moving quickly across areas and in small steps (as evidenced by the small mean saccade length), suggesting that the increase in response time can be mainly attributed to the increase in space users had to cover and their inefficient scan paths towards the target, as opposed to a problem with discriminating or processing information. In general, users can be considered to alter their scan patterns in the case of high clutter by trying to quickly sample as many areas as they can in a very random, unsystematic way.

Similarly, for image targets, the spread metrics all increased with clutter as well indicating that the user was looking across different parts of the screen, while the directness metrics, backtrack rate and scanpath ratio, increased with clutter, indicating less directness of scan patterns. The two metrics that proved to be different between words and images were backtrack rate and transition rate. The former was significantly higher in high clutter in the case of image targets, while the second was significantly higher in high clutter in the case of word targets. This makes for an interesting finding, as it suggests that there was more back and forth eye movements across large sections of the screen when there are images to search for, whereas in the case of word targets, the transitions are not necessarily at such a large scale. This once again can be attributed to the salience of the images, which would allow users to make larger saccades towards the target.

In general, as can be seen in Table 12, the results are consistent with a lot of previous literature on eye tracking metrics. However, mean fixation duration seems to be different than most previous literature. In the case of words, mean fixation duration decreased significantly with clutter but had no significant effect in the case of images; this could be explained due to the fact that the user followed search pattern as if reading words, thus explaining the increase in horizontal transitioning and the shorter mean saccades and scan path lengths per second (Amor et al, 2016); thus no difficult processing was needed to discriminate word targets from other objects in the display (Beck et al, 2010).

Table 12. Previous literature on eye tracking metrics where ↑ means increase with clutter and ↓ means decrease with clutter

	Previous Literature*	**Current Study for Words**	**Current Study for Images**
SPREAD METRICS			
Convex Hull Area[1,2]	↑	↑	↑
Spatial Density[1,2,6]	↑	↑	↑
Nearest Neighbor Index[1,2]	↑	↑	↑
Total Fixation Number[1,3,5,10]	↑	↑	↑
DURATION METRICS			

Mean Fixation Duration[1,2,5,6,7,8]	↑	↓	-
DIRECTNESS METRICS			
Scanpath length per second[2]	↓	↓	-
Scanpath Ratio[9]	↑	↑	↑
Mean Saccade Length[1,2,7]	↓	↓	-
Backtrack Rate[1]	-	-	↑
Transition Rate[2,5,6]	↑	↑	-

*Previous literature in this table is an aggregate of studies by the following: [1]Moacdieh, & Sarter (2015, 2016); [2]Moacdieh, & Sarter (2015); [3]Yoon, Lim, & Ji (2015); [4]Moacdieh, Ganji, & Sarter (2014); [5]Doyon-Poulin, Robert, & Ouellette (2014), [6]Moacdieh, prinet, & Sater (2013); [7]Beck et al (2012); [8]Zhu, & Su (2012); [9]Neider, & Zelinsky (2011); [10]Beck, Lohrenz, & Trafton (2010).

While these results are derived from the eye tracking metrics across the whole response time period, for the purposes of real-time display adjustment, what is necessary is to be able to obtain these results as early as possible in the response time period. A window of three seconds was used in this study and the same metrics were calculated in that period as well. What is evident is that the spread metrics show the same pattern as in the whole response time period, with convex hull area, spatial density, NNI, and the number of fixations higher in the presence of clutter; no other metrics were significantly different for the first three seconds. So even after the first few seconds of search, users had already scanned a significantly larger portion of the screen in an attempt to overcome the effects of clutter. A previous study done by Moacdieh and Sarter (2015) also showed that both convex hull area and NNI followed the same trend in the first four seconds of search. However, another study done by the same authors (2017) built a predictive model of clutter based on three different metrics: mean saccade length, scanpath length, and mean fixation duration; these metrics were not strong indicators of clutter in this thesis. The main difference between this paper and the earlier study by Moacdieh and Sarter (2015) is that performance decrements were largely accentuated by high stress and task difficulty, something that was not evident in this study under time pressure. Another difference is that in the first four seconds, there were no significant effects of clutter on the eye tracking metrics although convex hull area, NNI, scanpath length per second, and mean saccade length revealed similar trends over full response.

CHAPTER VI

CONCLUSION

In conclusion, regarding the three Rosenholtz metrics, all three were significantly correlated to performance decrements; especially in the case of image targets. However, this doesn't eliminate the importance of including the human-factors aspect, such as experience and knowledge, in clutter perception, calculation, and detection. This is evident as many other researchers had different results depending on the application domain and targets used.

Overall, with clutter, the user needed more time to check different parts of the screen and move his/her eyes around in order to perform a search task as highlighted by the increase in spread metrics in all conditions from low to high clutter. The user also had less sense of guidance to reach the target and took longer paths on the screen until he/she reached the target as concluded from scanpath ratio and backtrack rate as both increased with clutter. The search accuracy also did decrease with clutter although no visible effects of time pressure on search accuracy was witnessed. This indicated a less efficient search mechanism and poorer performance in the case of high clutter.

Nonetheless, in this study, spread metrics seem to be the most promising eye tracking metrics, to be used in future clutter and time critical prediction models as these metrics have, constantly across all targets, been affected in the presence of clutter, even in a short time window of 3 seconds. This could tremendously help in detecting clutter effects even under time critical situations such as emergency programs or even simple non-critical domains such as graphic design.

