UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Web Science

The Influence of Hyperlinks on Reading on the Web: An Empirical Approach.

by

Gemma Fitzsimmons

Thesis for the degree of Doctor of Philosophy

July 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Web Science

Thesis for the degree of Doctor of Philosophy

THE INFLUENCE OF HYPERLINKS ON READING ON THE WEB: AN EMPIRICAL APPROACH

Gemma Fitzsimmons

We increasingly spend a vast amount of time on the Web and much of that time is spent reading. One of the main differences between reading non-Web based text and reading on the Web is the presence of hyperlinks within the text, linking various related Web content and webpages together. Some researchers and commentators have claimed that hyperlinks hinder reading because they are a distraction that may have a negative effect on the reader's ability to process the text. However, very few controlled experiments have been conducted to verify these claims. In the experiments documented here we utilise eye tracking as a new methodology for examining how we read hyperlinked text. During reading we move our eyes in order to bring new information into our fovea where the highest visual acuity is present. There is a well-documented tight link between when and where we look and what we process. By measuring eye movements, we can gain insights into the ongoing cognitive processing that is occurring during a task. Eye movements have been used extensively to help us to understand the cognitive processing that occurs during reading, but there has been very little research into how our reading differs when we read information on the Web. Therefore, in this thesis we examine the influences of hyperlinks on reading on the Web.

Contents

List of Tables

List of Figures

Academic Thesis: Declaration Of Authorship

I, *Gemma Fitzsimmons* declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

The Influence of Hyperlinks on Reading on the Web: An Empirical Approach

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Either none of this work has been published before submission, or parts of this work have been published as:

Fitzsimmons, G., Weal, M. J. & Drieghe, D. (2014) Skim Reading: An Adaptive Strategy for Reading on the Web. In proceedings, of the 6th Annual ACM Web Science Conference held in Bloomington, IN, 23-26 June, (pp. 211-219).

Fitzsimmons, G., Weal, M. J. & Drieghe, D. (2013) On Measuring the Impact of Hyperlinks on Reading. In proceedings, of the 5th Annual ACM Web Science Conference held in Paris, France, 2-4 May, (pp. 65-74).

Signed: ……………………………………………………………………

Date: ……………………………………………………………………

Acknowledgements

"Ah, there's nothing more exciting than science. You get all the fun of...
sitting still, being quiet, writing down numbers, paying attention...
Science has it all."

Principal Skinner, 1995

To all of those who have been there for me during my PhD, for the programming strops and the lunch breaks and the hours spent standing in corridors waiting for participants. The future came and went and seemed to last forever and for a moment, but now the thesis is complete and I have many people to thank.

I would like to thank my supervisors Dr. Denis Drieghe and Dr. Mark J Weal for their help, support and valuable feedback throughout this project. I would also like to thank Dr. Tamaryn Menneer for the feedback during my annual progress report and Dr. Graeme Fairchild for his feedback during my upgrade. I would also like to thank Dr. Hayward Godwin for his assistance with programming and encouraging me into the field of Psychology. I would like to thank my colleagues in Finland, Dr. Johanna Kaakinen and Prof. Jukka Hyönä, for their time and assistance during my visit to the University of Turku in 2014 and my thanks also go to Hanna Martikainen for her valuable help with data collection for my final experiment.

I would like to thank my colleagues in Psychology and in Web Science for making my time as a PhD student an interesting and fun experience. I especially want to thank Charlotte for always being there and for her emotional support. I thank Lewis for his assistance with stimuli, proofreading and for making me laugh when I needed it most. Thank you to Olly and Mira for their continued support and copious amounts of pizza. Thank you to Vincent for all of the support, wine and much needed encouragement to finish this beast. Thank you all for making this time forever memorable.

This research was funded by the Research Councils UK Digital Economy Programme, Web Science Doctoral Training Centre, University of Southampton. EP/G036926/1.

Chapter 1

The Influence of Hyperlinks on Reading on the Web: An Empirical Approach

"The basic point I have to make is almost embarrassingly simple: hypertext is very different from more traditional forms of text. The differences are a function of technology, and so various, at once so minute and so vast, as to make hypertext a new medium for thought and expression - the first verbal medium, after programming languages, to emerge from the computer revolution…A new medium involves both the new practice and new rhetoric, a new body of theory."

Slatin, 1990

General Introduction: Reading Hypertext

The Web has had a long-lasting and transformative effect upon society. In the first quarter of 2016, 87.9% of adults in the UK (45.9 million) had, in the last 3 months, used the internet, compared with 86.2% in 2015 (Office for National Statistics, 2016). The internet was accessed every day, or almost every day, by 78% of adults (39.3 million) in Great Britain in 2015, compared with 35% (16.2 million) in 2006 (Office of National Statistics, 2015). The Web has opened up and continues to open up access to information and free exchange with other people across the globe. It has a substantial impact upon the global economy, and has a profound influence upon the lives of countless numbers of people. Users of the Web engage in a wide variety of different activities, such as searching for information, reading the news, sending and receiving email, online shopping and social networking, to name just a few (Office for National Statistics, 2015). Arguably, within the majority of these activities, the primary task that users engage in is reading text in some shape or form. Although there has been extensive research examining reading behaviour (Rayner, 1998, 2009), which will be discussed here, reading online text (termed hypertext:

Nelson, 1965) has some important differences to reading plain text that is not in a Web environment (such as a printed novel). One of the main differences between online reading and offline reading is the potential presence of *hyperlinks*. These hyperlinks, when clicked with the computer's mouse or finger on a touch screen, enables users to navigate from one webpage to another. Hyperlink documents consist of nodes and links which form the basis for linking one object to another. Nodes are connected to other nodes by links.

Hyperlinks are prolific online, either embedded in the text or listed in a sidebar as part of a navigation system (for an example see Figure 1.1). Hyperlinks are distinguished from the rest of the onscreen text by being displayed in a different colour (often blue) and are often underlined. Any piece of text or image can be made a hyperlink that once clicked will display the appropriate linked content. As the following Literature Review will highlight, the research that has been conducted to date has only scratched the surface of how we read hypertext. The primary goal of the present thesis is therefore to develop a theoretical framework explaining how we read, interact with and process hypertext at a fundamental level.

The goal of examining how we read, interact with and process hypertext is motivated by both practical and theoretical interests. Looking at real world tasks can inspire theoretical questions that serve as a motivation for novel empirical questions. To date, previous research that has examined reading behaviour has focused mainly on reading for comprehension, and the theoretical and computational models of reading typically model the reading of a single line of text rather than passages of text. The majority of this work also tends to focus on lexical processing rather than higher level issues such as discourse processing or task effects (how different instructions influence how people perform a task). Exploring how people read hypertext would explore these higher level issues such as task effects (by comparing it with different tasks) and how the high level information of knowing that a hyperlink links a word or phrase to another piece of content affects the reader. Hypertext and the Web therefore offer an ideal task/stimulus space for extending existing models and theories by allowing for an examination of those models and theories under new, but related conditions. Furthermore, by studying how people read hypertext on the Web, it will then be possible to generate suggestions for how to design and setup webpages

Figure 1.1 Screenshot of a webpage with hyperlinks (shown in blue) embedded in the main text content. The navigation on the sidebar also consists of hyperlinks.

in a way that most suits the demands and limitations of the human information processing system.

The experiments presented in this thesis will focus on considering what is known in terms of reading behaviour and will then explore how reading behaviour is modulated by reading and interacting with webpages. There are a number of differences between reading linear text documents and hypertext documents beyond the technological differences. Some of the main differences will be explored during this thesis. Firstly, hyperlinks are salient compared to the rest of the text on a webpage: given their colour, they 'stand out' from the rest of the text, and can be easily distinguished from the rest of the textual content (this is explored in Chapter 2). Hyperlinks are salient and this can suggest and signal to the reader that the pieces of hyperlinked text are more important or special in some way. We explore the importance of hyperlinks and other factors in Chapter 3 in order to understand the importance readers may put on hyperlinked words on the Web. Furthermore, readers do not just read for comprehension when using webpages. They engage in many different activities, such as searching for information or simply skim reading the

content rather than reading all the information in depth (explored in Chapter 4). Hyperlinks also represent a navigational link to another piece of content elsewhere on the Web. This means that the user needs to make a decision when reading hyperlinks. Specifically, they need to decide whether to follow the hyperlink or whether to continue reading (this issue is explored in Chapter 5 and 6).
Eye movements are an important methodological component of reading research. Therefore, this thesis will begin by explaining the methodology of eye movements and why they are instrumental in helping us understand how people process visual information. I will then describe previous research that has focused on eye movements and reading behaviour, and finally develop a grounding for the (mostly) eye movement based experiments presented here by outlining previous research that may be able to help shed light upon the differences between previous reading research and reading on the Web.