In terms of intellectual merit, this research helps add to the literature on clutter and eye tracking, and emphasizes the potential of using eye tracking for real-time adaptive displays that are adjusted in real time. By pinpointing which eye tracking metrics best reflected the effects of clutter, this research can provide a starting point for the creation of models of eye tracking metrics. These models can then be used to predict human search behavior while using websites so that the optimal display can then be provided. Moreover, the research can potentially generalize to other complex domains where visual search is common, search as the medical field, the military, and aviation. These domains are also characterized by time pressure, and since many of these metrics are consistently affected by clutter despite time pressure, they may be able to form the basis of robust adaptive displays. In turn, the adoption of adaptive displays in such domains that suffer from high clutter and time pressure can promote safety and efficiency, reducing errors and processing time.

REFERENCES

Alexander A. L., Stelzer E. M., Kim S.-H., Kaber D. B. (2008). Bottom-up and top-down contributors to pilot perceptions of display clutter in advanced flight deck technologies. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1180–1184). Santa Monica, CA: Human Factors and Ergonomics Society.

Alexander A. L., Wickens C. D. (2005). Synthetic vision systems: Flightpath tracking, situation awareness and visual scanning in an integrated hazard display. InJensen R. (Ed.), *Proceedings of the 13th International Symposium on Aviation Psychology* [CD-ROM]. Oklahoma City, OK: Federal Aviation Administration.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye Movements Do Not Reflect Retrieval Processes Limits of the Eye-Mind Hypothesis. Psychological Science, 15(4), 225-231.

Asher M. F., Tolhurst D. J., Troscianko T., Gilchrist I. D. (2013). Regional effects of clutter on human target detection performance. Journal of Vision, 13(5), 25.

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(4), 693-709.

Baldassi S., Megna N., Burr D. C. (2006). Visual clutter causes high-magnitude errors. PLoS Biology, 4(3), e56.

Barbu C., Lohrenz M., Layne G. (2006). Intelligent electronic navigational aids: A new approach. In *Proceedings of the 5th International Conference on Machine Learning and Applications (ICMLA '06)* (pp. 109–116). New York, NY: IEEE.

Barr, D. J. (2008). Analyzing 'visual world'eyetracking data using multilevel logistic regression. *Journal of memory and language*, *59*(4), 457-474.

Beck M. R., Lohrenz M. C., Trafton J. G. (2010). Measuring search efficiency in complex visual search tasks: Global and local clutter. Journal of Experimental Psychology: Applied, 16, 238–250.

Beck, M. R., Peterson, M. S., Boot, W. R., Vomela, M., & Kramer, A. F. (2006). Explicit memory for rejected distractors during visual search. *Visual Cognition, 14*(2), 150-174.

Beck, M. R., Trenchard, M., van Lamsweerde, A., Goldstein, R. R., & Lohrenz, M. (2012). Searching in clutter: Visual attention strategies of expert pilots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1411-1415). Boston, MA: Sage Publications.

Bergström, A., Jaksetic, P., & Nordin, P. (2000, January). Enhancing information retrieval by automatic acquisition of textual relations using genetic programming. In *Proceedings of the 5th international conference on Intelligent user interfaces* (pp. 29-32). ACM.

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective stimulus in visual search. *Perception & Psychophysics, 62*(3), 576-585. doi:10.3758/BF03212109

Blom, E., Paradis, J., & Duncan, T. S. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular–s in child L2 English. *Language Learning, 62*(3), 965-994.

Bravo M. J., Farid H. (2004). Recognizing and segmenting objects in clutter. Vision Research, 44, 385–396.

Bravo M. J., Farid H. (2006). Object recognition in dense clutter. Attention, Perception, & Psychophysics, 68, 911–918.

Bravo M. J., Farid H. (2008). A scale invariant measure of clutter. Journal of Vision, 8(1), 23, 1–9.

Bruneau, D., Sasse, M. A., & McCarthy, J. D. (2002, April). The eyes never lie: The use of eye tracking data in HCI research. In Proceedings of the CHI (Vol. 2, p. 25).

Castagnos, S., & Pu, P. (2010, February). Consumer decision patterns through eye gaze analysis. In Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction (pp. 78-85). ACM.

Chang H., Zhang J., Liu X., Yang C., Li Q. (2010). Color image clutter metrics for predicting human target acquisition performance. In *Proceedings of the 6th International Conference Wireless Communications Networking and Mobile Computing (WiCOM)* (pp. 1–4). New York, NY: IEEE.

Chetwood, A. S., Kwok, K. W., Sun, L. W., Mylonas, G. P., Clark, J., Darzi, A., & Yang, G. Z. (2012). Collaborative eye tracking: a potential training tool in laparoscopic surgery. Surgical Endoscopy, 26(7), 2003–2009.

Christ R. E. (1975). Review and analysis of color coding research for visual displays. Human Factors, 17, 542–570.

Chu X., Yang C., Li Q. (2012). Contrast-sensitivity-function-based clutter metric. Optical Engineering, 51(6), 067003-1-6.

Chukoskie, L., Snider, J., Mozer, M. C., Krauzlis, R. J., & Sejnowski, T. J. (2013). Learning where to look for a hidden target. *Proceedings of the National Academy of Sciences, 110*(Supplement 2), 10438-10445.