Eye Movement Methodology

Eye movements were initially assumed to offer no insight into how we process information. In the case of reading, it was assumed that the eyes simply glided across the text. However in 1879, Professor Emile Javal observed that a reader's eyes do not move smoothly across the text, but the eyes make a series of short movements with stationary pauses in between (reported in Huey, 1968). After this observation was made many questions were asked about when, where and why we move our eyes during reading. The answers to these questions can tell us about how we are processing the information we see.

The biology of the eye and how this affects eye movements

When we make eye movements, the movements themselves are called saccades. In between these saccades our eyes are relatively still, which is called a fixation. We take in information during fixations and vision is said to be suppressed during saccades to avoid seeing a blur or smear (Matin, 1974). Saccades are necessary due to the anatomy of the eye and more specifically the retina. The retina

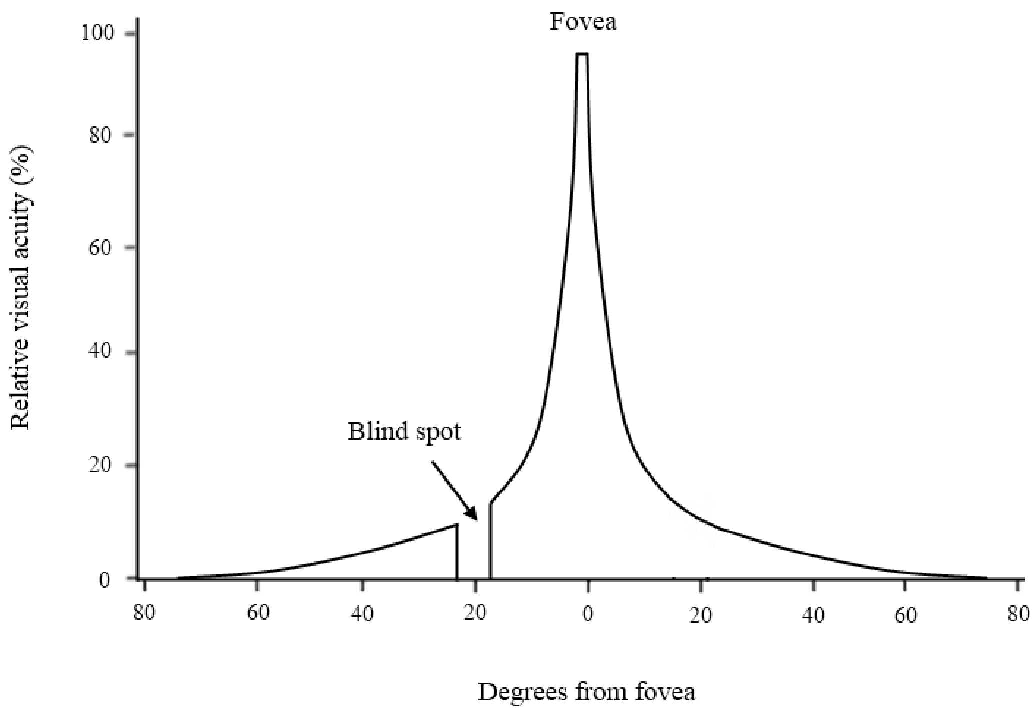


Figure 1.2 Visual acuity as a function of position on the retina. The blind spot is where the optic nerve leaves the eye to connect to the brain, hence why there are no photoreceptors present there.

contains a very high concentration of photoreceptors called rods and cones. Rods are particular useful for seeing in low-light levels and detecting motion. Cones are necessary for high visual acuity. The majority of the cones in the retina are in an area called the fovea which covers about 1° degree of visual angle on either side from the fixation point. Moving beyond the fovea, there is a large reduction in the number of cones and therefore a high acuity drop off (see Figure 1.2). Readers can still obtain some visual information from the parafovea (an area which extends to about 5° degree of visual angle either side of the fovea), but it does not have a level of acuity as high as the fovea. Beyond the parafovea is peripheral vision which has low acuity and readers may only be aware of relatively low level information, such as word length and spaces during reading.

In order to read we need the words to be on the high acuity fovea to be able to process them and therefore we need to move our eyes so that the fovea can be utilised to take in the most and best visual information while reading. Rayner and

Bertera (1979) conducted a reading study where the fovea information was masked and this made reading very difficult for the participants and they had difficulty reporting what they had read after the task because they could not pick up enough information to identify the words correctly. There is considerable variability in how far a saccade moves during reading (average 7-9 characters, range 1-15 characters) and how long a fixation lasts (average 200-250ms, range 100-500+ ms). This variability does not only occur between individuals, but also within an individual's own eye movements. Much of this variability can be related to how difficult the currently fixated text is to process (for reviews see Rayner, 1998, 2009) as it has been observed that the more difficult the text is the longer the fixations are and the shorter the saccades and there are also more regressions (eye movements going back to previously inspected text) being made (Rayner, 1978).

Eye tracking technology

In 1879, Professor Emile Javal observed individuals' eyes while they read, but in order to measure eye movements more accurately an eye tracker needed to be used. An early eye tracker was created by Edmund Huey, which resembled a contact lens with a hole in the centre for the pupil. The lens was connected to a pointer which moved when the individual wearing the lens moved their eye (Huey, 1968). However, this methodology was quite intrusive and lacked precision. Charles Judd developed the first eye movement camera which recorded eye movements by recording light reflected from the eyes of the participants. This technology was used by Guy Buswell to analyse eye movements during reading across different age groups and schooling (Buswell, 1922). The technology used now is very similar, but much more accurate. Naturally, eye tracking hardware and software has increased in sophistication thanks to the development of computer systems and camera equipment. The eye tracker used for the experiments within this thesis is the EyeLink 1000[1] (see Figure 1.3), which records the eye position every millisecond and has an accuracy of 0.25º - 0.5 º of visual angle. By using infrared light directed at the participant and a camera to record the light reflected back, the eye position can be calculated. The eye position is

[1] Eyelink 1000 developed by SR Research Ltd., Ontario, Canada.

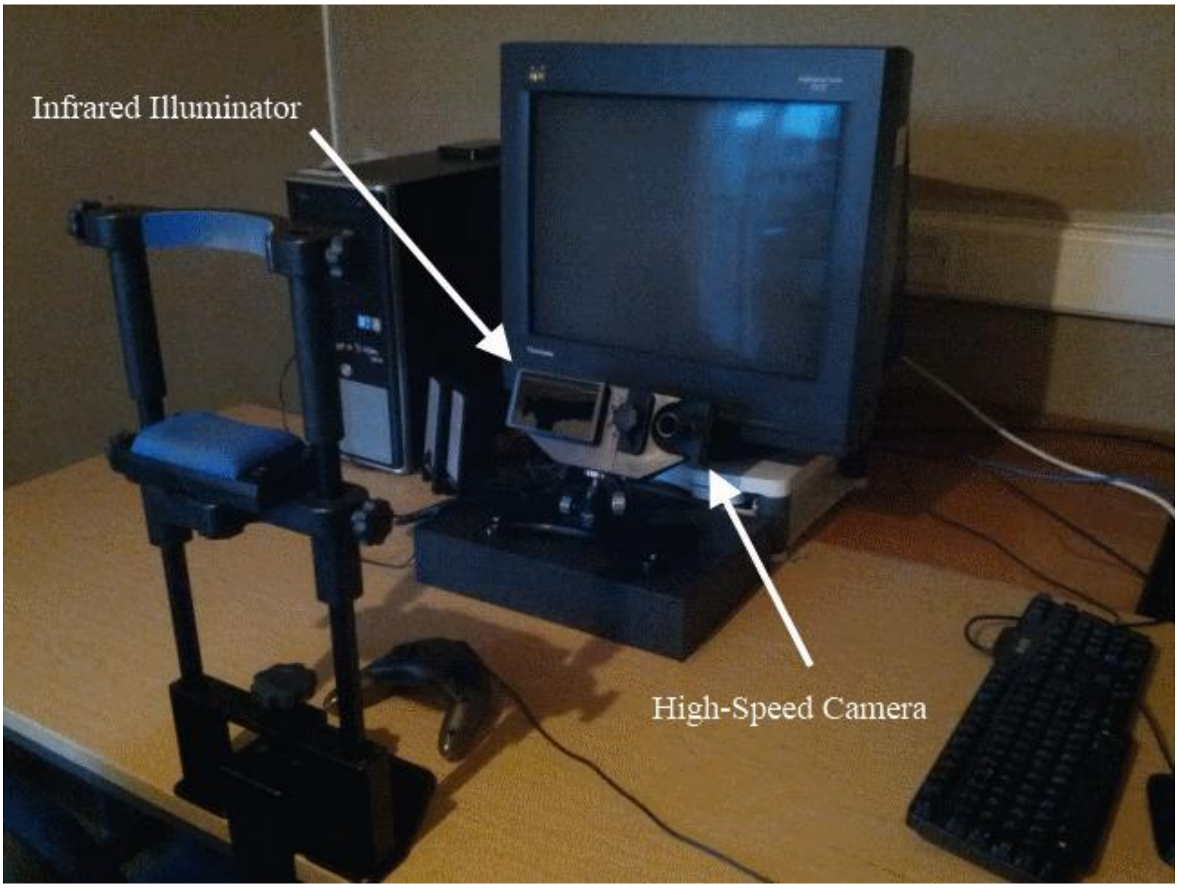


Figure 1.3 The EyeLink 1000 eye tracker attached to a 23 inch CRT monitor. The chin/head rest is adjusted to an appropriate height for the participant and they use the controller to make responses throughout the experiment.

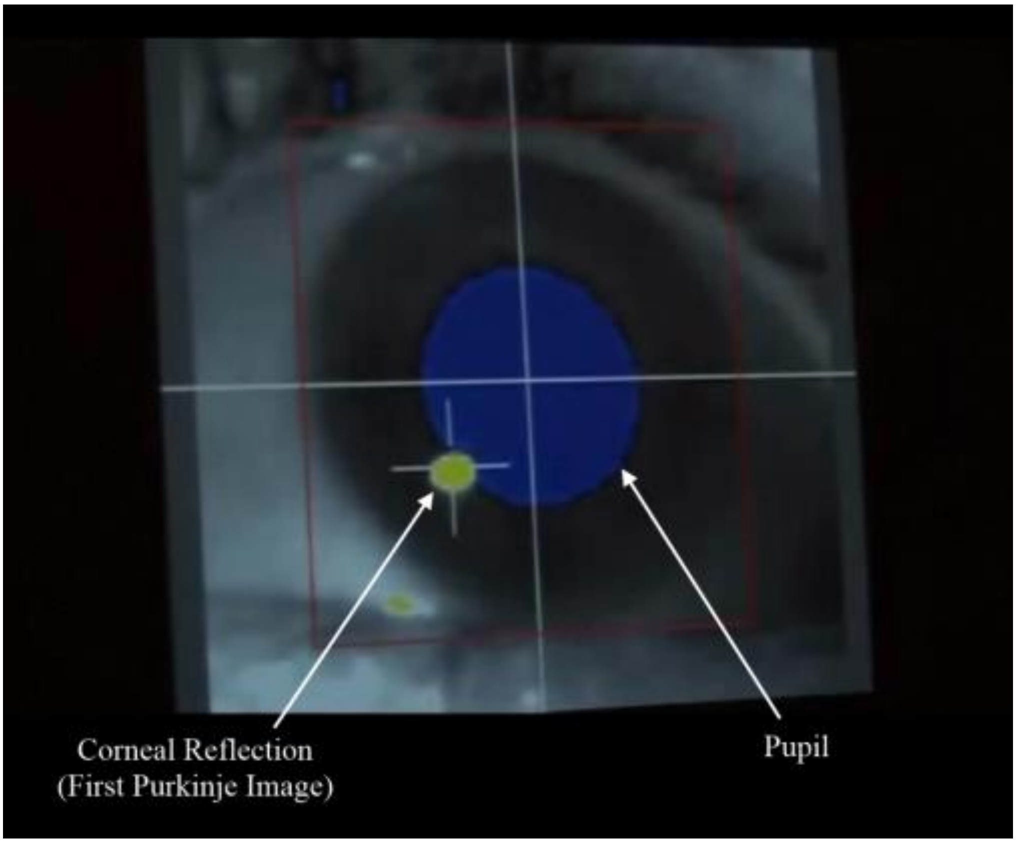


Figure 1.4 An image of an individual's eye and the infrared light reflected back to the camera of the eye tracker. The large blue circle is the pupil and the small yellow circle with the cross hair is the corneal reflection (the first Purkinje image, which is the reflection from the outer surface of the cornea). By using infrared light directed at the participant and a camera to record the light reflected back, the eye position can be calculated.

Figure 1.5 An example scanpath from Data Viewer (software used to visualise and process EyeLink data) showing the fixations as turquoise circles that increase in size as fixation duration increases. The saccades (in orange) are movements between the fixations and they are denoted by arrows indicating the direction of the saccade. The numbering of fixations is to clarify order.

calculated by asking the participants to look directly at points on the screen and by recording both the position of the pupil and the position of the corneal reflection (the first Purkinje image, which is the reflection from the outer surface of the cornea) when looking at these points (see Figure 1.4). An algorithm is then used to extrapolate eye positions across the entire display.

The example scanpath shown in Figure 1.5 shows short movements (saccades) and pauses (fixations) where the eyes are still and taking in information. The turquoise circles denote a fixation and the larger the circle, the longer the fixation on that position. Saccades are denoted by the orange arrows indicating the direction of the saccade. The contents of the text being read makes a significant difference to when and where the eyes move, indicating that eye movements are linked to our online cognitive processing. Note that the term "online" used in the context of eye movement research refers to the current and ongoing processing occurring, not to be confused with the term when used to refer to a computer being connected to the Web. Due to the topic of this thesis this results in the addition of an extra level of confusion.

Therefore, when addressing how anyone is online/connected to the Web, it will be referred to as using the Web. However, how eye movements are linked to online cognitive processes will always be referred to in the context of current and ongoing processes. This is discussed further in the next section.

Eye movements are linked to current and ongoing cognitive processes

The cognitive processes involved in performing a specific task can be difficult to determine. One way to find out would be to ask the individual performing the task. However, there are a number of problems with this method. Firstly, the experimenter would have to wait until after the task had finished or risk interrupting the task and confounding the results. Another problem is that individuals have a tendency to rationalise their behaviours, making self-reports unreliable (Ericsson & Simon, 1993). Recording eye movements is a more objective way of collecting data about behaviour and a number of studies have shown that eye movements provide an unobtrusive, real-time behavioural index of visual and cognitive processing (Liversedge & Findlay, 2000; Rayner, 1998, 2009). It has also been observed that when eye movements are absent in an individual their visual perception can be surprisingly normal. For example, Gilchrist, Brown and Findlay (1997) examined an individual referred to as AI who was born without the ability to make eye movements and who compensates for this lack of eye movements by making saccadic-like movements of the head. This suggests that saccadic movements of the eye, or head in this case, form an optimal sampling method for the brain. The head movements made by AI were present not only in reading, but also in visual scene viewing, orientating her head movements to locations of interest in the scenes. This kind of orientating of eye movements to locations of interest in a scene was first observed by Yarbus (1967) showing that different eye movement patterns were observed dependent on the task being completed. Participants viewed a painting by Repin entitled “An Unexpected Visitor” that depicts a family in the room with another person entering the room. The participants were given different instructions while observing the picture and having their eyes tracked. A different pattern of eye movements was observed for the different instructions. For example, when asked to remember the clothes worn by the people in the room, the participants focused on the people in the room and looked at the whole person to observe all of their clothes. However, when asked to estimate the age of the people in the room, the participants tended to focus on the faces of the people in the room. This suggests that the eye movements are driven by the task at hand. Together, this evidence, along with many other studies (Rayner, 1998, 2009) points to the notion that eye movements can be used to gain an insight into cognitive processes.