Clay M. C. (1993). *Key cognitive issues in the design of electronic displays of instrument approach procedure charts* (Department of Transportation Report DOT-VNTSC-FAA-93-18). Washington, DC: U.S. Department of Transportation.

Coco M. I, Keller F. (2009). The impact of visual information on reference assignment in sentence production. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society (CogSci)* (pp. 274–279). Amsterdam, Netherlands:Cognitive Science Society.

Cowen, L., Ball, L. J., & Delin, J. (2002). An eye movement analysis of web page usability. In People and Computers XVI-Memorable Yet Invisible (pp. 317-335). Springer London.

Cummings, M. L., Brzezinski, A. S., & Lee, J. D. (2007). The impact of intelligent aiding for multiple unmanned aerial vehicle schedule management.

DeWitt, A. J. (2010). Examining the order effect of website navigation menus with eye tracking. *Journal of Usability Studies, 6*(1), 39-47.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making, 1*(3), 271-285.

Di Stasi, L. L., Contreras, D., Cándido, A., Cañas, J. J., & Catena, A. (2011). Behavioral and eye-movement measures to track improvements in driving skills of vulnerable road users: First-time motorcycle riders. *Transportation research part F: traffic psychology and behaviour, 14*(1), 26-35.

Doyon-Poulin P., Robert J., Ouellette B. (2012). Review of visual clutter and its effects on pilot performance: A new look at past research. In *Proceedings of the 2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC)* (pp. 2D1-1–2D1-11). New York, NY: IEEE.

Duchowski, A. T., & SpringerLink (Online service). (2007). *Eye tracking methodology: Theory and practice* (2nd;2;Second; ed.). London: Springer. doi:10.1007/978-1-84628-609-4

Duncan J., Humphreys G. W. (1989). Visual search and stimulus similarity. Psychological Review, 96, 433.

Durkee, K., Geyer, A., Pappada, S., Ortiz, A., & Galster, S. (2013). Real-Time Workload Assessment as a Foundation for Human Performance Augmentation. In *Foundations of Augmented Cognition* (pp. 279-288). Springer Berlin Heidelberg.

Edwards S. M., Li H., Lee J. H. (2002). Forced exposure and psychological reactance: antecedents and consequences of the perceived intrusiveness of pop-up ads. J. Advert. 31, 83–95. 10.1080/00913367.2002.10673678

Ehmke, C., & Wilson, S. (2007, September). Identifying web usability problems from eye-tracking data. In Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1 (pp. 119-128). British Computer Society.

Ellis G., Dix A. (2007). A taxonomy of clutter reduction for information visualisation. IEEE Transactions on Visualization and Computer Graphics, 13, 1216–1223.

Ellis, K. K. E., Kramer, L. J., Shelton, K. J., Arthur, J. J. (., & Prinzel, L. J. (2011). Transition of attention in terminal area NextGen operations using synthetic vision systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 46-50.

Ewing G. J., Woodruff C. J., Vickers D. (2006). Effects of “local” clutter on human target detection. Spatial Vision, 19, 37–60.

Fabrikant, S. I., Hespanha, S. R., & Hegarty, M. (2010). Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making. *Annals of the Association of American Geographers, 100*(1), 13-29.

Fadden S., Ververs P. M., Wickens C. D. (1998). Costs and benefits of head-up display use: A meta-analytic approach. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 16–20). Santa Monica, CA: Human Factors and Ergonomics Society.

Findlay J. M. (2004). Eye scanning and visual search. In Henderson J. M., FerreiraF. (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 135–160). New York, NY: Psychology Press.

Findlay, J. M., & Gilchrist, I. D. (2005). Eye guidance and visual search. In G. Underwood, Cognitive processes in eye guidance, 282-295. New York: Oxford University Press, USA

Geyer, T., Von Mühlenen, A., & Müller, H. J. (2007). What do eye movements reveal about the role of memory in visual search?. *The Quarterly Journal of Experimental Psychology, 60*(7), 924-935.

Gibson, C. N. (n.d.). Introduction to Deep Neural Networks. Retrieved March 23, 2017, from https://deeplearning4j.org/neuralnet-overview

Global B2C Ecommerce Sales to Hit $1.5 Trillion This Year Driven by Growth in Emerging Markets - eMarketer. (2014). Retrieved October 11, 2016, from http://www.emarketer.com/Article/Global-B2C-Ecommerce-Sales-Hit-15-Trillion-This-Year-Driven-by-Growth-Emerging-Markets/1010575

Goldberg J. H., Kotval X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. International Journal of Industrial Ergonomics, 24, 631–645.

Goldberg, H. J., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner’s guide. In J. Hyönä, R. Radach, & H. Deubel (Eds.), The mind's eye: Cognitive and applied aspects of eye movement research (pp. 493-516). Amsterdam: Elsevier

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002, March). Eye tracking in web search tasks: design implications. In *Proceedings of the 2002 symposium on Eye tracking research & applications* (pp. 51-58). ACM.