Analysis of eye movement data

When analysing eye movement data there are a number of issues to consider. Although previous research has shown that eye movement data gives us a good indication of moment-to-moment cognitive processes during reading (Liversedge & Findlay, 2000; Rayner, 1998, 2009), simply averaging fixation durations will only give us a global measure of the processing that is taking place. For instance, not all words are fixated during reading, in fact a third of all words are skipped, with function words skipped more frequently than content words (for review see Rayner, 1998; Rayner 2009). Additionally, longer words may have more than one fixation on them. Thus, if readers only fixated each word once, average fixation durations would be an appropriate measure to use, but due to the issue of skipping and multiple fixations on a single word, other local measures are used to provide a reasonable estimate of how long it takes to process each word.

The local measures are normally conducted on a target word region within the sentence (or sometimes a region which comprises a fixed set of multiple words). An experimental sentence will be created and it often contains a target word that is manipulated between conditions. For example, if the experimenter wants to explore the impact of word length, the target word might be a 3 letter short word such as "*cat*" or a 7 letter long word such as "*panther*". These target words could both fit into the sentence "The man saw a large *cat/panther* hiding in the woods." In order to explore the impact of the target word manipulation we can take the fixations from the region where the target word is located and analyse a number of local eye movement measures.

These local measures are based on the first-pass forward reading of the word (the first set of forward saccades through the sentence, excluding re-reading fixations).

There are a number of eye movement measures that have been developed to investigate and understand reading behaviour (Rayner, 1998), and a number of relevant measures are listed below.

- *Skipping probability* is the probability that the target word was skipped (i.e. did not receive a direct fixation on it) in first-pass reading. Skipping rates are influenced by the ease of processing a word. If a word is very easy to process than it will be skipped more often in the first pass of reading (Rayner, Slattery, Drieghe, & Liversedge, 2011).
- *First fixation duration* is the duration of the first fixation on a word. It does not matter if the target word was fixated once or more than once, only the first fixation is used. This measure is used to examine the early effects of the target word on eye movements. If first fixation durations are short then the word is easier to process than if the first fixation durations are long (Rayner, 1998).
- *Single fixation duration* is where the reader made exactly one first-pass fixation on the target word. Single fixation durations are similar to first fixation duration in that they show the ease of processing the target word (Rayner, 1998). However, they only include occasions where the target word was fixated once in the first pass of reading, giving a good estimate of the difficulty or ease in processing that particular word.
- *Gaze duration* is the sum of all first-pass fixations on the target word before moving to another word. Longer gaze durations indicate that the participant is having difficulty processing the information and may make more than one fixation of the target word if needed (Rayner, 1998).
- *Go-past time*, which is the accumulated time from when a reader first fixated on the target word until their first fixation to the right of the target word. This measure includes any regressions the reader made before moving forward past the word. Go-past times are often used to examine when the reader has difficulty integrating a word and has to go back and re-read the preceding content, thus increasing the go-past times measure.
- *Total reading time* is the sum of all fixation durations on the target word regardless of whether this reading happened during first pass or later. This measure examines any disruption seen in reading the target word because it includes all fixations on the target word during the trial. Again, if the total

reading time is longer in one condition compared with another it suggests that the text in that condition was more difficult to process.

- *Regressions* during reading occur when the eyes move backwards through the text, back to previous words/sentences/sections of text. Most saccades when reading English move from left to right in the direction of the text, but readers do not always move forward in the text. 10-15% of saccades are regressions, meaning they move from right to left during reading. Most regressions are only a few letters long and occur when the reader has made too short or too large a saccade and moves their eyes back to a more optimal position for recognising a specific word. (Rayner, 1998). Some regressions are longer and may move the eyes back several words or to another line. This occurs when the reader is having difficulty understanding or parsing the text being read (Frazier & Rayner, 1982; Murray & Kennedy, 1988). Poor readers engage in more regressive eye movements in order to re-read the text (Murray & Kennedy, 1988). Regressions can be split into two types of regressions, regression in and regressions out. Regressions in are those that occur when the target area has been left via a forward saccade and then a backwards saccade is made with a fixation in the target area, so the reader has left the target region and then decided to re-visit it. Regressions out are those that originate in the target region and then a backward saccade is made to visit regions preceding the target region.

These measures are listed above roughly according to their position in the time course of processing a sentence. Some of these measures are considered early measures and some are considered late measures. Early processing measures, which are sometimes referred to as first-pass measures, include skipping probability, first fixation duration and gaze duration. These are the measures that explore the effects of any manipulation on the first-pass of reading. Later processing measures such as go-past times, total times and regression rates are classed as late measures because they include the re-reading that occurs after first-pass.

If a difference between conditions is observed in first-fixation duration, i.e. a short word has shorter first fixation times on it than a long word; then the effect is considered an early effect because the reader has shown that they spend less time processing a short word when they first fixate. However, if there is a difference between conditions that is observed in go-past times, i.e. a short word has shorter go-past times than a long word; then the effect is considered a late measure because it includes re-reading of the text before moving to the right of the target word. Additional re-reading could suggest there is a problem with integrating and understanding the content of the text. The approach of explaining these measures in terms of their time course is an accepted practice within eye movement and reading research.

Task effects

Reading for comprehension is not the only task someone can engage with while processing visually presented stimuli, and with each task a different pattern of eye movements is typically observed. Much of the research into eye movements and reading focuses on reading for comprehension. However, on the Web the readers may need to browse or search for specific information as opposed to reading the entire text for comprehension. For example, Shrestha and Lenz (2007) observed participants' scan paths of webpages that contained text and/or pictures while completing browsing and searching tasks. They found participants searched in an 'F' shaped pattern. This 'F' shaped pattern was first observed by Nielsen (2006) whereby users focused on the left hand side of the webpage making longer horizontal scans at the top of the page and longer vertical scans down the page on the left hand side later (resulting in a pattern similar in shape to the letter F). However, this cannot be generalised to reading for comprehension where readers need to read the majority of text on the screen, not just browse it. This makes it difficult to assume that the same findings will be found for both tasks.

The task at hand will affect the way users engage with the webpage and their eye movement behaviour. We again refer to the seminal work reported by Yarbus (1967) who used the famous painting by Repin entitled "An Unexpected Visitor" and gave participants different tasks to complete while observing the picture and having their eyes tracked. A different pattern of eye movements was observed with each different task. For example when asked to give the ages of the people within the picture eye movements were focused on the faces of the family. However, when asked