Grahame, M., Laberge, J., & Scialfa, C. T. (2004). Age differences in search of web pages: The effects of link size, link number, and clutter. Human Factors, 46(3), 385-98. Retrieved from http://search.proquest.com/docview/216438369?accountid=8555

Granka, L., Feusner, M., & Lorigo, L. (2008). Eye monitoring in online search. In Passive eye monitoring (pp. 347-372). Springer Berlin Heidelberg.

Greene, H. H., & Rayner, K. (2001). Eye movements and familiarity effects in visual search. *Vision Research, 41*(27), 3763-3773. doi:10.1016/S0042-6989(01)00154-7

Hammond K. W., Efthimiadis E. N., Laundry R. (2012). Efficient de-identification of electronic patient records for user cognitive testing. In *Proceedings of the 45th Hawaii International Conference on System Sciences* (pp. 2771–2778). New York, NY: IEEE.

He G., Zhang J., Liu D., Chang H. (2008). Clutter metric based on the Cramer-Rao lower bound on automatic target recognition. Applied Optics, 47, 5534–5540.

Hegarty, M., De Leeuw, K., & Bonura, B. (2008, November). What do spatial ability tests really measure. In Proceedings of the 49th Meeting of the Psychonomic Society. Chicago, IL.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. *Eye movements: A window on mind and brain*, 537-562.

Henderson, J., Chanceaux, M., & Smith, T. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. Journal of Vision, 9(1), 32-32. doi:10.1167/9.1.32

Henderson, J.M., Weeks, P.A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during scene viewing. Journal of Experimental Psychology: Human Perception and Performance, 25, 210–228.

Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? Visual Cognition, 10, 257–287.

Horrey W. J., Wickens C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. Human Factors, 46, 611–624.

Hou, M., Kobierski, R. D., & Brown, M. (2007). Intelligent adaptive interfaces for the control of multiple UAVs. *Journal of Cognitive Engineering and Decision Making, 1*(3), 327-362.

Huang, X., Shen, C., Boix, X., & Zhao, Q. (2015). SALICON: Reducing the semantic gap in saliency prediction by adapting deep neural networks. Paper presented at the 262-270. doi:10.1109/ICCV.2015.38

Hyrskykari, A. (2006). Utilizing eye movements: Overcoming inaccuracy while tracking the focus of attention during reading. *Computers in human behavior,22*(4), 657-671.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. *Handbook of cognitive task design, 8*, 147-169.

Inhoff, A.W., & Radach, R. (1998). Definition and computation of EYE MOVEMENTS IN READING 411 oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-54). Oxford, England: Elsevier.

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task-evoked pupillary response to mental workload in human-computer interaction. In E. DykstraErickson & K.Manfred Tscheligi (Eds.), *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1477–1480). New York: ACM Press.

Jack Rotfeld, H. (2006). Understanding advertising clutter and the real solution to declining audience attention to mass media commercial messages. *Journal of Consumer Marketing,23*(4), 180-181. doi:10.1108/07363760610674301

Jacob R. J. K., Karn K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Radach J. H., Deubel H. (Eds.), *In the mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Amsterdam, Netherlands: Elsevier Science.

Jin, L., Niu, Q., Jiang, Y., Xian, H., Qin, Y., & Xu, M. (2013). Driver sleepiness detection system based on eye movements variables. *Advances in Mechanical Engineering, 2013.*

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009, September). Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on* (pp. 2106-2113). IEEE.

Just, M. A., & Carpenter, P. A. (1978). Inference processes during reading: Reflections from eye fixations. In J. W. Senders, D. F. Fisher, & R. A. Monty (Eds.), Eye movements and the higher psychological functions. Hillsdale, N.J.: Erlbaum.

Kaber, D. B., Naylor, J. T., Gil, G. H., Pankok, C., & Kim, S. H. (2013). Influence of flight domain and cockpit display dynamics on pilot perceived clutter. *Journal of Aerospace Information Systems, 10*(12), 550-559.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors: The Journal of the Human Factors and Ergonomics Society,47*(4), 730-741.

Kardan, S., & Conati, C. (2012). Exploring gaze data for determining user learning with an interactive simulation. In *User Modeling, Adaptation, and Personalization* (pp. 126-138). Springer Berlin Heidelberg.

Katsanos, C., Tselios, N., & Avouris, N. (2010). Evaluating website navigability: validation of a tool-based approach through two eye-tracking user studies. *New review of Hypermedia and Multimedia, 16*(1-2), 195-214.

Kaufmann K., Kaber D. B. (2010). The influence of individual differences in perceptual performance on pilot perceptions of head-up display clutter. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp. 70–74). Santa Monica, CA: Human Factors and Ergonomics Society.

Keeble, R. J., & Macredie, R. D. (2000). Assistant agents for the world wide web intelligent interface design challenges. *Interacting with Computers, 12*(4), 357-381.

Kim N. Y., Sundar S. S. (2010). Relevance to the rescue: Can "smart ads" reduce negative response to online ad clutter? Journalism & Mass Communication Quarterly, 87, 346–362.

Kim S.-H., Kaber D. B. (2009). Assessing the effects of conformal terrain features in advanced head-up displays on pilot performance. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 36–40). Santa Monica, CA: Human Factors and Ergonomics Society.

Klein, R. M. (2000). Inhibition of return. *Trends in cognitive sciences, 4*(4), 138-147.

Konstantopoulos, P., Chapman, P., & Crundall, D. (2010). Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye 40 movements in day, night and rain driving. *Accident Analysis & Prevention, 42*(3), 827- 834.

Kozma, L., Klami, A., & Kaski, S. (2009, November). GaZIR: gaze-based zooming interface for image retrieval. In *Proceedings of the 2009 international conference on Multimodal interfaces* (pp. 305-312). ACM.

Kroft P., Wickens C. D. (2002). Displaying multi-domain graphical database information: An evaluation of scanning, clutter, display size, and user activity. Information Design Journal, 11, 44–52.

Kruthiventi, S. S. S., Ayush, K., & Babu, R. V. (2015). DeepFix: A fully convolutional neural network for predicting human eye fixations

Kumar, M., Winograd, T., & Paepcke, A. (2007, April). Gaze-enhanced scrolling techniques. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems* (pp. 2531-2536). ACM.

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *Intelligent Transportation Systems, IEEE Transactions on, 8*(2), 340-350.

Liao M. J., Wu Y., Sheu C. F. (2013). Effects of perceptual complexity on older and younger adults' target acquisition performance. Behaviour & Information Technology. Advance online publication. 10.1080/0144929X.2013.847974.

Lohrenz M. C., Layne G. L., Edwards S. S., Gendron M. L., Bradley J. T. (2006, March). *Feature clustering to measure clutter in electronic displays*. *Paper* *presented at the 12th Industry, Engineering, and Management Systems Conference,* Cocoa Beach, FL.

Lohrenz M. C., Trafton J. G., Beck M. R., Gendron M. L. (2009). A model of clutter for complex, multivariate geospatial displays. Human Factors, 51, 90–101.

Lopes, A., & Galletta, D. (2006). Consumer Perceptions and Willingness to Pay for Intrinsically Motivated Online Content. *Journal of Management Information Systems, 23*(2), 203-231. Retrieved from http://www.jstor.org/stable/40398847

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., ... & Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. Journal of the American Society for Information Science and Technology, 59(7), 1041-1052.

Mack M. L., Oliva A. (2004, November). *Computational estimation of visual complexity*. *Paper presented at the 12th Annual Object, Perception, Attention, and Memory Conference*, Minneapolis, MN.

Marquard, J. L., Jo, J., Henneman, P. L., Fisher, D. L., & Henneman, E. A. (2012). Can Visualizations Complement Quantitative Process Analysis Measures? A Case Study of 41 Nurses Identifying Patients Before Administering Medications. *Journal of Cognitive Engineering and Decision Making,* 155534341245768

McKenzie M., Wong S., Gibbins D. (2013). An empirical evaluation of infrared clutter for point target detection algorithms. In Sadjadi F. A., Mahalanobis A. (Eds.), *Proceedings of SPIE Vol. 874409: Automatic Target Recognition XXIII*. Bellingham, WA: International Society for Optics and Photonics (SPIE). 10.1117/12.2015668

Moacdieh, N., & Sarter, N. (2015). Display clutter: A review of definitions and measurement techniques. *Human Factors: The Journal of Human Factors and Ergonomics Society, 57*(1), 61-100. doi:10.1177/0018720814541145

Mokhtar, R., Abdullah, S. N. H. S., & Zin, N. A. M. (2011, June). Classifying modality learning styles based on Production-Fuzzy Rules. In *Pattern Analysis and Intelligent Robotics (ICPAIR), 2011 International Conference on* (Vol. 1, pp. 154-159). IEEE.

Moore R. K., Camp H. A., Moyer S., Halford C. E. (2010). *Masked target transform volume clutter metric applied to vehicle search*. In Holst G. C., Krapels K. A. (Eds.), *Proceedings of SPIE Vol. 7662: Infrared Imaging Systems. Design, Analysis, Modeling, and Testing XXI.* Bellingham, WA: International Society for Optics and Photonics (SPIE). 10.1117/12.850429

Munn S. M., Stefano L., Pelz J. B. (2008). Fixation-identification in dynamic scenes: Comparing an automated algorithm to manual coding. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization* (pp. 33–42). New York, NY: ACM Press.

Naylor J. (2010). The influence of dynamics, flight domain and individual flight training & experience on pilot perception of clutter in aviation displays (Master's thesis). Available from http://repository.lib.ncsu.edu/ir/handle/1840.16/6103

Nielsen J. (1993). *Usability engineering*. Boston, MA: Academic Press.

Noton D., Stark L. (1971). Scanpaths in eye movements during pattern perception. Science, 171, 308–311.

Oliva A., Mack M. L., Shrestha M., Peeper A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 1041–1046). Chicago, IL: Cognitive Science Society.

Olson, W. A., & Sarter, N. B. (2000). Automation management strategies: Pilot preferences and operational experiences. *The International Journal of Aviation Psychology, 10*(4), 327341.

Olsson P.-O., Pippig K., Harrie L., Stigmar H. (2011). Identifying map regions that are difficult to read. Mapping and Image Science, 3, 22–29.

Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. (2004, March). The determinants of web page viewing behavior: an eye-tracking study. In Proceedings of the 2004 symposium on Eye tracking research & applications (pp. 147-154). ACM.

Parasuraman, R., Warm, J. S., & See, J. E. (1998). Brain systems of vigilance. In R. Parasuraman (Ed.), The attentive brain (pp. 221–256). Cambridge, MA: MIT Press

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research, 42*(1), 107-123.