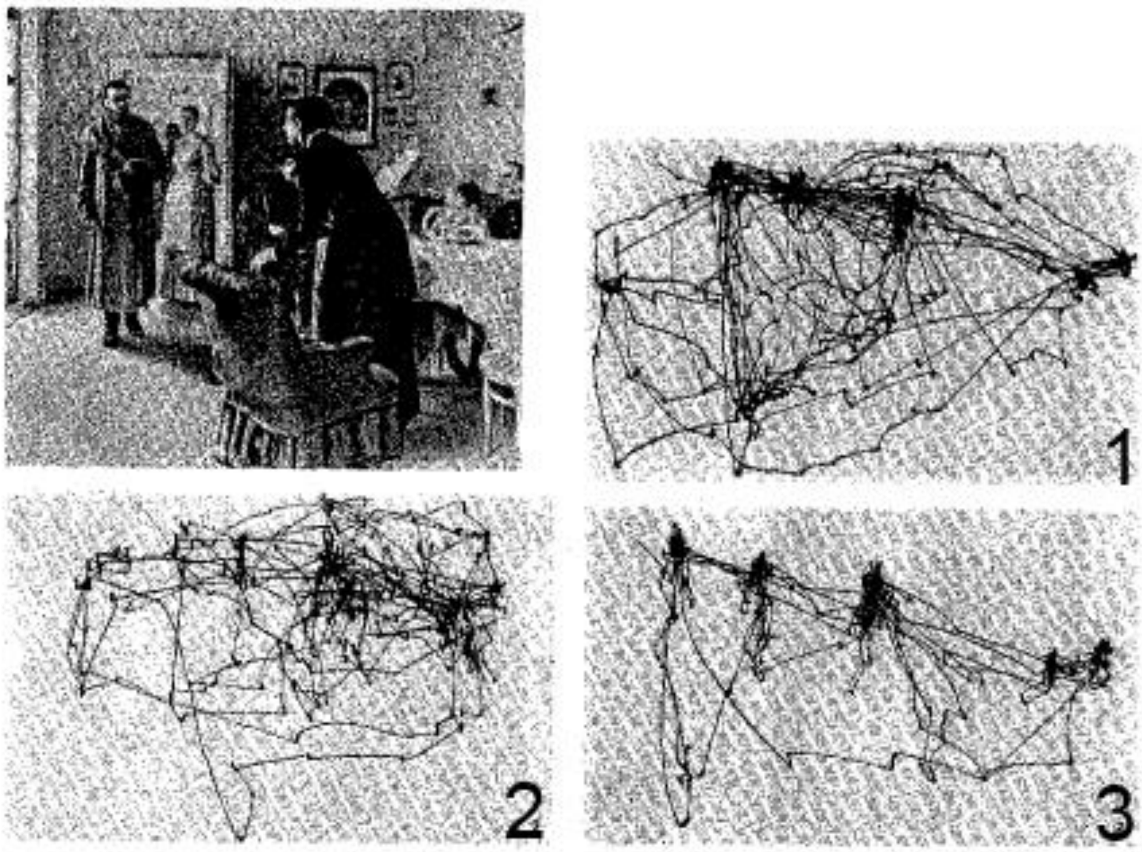


Figure 1.6 Scanpaths from one participant on the painting "An Unexpected Visitor" from Yarbus (1967). 1) A scanpath during free examination of the picture. 2) A scanpath when the participant was asked to estimate the material circumstances of the family before viewing. 3) A scanpath when the participant was asked to estimate the ages of the family members in the painting.

to estimate the material circumstances of the family in the picture, fixations were made on the family, their clothes and the objects around the room (see Figure 1.6). This demonstrates that eye movements are not made arbitrarily and are dependent on the demands of the task at hand. Rayner and Raney (1996) found that when searching for a target word in a passage of text participants made longer saccades and the average fixation durations were shorter than those observed during reading. Also, a frequency effect seen in normal reading, where fixation times are shorter on high frequency words compared to low frequency words (Henderson & Ferreira, 1990; Inhoff & Rayner, 1986; Rayner & Fischer, 1996), was not observed in the visual search task (where the participant had to search the text for a certain word and say if it was present or absent, but did not have to comprehend the text). When participants were reading the text, they were more likely to skip the high frequency words and had shorter fixation durations on the high frequency words compared to the low frequency words. However, when asked to search the text the participants showed no differences between the high and low frequency words in eye movements measures. Both findings indicate that during the visual search tasks the words were not lexically processed up to the same extent as during the reading for comprehension task. The suggested explanation for this is that the meaning of the text is not, or at least to a lesser extent, important during a search task.

These task effects are also present when the reader has to read from a certain 'perspective'. For example, participants might be instructed to read a passage of text about a house that included information such as the side door was always unlocked and that the roof of the house leaked. The participants could be asked to read from the perspective of a burglar or as a person looking to buy a house. The participants' perspective influences what parts of the text the reader focuses on. The perspective adopted by the reader affects the ongoing processing of text by guiding the readers to invest extra processing time on relevant information, resulting in better memory for relevant than for irrelevant information (Anderson, 1982). So readers who had to read from the perspective of a burglar would focus more on the details about the security of the house and the expensive contents and the homebuyer would focus on the house infrastructure. This suggests that the strategy for processing text is affected by the readers' goal and task at hand.

Although people do engage in a number of different tasks on the Web such as searching for information, reading the news, sending and receiving email, etc., these tasks generally involve reading. Therefore, in the next section of this literature review, I will address the current research on eye movements and reading. I will also include research exploring how reading on the Web is different from reading plain text exploring issue such as the saliency of a hyperlink, the signalling of content, skim reading and foraging for information in a hypertext environment.

Reading for comprehension

This section of the literature review covers relevant research on reading. First, research will be explored that outlines the eye movements and reading for comprehension literature. Followed by, the research specific to reading on the Web. The Web contains hyperlinks which make it a different reading experience compared to reading plain text. One of the important issues to consider is the saliency of hyperlinks and the convention that hyperlinks are coloured which will impact the saliency. Previous research relating to these issues will be explored. Next, I will explore signalling research and how important signals, such as text demarcation (bold, underline, etc), in the text might influence the reader. This is especially important to understand considering that hyperlinks themselves act as salient signals to the reader. Then I will move onto discussing models of comprehension which relate to how individuals comprehend the text content. Finally, research on skim reading and information foraging will be explored. The current literature suggests that people reading on the Web often engage in skim reading (Liu, 2005; Morkes & Nielsen, 1997). It is therefore important to explore this research in order to understand how people read on the Web.

Eye movements during reading

This section focuses on the theoretical background of the research presented in this thesis, mostly exploring cognitive processes during reading. The main methodology used for examining the theoretical questions in this area is eye tracking, which has been explored in-depth in the previous section.

Although reading could be said to be one of the most fundamental tasks we engage in while using the Web, thus far very little research has directly investigated how the presence of hyperlinks influence reading behaviour on the Web. This is an important point considering the large number of hyperlinks on webpages throughout the entire Web. It is therefore vital that this fundamental aspect of Web use be investigated in detail. It is especially surprising that so little research has investigated the presence of hyperlinks during reading on the Web considering the extensive research in the field of eye movements and reading behaviour which will now be described in detail.

The eye movement methodology has been used extensively to investigate how individuals process text during reading. There are a number of factors to consider that influence when and where we move our eyes and investigating these factors can

```
1) The cat sat on the mat by the door.

2) xxx xxx xxx xx xxx mat xx xxx xxxx.
                       *
3) xxx xxx xxx xx the mat by xxx xxxx.
                       *
4) xxx xxx xxx xx the mat by the door.
                       *
```

Figure 1.7 Example of the gaze-contingent moving window paradigm developed by McConkie and Rayner (1975). The asterisk below the sentence represents a fixation. 1) The sentence without a moving window. 2) The sentence with a small symmetrical moving window. 3) The sentence with a larger symmetrical moving window. 4) The sentence with a large asymmetric moving window.

tell us a great deal about how text is processed during reading. Although visual acuity drops off symmetrically moving further away from the fovea, the area from which information is picked up and utilised during reading is not symmetrical. This asymmetry shows the effect of attention rather than the biology of the eye and it is called the perceptual span, which is defined as the region from which the reader picks up useful information. The perceptual span in readers who read left to right is about four characters to the left and 15 characters to the right of a fixation (McConkie & Rayner, 1976). The perceptual span was measured by the use of a gaze-contingent moving window paradigm developed by McConkie and Rayner (1975). This paradigm involves controlling what is seen during each fixation so that only a pre-defined area of information around the fixation is visible to the reader (see Figure 1.7). McConkie and Rayner also manipulated the size of the moving window and the participants' eye movements were recorded and monitored. When the moving window was reduced below 13 characters fixation times increased suggesting that the reader was having difficulty reading normally. The moving window size was manipulated and the perceptual span is the size that participants could read normally with no disruption. The effect is reversed in those readers who read right to left (Hebrew readers, for example), showing a perceptual span skewed to the left instead (Pollatsek, Bolozky, Well, & Rayner, 1981). This demonstrates that not all readers pick up information to a similar degree from the right and left of the visual field. Instead it demonstrates that readers are biased to pick up information in the direction of where the specific language presents novel information.

Where we move our eyes

Research has shown that the decision of where to move the eyes is largely driven by low-level cues such as word length and space information. If the word to the right of fixation is a long word of 9 characters, the next saccade will be longer than if it were a medium sized 5 character word (for a review, see Rayner, 1998). This reflects the tendency of the system to land on a position in the word that is optimal for recognising this word, which will be more character positions into a long word compared to a short word. Also spaces between words are used as cues to where the word ends and can be used to help target saccades. When spaces are removed reading speed is reduced substantially by as much as 50% (Morris, Rayner, & Pollatsek, 1990). Having spaces between words gives the reader useful information about the length of upcoming words. Rayner (1979) found that readers have a preferred viewing location (PVL) that falls halfway between the beginning of the word and the middle of the word. If the reader does not fixate the PVL then they are more likely to refixate that word and increase reading time. However, this can vary as a function of the launch site of the saccade. The further away the saccade is launched, the landing position moves further to the left and the closer the saccade is launched to the word the landing position moves further to the right (McConkie, Kerr, Reddix, & Zola, 1988).