Pavlas, D., Lum, H., & Salas, E. (2012). How to build a low-cost eye-tracking system. *Ergonomics in Design: The Quarterly of Human Factors Applications, 20*(1), 18-23.

Peng W., Ward M. O., Rundensteiner E. A. (2004). Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS)* (pp. 89–96). New York, NY: IEEE.

Phillips R. J., Noyes L. (1982). An investigation of visual clutter in the topographic base of a geological map. Cartographic Journal, 19, 122–132.

Poole A., Ball L. J. (2005). Eye tracking in human-computer interaction and usability research. In Ghaoui C. (Ed.), *Encyclopedia of human computer interaction*(pp. 211–219). Hershey, PA: Idea Group.

Pope, A.T., Bogart, E.H., Bartolome, D.S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology 40, 187–195.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closedloop system for examining psychophysiological measures for adaptive task allocation. *The International journal of aviation psychology, 10*(4), 393-410.

Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2008, April). Predicting postcompletion errors using eye movements. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 539-542). ACM.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372.

Rosenholtz R., Li Y., Mansfield J., Jin Z. (2005). Feature congestion: A measure of display clutter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 761–770). New York, NY: ACM Press

Rosenholtz R., Li Y., Nakano L. (2007). Measuring visual clutter. Journal of Vision, 7(2), 1–22. 10.1167/7.2.17

Roth, S. P., Tuch, A. N., Mekler, E. D., Bargas-Avila, J. A., & Opwis, K. (2013). Location matters, especially for non-salient features–An eye-tracking study on the effects of web object placement on different types of websites. *International Journal of Human-Computer Studies, 71*(3), 228-235. doi:10.1016/j.ijhcs.2012.09.001

Russell B. (2005). Neural network applications in geophysics. Hampson-Russell Software, a Veritas Company, Calgary, Canada, 1-3

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eyetracking protocols. In Proceedings of the 2000 symposium on Eye tracking research & applications (pp. 71-78). ACM.

Schnell, T., Kwon, Y., Merchant, S., & Etherington, T. (2004). Improved flight technical performance in flight decks equipped with synthetic vision information system displays. *The International Journal of Aviation Psychology*, *14*(1), 79-102. doi:10.1207/s15327108ijap1401_5

Schons, V. W., and Wickens, C. D. (1993). Visual separation and information access in aircraft display layout (Technical Report ARL-93-7/NASA-A3 I-93-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Shen, S., Zhang, Y., Peng, L., Du, Y., & Wang, Q. (2011). Matchstick-shaped Ag2S-ZnS heteronanostructures preserving both UV/blue and near-infrared photoluminescence.Angewandte Chemie *(International Ed. in English), 50*(31), 7115-7118. doi:10.1002/anie.201101084

Shneiderman, B. (1997). Designing information-abundant web sites: issues and recommendations. *International Journal of Human-Computer Studies, 47*(1), 5-29.

Shneiderman, B. (1998). Educational journeys on the web frontier. *Educom Review, 33*(6), 10.

Spiekermann, S., & Korunovska, J. (2014). The importance of interface complexity and entropy for online information sharing. *Behaviour & Information Technology*, *33*(6), 636-645. doi:10.1080/0144929X.2013.845910

Steichen, B., Carenini, G., & Conati, C. (2013, March). User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities. In *Proceedings of the 2013 international conference on Intelligent user interfaces* (pp. 317-328). ACM.

Tang, A., Lu, K., Wang, Y., Huang, J., & Li, H. (2015). A real-time hand posture recognition system using deep neural networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *6*(2), 1-23. doi:10.1145/2735952

The visual presentation of quantitative data, by edward tufte. cheshire, CT: Graphics, 1983. (1984). *Journal of Policy Analysis and Management*, *3*(2), 320-321. doi:10.1002/pam.4050030255

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379-385.

Today's chart shows a platform breakdown of U.S. online and mobile shopping revenue alongside a forecast of mobile commerce revenue in the United States. (n.d.). E-Commerce | Statista. Retrieved October 11, 2016, from https://www.statista.com/markets/413/e-commerce/

Toet A. (2010). Structural similarity determines search time and detection probability. Infrared Physics & Technology, 53, 464–468.

Trukenbrod, H. A., & Engbert, R. (2007). Oculomotor control in a sequential search task. *Vision research*, *47*(18), 2426-2443.

Tufte. (1991). Envisioning information. *Program*, *25*(1), 88.

Tullis T. S. (1983). The formatting of alphanumeric displays: A review and analysis. Human Factors, 25, 657–682.

Van den Berg R., Cornelissen F. W., Roerdink J. (2009). A crowding model of visual clutter. Journal of Vision, 9(4), 1–11. 10.1167/9.4.24

Vlaskamp, B. N., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. Vision Research, 46, 417–425.

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.

Ware, C. (2004). *Information visualization, 2nd edition* (2nd ed.) Morgan Kaufmann.

Westerbeek H. G. W., Maes A. (2011). Referential scope and visual clutter in navigation tasks. In van Deemter K., Gatt A., van Gompel R., Krahmer E. J. (Eds.),*Proceedings of the Workshop on the Production of Referring Expressions (PRE-CogSci 2011)* (pp. 1–6). Boston, MA: Cognitive Science Society.

Wickens C. D., Nunes A., Alexander A. L., Steelman K. (2005). 3D navigation and integrated hazard display in advanced avionics: Workload, performance, and situation awareness display clutter issues in SVS design (AHFD-05-19/NASA-05). University of Illinois Human Factors Division technical report.

Wickens, C. (1994). Designing for situation awareness and trust in automation. In Proceedings of the IFAC Conference. Baden-Baden, Germany, pp. 174-179.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37*(3), 473.

Wickens, C. D., Hollands, J. G., Banbury S., Parasuraman R. (2013). *Engineering psychology & human performance* (Fourth international (cover).;4th; ed.) Psychology Press - M.U.A. doi:10.4324/9781315665177

Williams, C. C., & Pollatsek, A. (2007). Searching for an O in an array of Cs: Eye movements track moment-to-moment processing in visual search. *Perception & psychophysics, 69*(3), 372-381.

Wingfield, N. (2001, Dec 10). E-commerce (A special report): Cover story --- the people behind the sites: Amazon: Fighting the clutter. *Wall Street Journal* Retrieved from http://search.proquest.com/docview/398883564?accountid=8555

Wolfe J. M., Oliva A., Horowitz T. S., Butcher S. J., Bompas A. (2002). Segmentation of objects from backgrounds in visual search tasks. Vision Research, 42, 2985–3004.

Wuaters, R. (2011). Forrester: online retail industry in US will be worth $279 billion in 2015. Retrieved from http://techcrunch.com/2011/02/28/forrester-online-retail-industry-in-the-us-will-be-worth-279-billion-in-20 15/

Xun Huang, Chengyao Shen, Xavier Boix, Qi Zhao; The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 262-270

Yamani Y., McCarley J. S. (2010). Visual search asymmetries within color-coded and intensity-coded displays. Journal of Experimental Psychology: Applied, 16, 124.

Yang, C., Zhang, J., Xu, X., Chang, H., & He, G. (2007). Quaternion phase-correlation-based clutter metric for color images. *Optical Engineering, 46*(12), 127008. doi:10.1117/1.2823489

Yarbus A. L., Haigh B., Rigss L. A. (1967). *Eye movements and vision*. New York, NY: Plenum Press.

Yeh M., Brandenburg D., Wickens C. D., Merlo J. (2000). *Up or down? A comparison of helmet mounted display and hand held display tasks with high clutter imagery* (No. ARL-00-11/FED-LAB-00-3). Champaign: Illinois University at Urbana-Champaign, Savoy Aviation Research Lab. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA436770

Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding, and decluttering techniques.*Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 543-562.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological review*, *115*(4), 787.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel–serial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 244.

Zelinsky, G., & Neider, M. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. Perception, 40, 67-67.

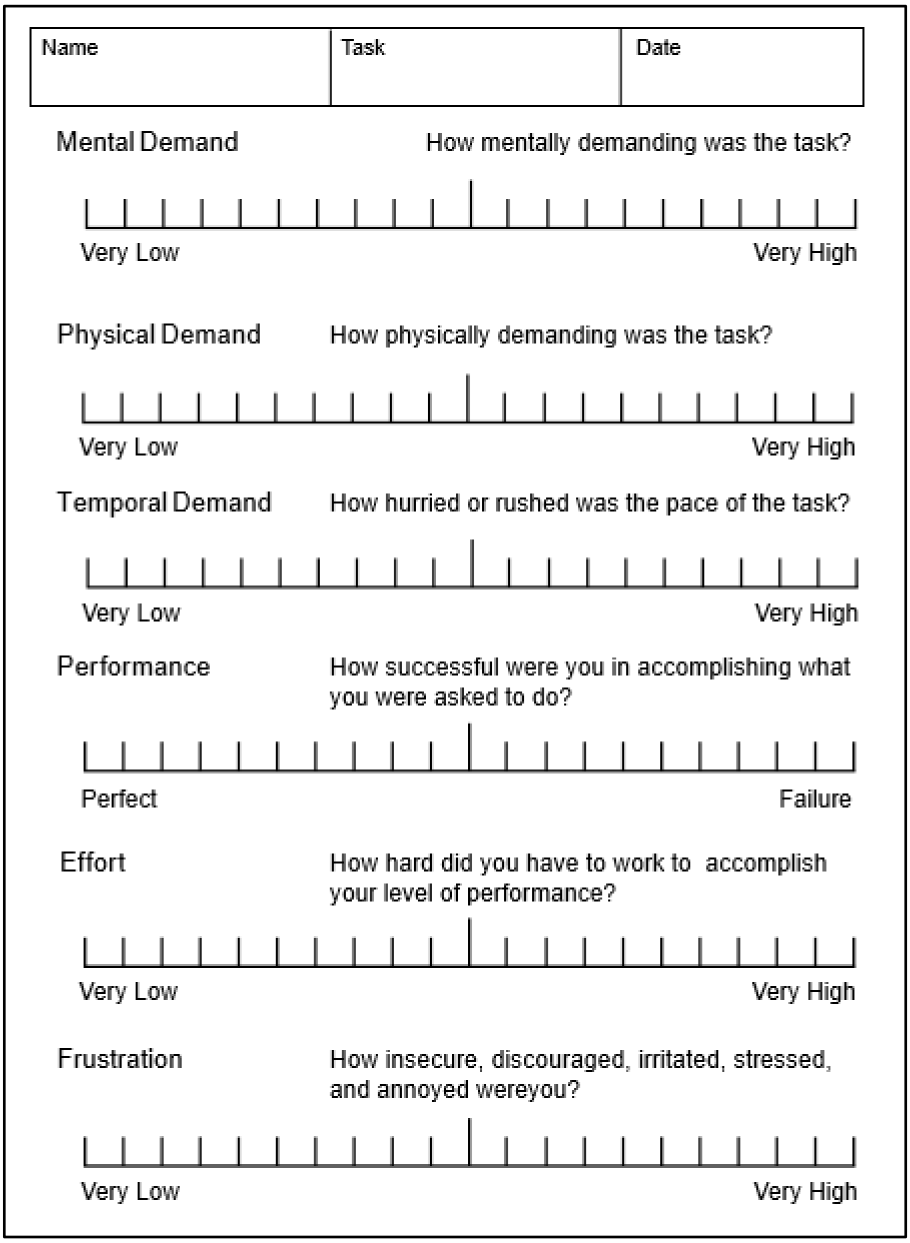
Zelinsky, G.J. (2001). Eye movements during change detection: Implications for search constraints, memory limitations, and scanning strategies. Perception & Psychophysics, 63, 209-225.