In contrast to the PVL, the optimal viewing position (OVP) represents the location in a word where it can be recognised in the quickest time. The OVP is in the centre of the word. If the OVP is not fixated then there is a greater chance of refixations on the word and also there has shown to be a processing cost if the OVP is not fixated. For every letter the eye deviates from the OVP, a 20 ms cost takes place on average (O'Regan, Lévy-Schoen, Pynte, & Brugaillère, 1984).

As already mentioned, not every word is fixated during reading, some words are skipped. Research suggests that skipped words are processed up to a certain extent in the parafovea (the area outside the fovea where visual information can still be utilised). Fisher and Shebilske (1985) found that skipped words must undergo some degree of processing, even though they are not directly fixated. They conducted an experiment where the participants' eyes were tracked during reading. The words that were skipped were then deleted and another participant tried to read the passage. However, with the missing words it proved very difficult, therefore the skipped words must have been processed by the original participants, even though they were not fixated. A number of factors affect whether a word is skipped or not. Research has shown that low-level factors such as launch site (the location the saccade originated from prior to the skipping of or the fixating on the target word) have an effect on skipping rates. The closer the fixation is to the parafoveal word, the more likely it will be skipped due to increased parafoveal vision from a close-by launch site. (Rayner, Sereno, & Raney, 1996). In addition, word length is very important because it is a strong predictor of whether a word will be skipped or not, with short words being skipped more often than long words (Brysbaert, Drieghe, & Vitu, 2005; Rayner & McConkie, 1976). However, it is difficult to pull apart the reason why short words are skipped more often. Short words may be skipped more often because they are shorter and visually easier to process fully in the parafovea. Alternatively, short words could be skipped more often because they tend to be more frequent and are therefore easier to process.

Ease of processing has been shown to influence word skipping ratios, as seen in experiments examining the effects of more high-level linguistic factors on word skipping. Highly predictable words are skipped more often than those that are not predictable (Balota, Pollatsek, & Rayner, 1985; Rayner & Well, 1996), for example, in the sentence, "Since the wedding was today, the baker rushed the wedding *cake/ pies* to the reception." The word *cake* is more predictable than the word *pies* and will therefore be more likely to be skipped. In addition, high frequency words are skipped more often than low frequency words (Henderson & Ferreira, 1990; Inhoff & Rayner, 1986; Rayner & Fischer, 1996). For example, in the sentence, "Mary bought a *chest/trunk* despite the high price.", the high frequency word *chest* would be more likely to be skipped compared to the low frequency word *trunk*. Furthermore, monosyllabic words are skipped more than disyllabic words suggesting that phonological information can also be utilised early enough in the parafovea for it to impact on word skipping (Fitzsimmons & Drieghe, 2011). For example, in the sentence, "The workers were quick at loading the *grain/cargo* onto the ship." The monosyllable word *grain* is more likely to be skipped than the disyllable word *cargo*, because it is phonologically less complex.

When we move our eyes

The decision of when to move the eyes is mostly driven by lexical factors and ease of processing (Liversedge & Findlay, 2000). Fixation times can be influenced by a range of lexical variables (for a review, see Juhasz & Pollatsek, 2011). Some examples of the types of lexical variables that can influence when we move our eyes will now be discussed.

How common a word is in the language is referred to as word frequency and the frequency of a word can influence eye movement control, with high frequency words being fixated for less time than low frequency words (Inhoff & Rayner, 1986; Rayner & Fischer, 1996). For example, in “The cold *water/tonic* tasted stale” the high frequency word water will be fixated for a shorter time than the low frequency word tonic.

The letters a word contains and how similar that word is to another word can also influence when we move our eyes. This is referred to as orthographic similarity and is often measured be assessing orthographic neighbours. Orthographic neighbours are words that differ from the target word by a single letter, such as the word *house* has neighbours such as *mouse* and *horse*. In lexical decision tasks there is evidence that there is a facilitative effect of having a large number of orthographic neighbours (Andrews, 1997) where those words with more orthographic neighbours may be identified more quickly. However, when the words are inserted into sentences, inhibition effects were observed (Pollatsek, Perea, & Binder, 1999). Perea and Pollatsek (1998) found inhibition effects when comparing target words with high-frequency neighbours (target: *plate*/high frequency neighbour: *place*) with target words without high frequency neighbours (e.g. target: *spoon*). There is also evidence suggesting that changes to orthography affect eye movements. Rayner, White, Johnson and Liversedge (2006) found that when letters within words were transposed reading times increased, for example the word *problem* could become *rpoblem* or *prolbem* by transposing two letters. Reading times increased the most when the first few letters of words were transposed. This suggests that the beginning letters are crucial in word identification.

There are also semantic influences where the meaning of the word affects when we move our eyes. One semantic influence is predictability, where highly predictable words have shorter fixation durations than unpredictable words (Balota et al., 1985; Ehrlich & Rayner, 1981; Fitzsimmons & Drieghe, 2013; Rayner & Well, 1996). For example, in the sentence “The baker rushed the wedding *cake/pies* to the reception.” the more predictable word *cake* will be fixated for a shorter time than the less predictable word *pies*. Word predictability is measured by the use of a cloze task. Separate participants who do not take part in the actual experiment are given the beginning of a sentence up until just before the target word in the sentence and asked to fill in the next word. The percentage of participants that correctly match the target word is used as a measure of predictability. The more predictable the target word the higher the percentage of participants matching the target word in the cloze task. There is also evidence of plausibility influencing when we move our eyes. Rayner, Warren, Juhasz and Liversedge (2004) found that if a target word was anomalous (i.e. the target word does not make sense) given the prior context, gaze duration increased for that word when compared to a plausible or implausible word. For example, “Bill used the knife to cut the hard cheese” is plausible, “Bill used the scissors to cut the hard cheese” is implausible, but “Bill used the calculator to compute the hard cheese” is anomalous and would increase gaze durations.

Models of eye movement control

Over the years a number of models of eye movement control have been developed. The most influential and highly cited model is the E-Z Reader model (Reichle, Rayner, & Pollatsek, 2003). This model suggests that words are serially processed, meaning that only one word is being lexically processed at any one time. In contrast, other models such as SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) suggest that words can be processed in parallel. However, most models do agree that lexical processing has a strong influence on when we move our eyes. Current models of eye movement control do not take into account saliency or high level information concerning hyperlinks. This means that the models cannot make any in-depth predictions for how saliency and hyperlinks could affect reading behaviour. The only realistic prediction that could perhaps be applied to hyperlinks is from the E-Z Reader model (Reichle, Warren, & McConnell, 2009) where it is suggested that higher-level processes intervene in eye movement control only when "something is wrong" and either send a signal to stop moving forward or a signal to execute a regression. This is exclusively seen to impact the later eye movement measures (see previous discussion of early versus late eye movement measures in section: Analysis of eye movement data). There has been comparatively little focus on high-level processes within reading and eye movement research. Therefore, the current research is important in providing additional information for the modification of the current models of eye movement control to account for the effects we find. The serial versus parallel debate is of lesser importance to the work shown in this thesis, but it is important to be able to consider these models and there approaches to eye movement control. However, the serial vs parallel debate will not be discussed in depth here as it is outside the scope of this thesis.

Saliency of hyperlinks

One of the main differences between reading plain text and reading on the Web is the fact that online text may contain hyperlinks, which are usually salient in comparison to the surrounding text. Salient items are those that stand out from the rest of the items. The majority of studies on eye movements and reading do not manipulate the saliency of the text being read and instead, to maintain a high level of experimental control, often involve participants reading exclusively black text on a white background. It is therefore interesting for both theoretical and applied reasons to ask the question: does the coloured nature of hyperlinks and their saliency in itself influence reading behaviour?