Zeng, Q., Cimino, J. J., & Zou, K. H. (2002). Providing concept-oriented views for clinical data using a knowledge-based system: An evaluation. *Journal of the American Medical Informatics Association, 9*(3), 294-305. doi:10.1197/jamia.M1008

Zhu, B., & Sun, X. (2012). Effects of Superposition on Oculomotor Guidance and Target Recognition. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (pp. 333-337). Boston, MA: Sage Publications. doi: 0.1177/1071181312561077

Appendix

I. APPENDIX A: NASA-TLX QUESTIONNAIRE



Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

II. APPENDIX B: POST-EXPERIMENT QUESTIONNAIRE

Subject ID [Filled out by experimenter]: ……………
Questionnaire Date and Time [Filled out by experimenter]: ……………

1. What is your age? ____________________

2. What year is this for you at AUB? __________________

3. What is your major? ___________________

4. How often do you use online shopping websites?
   - o Every day
   - o Once a week
   - o Once or twice a month
   - o A couple of times a year
   - o Never

5. Please **rank** the screenshots presented on the screen from least to most cluttered, based on your own definition of clutter.

Most Cluttered

Rank One:
Screenshot ________.
Rank Two:
Screenshot ________.
Rank Three:
Screenshot ________.
Rank Four:
Screenshot ________.
Rank Five:
Screenshot ________.
Rank Six:
Screenshot ________.
Rank Seven:
Screenshot ________.
Rank Eight:
Screenshot ________.
Rank Nine:
Screenshot ________.
Rank Ten:
Screenshot ________.

Least Cluttered

6. Please rate the amount of clutter you believe is in each image (put an X at the correct level):

Screenshot 1:

Very Low Very High

Screenshot 2:

Very Low Very High

Screenshot 3:

Very Low Very High

Screenshot 4:

Very Low Very High

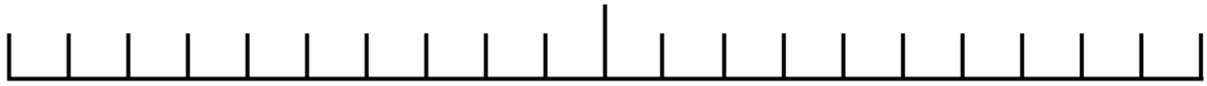
Screenshot 5:

Very Low Very High

Screenshot 6:

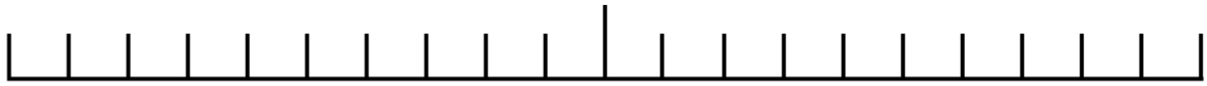
Very Low Very High

Screenshot 7:



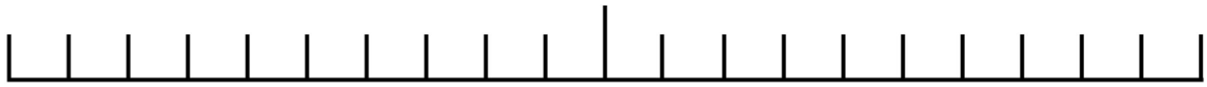
Very Low Very High

<u>Screenshot 8:</u>



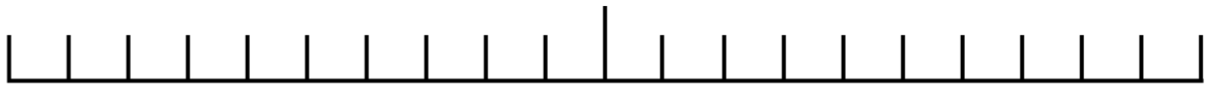
Very Low Very High

<u>Screenshot 9:</u>



Very Low Very High

<u>Screenshot 10:</u>



Very Low Very High