Although there has been limited research exploring reading for comprehension when there are salient hyperlinked words present in the text, there is some relevant research exploring related issues. Leyland, Kirkby, Juhasz, Pollatsek and Liversedge (2013) found that shading words had an effect upon saccadic targeting (i.e., influencing where the eyes move to). If the first half of the word was shaded, fixations landed closer to the beginning of the word. Furthermore, partially shaded words were fixated for longer than fully shaded words, or non-shaded words, suggesting that visual non-uniformity (where the word is not fully shaded creating a non-uniform word) also influences when we move our eyes. These findings suggest that saliency can influence both when and where we move our eyes during reading, demonstrating the importance of considering saliency in relation to reading on the Web. However, the shading of words within the text does not appear very frequently in normal reading or reading on the Web. Elsewhere, White and Filik (2009) examined eye movements while participants read passages of text that contained normal, bold or italic target words, which are much more common in everyday reading. Italic text showed no differences in fixation times when compared to normal text, however bold text had significantly shorter fixation durations on the target word. It was suggested that the bold words attracted attention and the shorter fixation times reflect the improved visual discrimination of the text, perhaps making it easier to identify. Additionally, an attraction effect was also observed by Hyönä (1995b) where participants' eyes were drawn to irregular letter clusters that were unusual and therefore attracted the participants' attention. This attraction may be also likely to occur for salient text in a passage.

Recently, Gagl (2016) asked participants to read text that featured target words that were highlighted (either coloured blue or underlined) or the target words were degraded in quality to make them more difficult to read. Gagl found an effect of degradation with increased fixations times in first fixation durations, gaze durations and total viewing times. However, there was no effect of colour or underlining in first-pass reading. In total viewing times there was an effect of whether the target was underlined and an interaction between whether the target was coloured and whether the target was underlined. This interaction was due to the un-highlighted black words having reduced viewing times in contrast to all other conditions. This suggests that highlighting with colour or underlining the word increases re-reading times reflecting the allocation of attention to highlighted words after first pass reading. Gagl suggests that this means having hyperlinks coloured in blue is a good choice because it does not disrupt first pass reading, but attention is drawn to the highlighted words in re-reading so it serves the functions of highlighting important information without disrupting reading.

Other research has shown that the reader can use salient text to assist with the retention of these salient words. Nikolova (2004) examined participants who were studying a foreign language and found that visible hyperlinked words encourage participants to click on the highlighted words to seek their meaning. It was suggested that the saliency of the word ensured better acquisition. In this study the hyperlinked words did not distract the readers, but helped with their acquisition. Brett (1998) commented that saliency could be important in learning and suggested that if language items are made salient in any way, such as by highlighting them, they are more likely to be remembered than non-salient language items. These findings may be an example of the Von Restorff effect, where items that 'stand out' are more likely to be remembered (von Restorff, 1933). This idea is discussed further in the next section on signalling cues.

From the above research it could be expected that hyperlinks do not disrupt reading behaviour and can, in fact, help with remembering the words that were hyperlinked. Hyperlinks are salient from the rest of the text so that the reader can discriminate the hyperlinks from the rest of the content and easily observe that the hyperlink can be clicked on. Hyperlinks are typically shown in a blue font colour so readers may learn this convention and take advantage of it when reading on the Web. In the next section, this possibility is discussed in detail.

The culture of the blue link

The formatting of hypertext has its own convention used across the modern Web, with hyperlinks often being denoted in blue and changing colour (normally to purple) when a destination has been visited. At the time of writing, the default hyperlink style for all major browsers (Chrome, Firefox, Internet Explorer and Safari) is to render hyperlinks in blue without an underline. Displaying hyperlinks in blue and whether this is beneficial has been the source of much debate, which will now be discussed in detail.

It has been argued that using the colour blue for hyperlinks is a poor choice. Nielsen (1999) claimed that it is 'the mother of bad design conventions is the decision to make hypertext links blue'. In support of this claim, Nielson cites the fact that a sharp blue image is difficult to bring into focus due to the fact that only 2% of the cones on the retina are blue sensitive and blue therefore makes for a poor colour choice in terms of usability (Galitz, 1997).

Elsewhere it has been argued that using the colour blue for hyperlinks does not negatively affect usability. Pearson and van Schaik (2003) suggest that the convention of the blue hyperlink should continue after they found, in a visual search task, that participants responded more rapidly to blue hyperlinks than red hyperlinks. Ling and van Schaik (2004) examined participants' performance in a visual search task in which the participants had to look for hyperlinks on a page and also asked about the participants' preferences of the format of hyperlinks and found that although the participants preferred bold and underlined hyperlinks, the format of the hyperlink had no effect on accuracy in the visual search task. Further evidence suggesting that blue hyperlinks are of benefit to the user comes from van Schaik and Ling (2003) who observed that the colour of a hyperlink also has an effect on accuracy in information retrieval tasks. Participants had greater accuracy in the tasks and rated the display more positively when the hyperlinks were presented in blue.

Although as stated by Nielsen (1999), hyperlinks being shown in blue should be a poor choice in term of usability, the research presented above converges upon the idea that neither the saliency of hyperlinks nor their blue colour has a negative influence upon information processing. If anything, it seems that hyperlinks can improve recall performance. Why might this be the case? Below we explore two potential accounts for why this might occur: namely, automatic attention and schema theory. It is important to note that these accounts are not mutually exclusive and both may contribute to the fact that the saliency of hyperlinks does not disrupt reading.

Automatic attention

Nielsen (1999) argued that a move away from blue hyperlinks could improve usability, but he also recommends that the convention of the blue hyperlink should remain because most users know that blue text denotes a hyperlink, making it easier for users to recognise which words are hyperlinks more rapidly. This idea is supported by research on automatic attention which suggests that when a user consistently searches the same environment for the same information which is always represented in the same way, then the processing becomes automatic (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). In the case of hyperlinks, blue text is automatically processed as being a hyperlink because blue text in a webpage context very often represents a hyperlink. This effect was observed by Campbell and Maglio (1999) where participants were faster to detect the target when searching for a target word in a webpage that was a blue underlined hyperlink compared to when the target words were underlined black text, even though the black text should be easier to read in terms of usability.

Schema theory

An alternative account for why hyperlinks being shown in blue may not negatively influence reading behaviour comes from schema theory (Bartlett, 1932). A schema is a hypothetical mental structure used to represent generic concepts, such as a framework or a plan. We can use a schema based on past experiences to make assumptions about what will happen in similar, future experiences. For example, Smith and Swinney (1992) presented participants with vague passages to read. The passages were either presented with a title, which helped prompt a schema of the task the passage was about or the passages were presented on their own with no schema prompt. Participants read significantly faster when they had a schema prompted as it assisted with the understanding of the passages of text. For example, an extract from one of the passages used by Smith and Swinney (1992) reads "Only one substance is necessary for this process. However, the substance must be quite abundant and of suitable consistency. The substance is best used when it is fresh, as its lifespan can vary." On its own the text is vague and confusing, hindering comprehension. However, if the schema for 'building a snowman' is triggered, the past experience and what it involved assist with the comprehension of the passage and it can be processed more easily.

Schema theory can be used to explain why even though blue is considered a poor choice for hyperlinks we do not see evidence of hyperlinks impairing information processing or recall performance. If the webpage conforms to what the user expects and fits their mental model or schema of how a webpage functions, they may find it easier to use than a webpage that does not conform to the user's schema. For example, Benway and Lane (1998) asked participants to find information on a webpage that was either contained in the text or in a large banner. They found that users often missed the information in the 'obvious hyperlinks', such as large banners, because they did not conform to their schema of the blue hyperlink (i.e. they location of the hyperlink was not in the body of the text and the hyperlink was not displayed as blue text) and they instead looked for hyperlinks presented in the way that fitted their schema (i.e. hyperlinks displayed as blue text) even though those hyperlinks did not stand out as much as the more obvious large banner hyperlinks. These findings suggest that we should not necessarily assume that the 'best' way to present information is to make it 'obvious' or salient from the background. It seems that saliency can be overridden by schemas. However, it is also worth noting that although saliency was a well-established phenomenon for a period of time (Itti, Koch, & Niebur, 1998; Itti & Koch, 2000), there is now substantial evidence that even highly salient objects can be ignored in favour of less salient objects. The individual observing the scene can override the saliency and instead focus on the task they are currently completing (for reviews see Henderson, Brockmole, Castelhano, & Mack, 2007; Tatler, Hayhoe, Land, & Ballard, 2011). In the case of hyperlinks this means it may be possible to override salient areas of a webpage, such as banners, and focus instead on searching for hyperlinks and reading text.

The research in this section suggests that hyperlinks are processed differently to normal text and suggests that reading hypertext on the Web is different from normal reading. It appears that although blue may not be the most appropriate colour in terms of usability (Nielsen, 1999), the fact is that blue is the known convention for hyperlinks and this has an influence on how people perceive hyperlinks and interact with them. There is an on-going debate about whether hyperlinks have a positive or negative affect on reading behaviour. This is examined in Chapter 2 where the saliency of a hyperlink is investigated further.

The fact that hyperlinks are salient is an important issue to investigate. However, the hyperlinks also represent something more. The blue text does not just mean that a particular word is important or that it is a hyperlink. It may convey that there is additional content at the other end of the link. This idea can be explored for a branch of reading research that focuses on the signals an author can put into text content and how the reader processes these signals. This is explored in the next section.

Signalling

Written text contains numerous different types of signals that the reader can make sense of and utilise to assist in their comprehension of the text, for example, including headings to break up the text content and signal the content contained within the following section or highlighting important words in the text by underlining them. Signals of this kind presumably evolved as a means for an author to guide the reader's processing of the text. The author can make the structure of the text more salient to the reader by including headings and subheadings (like in this thesis), or make an individual word or small section more salient to the reader by underlining a word or highlighting a word by displaying it in italics or boldface. The following section will explore signalling further and discuss the use of signals in hypertext on the Web.

Text information in expository texts tends to be hierarchically organised around a set of topics that might be related (Meyer, 1984). For example, a text could be organised by interlinking themes, with headings and subheadings such as in this thesis. Adult readers are typically experienced in reading text and are aware of the presence of topics within long texts. For example readers read sentences that introduce a topic at a slower pace than those that are not introductory sentences (Lorch, 1989). Also, topic sentences that are difficult to relate to the immediately preceding text are also read more slowly compared to sentences that are easy to relate to the preceding content (Hyönä, 1995a; Lorch, Lorch, & Matthews, 1985). This is an important point to consider for the readers of hypertext because once they click on a hyperlink they may end up on a webpage that does not easily link to the webpage they came from in terms of topic. Often the linking might be to a page written by a different author who is unaware of the source of the link to their content.

Lorch (1989) wrote a review paper on text-signalling devices and their effects on reading and memory. He defines signals as "writing devices that emphasise aspects of a text's content or structure without adding to the content of the text." The writer of a text has the role of organising thoughts and information in their head into written text. The writers have numerous signalling devices that they can utilise in order to highlight important pieces of text (e.g. via boldface, underline, etc.) or simply to assist with the organisation of the text (numbering, heading, etc.). All devices share the goal of directing the reader's attention during reading. Bold or underlining may highlight an important term in the text, whereas headings label the dominant theme for the following section of text. These both help to guide the reader to where the information is and help the reader to infer the importance of the text sections and their content. Signals are a diverse set of writing devices (Lemarié, Eyrolle, & Cellier, 2006) and they have a variety of means to emphasise and help organise the text.

Typographical cues such as boldface or underline may be used to visually emphasise a word or section of text, but there are also discursive phrases that can be used to help alert the reader. Including the phrases “Let me stress…” or “In conclusion…” signals to the reader that the following text will either be important or contain a summarised conclusion of previous points, respectively (Lemarié, Lorch, Eyrolle, & Virbel, 2008). These different types of signal can be broken up into two types: organisational cues and typographical cues.

Organisational cues

Organisational cues can assist with the organisation of the text content, for example, titles, headings, topic introductory sentences, indentation etc. Organisational signals can communicate the topic structure to the reader which can assist in helping the reader to process the text content and even facilitate memory for the text (Brooks, Dansereau, Spurlin, & Holley, 1983; Dee-Lucas & Di Vesta, 1980; Doctorow, Wittrock, & Marks, 1978; Holley et al., 1981; Spyridakis & Standal, 1987; Wilhite, 1986).

The reason behind this observed benefit for the presence of titles and headings has been explored. Titles and headings are distinguished from the rest of the text by both their spatial location (they appear at the top of a section of text, normally some distance away from the main body of text) and by their visual appearance (titles are often larger or in a different typeface). Titles signal the content of the text and help increase recall of the text (Bransford & Johnson, 1972). Lorch (1989) suggests this phenomenon is not due to the title cueing the reader to remember more, but instead by indicating a knowledge structure for organising the text into, and this is what assists with the recall, much like schemas (described in the previous section: Schema theory). Sanchez, Lorch and Lorch (2001) suggest that the benefit from the presence of headings is because they can assist the reader in developing an appropriate strategy for reading the text. They explored how headings influence the readers’ memory of the text. Expository text was read with or without headings and participants either received training on how to construct a mental outline to recall information or they received no training. They found that those who received training or those who had headings performed better at the recall task then those who had no training or those without headings in the text. This suggests that headings are a signal that can help induce change in the readers' strategies for encoding and recalling the text.

It is not just titles and headings that can assist with the organisation of the text. Topic introductory sentences seem to be important to the reader when reading the text and authors are aware of this and often use signalling devices to help highlight the topic structure of the text (Lorch, 1989). Topic sentences may require extra attention from readers because topic sentences tend to introduce a reasonably large amount of new information (Haberlandt & Graesser, 1989). However, time spent on topic sentences can be manipulated by other factors. For example, if topic boundaries are signalled (e.g. with headings or subsection headings) then more processing time is spent on topic sentences (Lorch & Lorch, 1996b).

However, organisational signals are not always helpful. The familiarity of the topic also has an impact. Organisational signals can benefit recall of unfamiliar text topics, but not recall of familiar topics (Lorch & Lorch, 1996a). They can also benefit the recall of relatively unelaborated topics more than elaborated topics (Lorch, Lorch, & Inman, 1993), and they benefit the recall of badly organised texts more than well-organised texts (Lorch & Lorch, 1985). Finally, they benefit the recall of texts that are organised in a complex manner more than those organised simply (Lorch et al., 1985). Basically, if a text is difficult for the reader to understand, encode or recall the text's organisation, then the reader will receive some benefit from any organisational signalling that makes it easier for them to comprehend the text.

When organisational signals are omitted from a text, readers tend to recall the initial content of the text and perhaps some of the concluding content, but they forget entire sections of the text (Kintsch & Dijk, 1978). In contrast, when organisational signals are added to the same text the reader can remember more of the text topics and their organisation. This suggests that the reader is not just remembering the text, but also something more about the structure and the way the text topics fit together (Loman & Mayer, 1983; Lorch et al., 1993; Lorch & Lorch, 1996a, 1996b; Mayer, Dyck, & Cook, 1984). From these findings it has been suggested that organisational signals allow the readers to use a different text processing strategy than they would otherwise use. The signals are believed to cause more attention to the text topics and their organisation, which could result in the readers creating a mental model of the outline of the text. This can then guide the readers' memory retrieval at the time of recall.

The presence of signals coupled with the reader's goal can assist the reader in choosing a strategy which best suits the task at hand. For example novices in a domain depend more on signalling contained in a text whereas experts with a large amount of background knowledge on the topic are not influenced by the signalling (Dee-Lucas & Larkin, 1988). This effect of expertise could also apply to longer or more complex texts making them a challenge to comprehend easily. The text could have a complicated structure or a complicated subject matter, or the text could seem complicated to the reader because they lack familiarity with the text content (Dee-Lucas & Larkin, 1988). Authors recognise these difficulties and can insert aids to help with text organisation or to identify important aspects of the text.

In terms of reading on the Web, the presence of navigational hyperlinks (such as navigational menu) may help the user with understanding the organisation of the whole text, perhaps in a similar way to a contents page or index. Also, the embedded hyperlinks within the text may assist the reader and serve as titles of other pages or at the least serve as a topic heading for what other webpages can be accessed from the current webpage.

Typographical cues

One of the main differences between organisational cues such as titles/headings and typographical cues such as underlining or boldface, is that typographical cues highlight the importance of a small region of text embedded within the main text. Typographical cues are more specific than titles or headings and are unambiguous with respects to what information they are signalling. In contrast to typographical cues, headings tell the reader about the general gist of the content that is below the heading and the reader is also left to interpret the heading on their own and decide whether that section of text may contain something relevant to their task (Lorch, 1989).

Typographical cues have also been shown to improve memory for the signalled content (Cashen & Leicht, 1970; Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch, Lorch, & Klusewitz, 1995). However, simply putting text in bold or underlining the text does not automatically mean it will be remembered. The reader has to assume what the typographical cues mean in order to utilise them. Golding and Fowler (1992) found that when the reader expected "questions on specific details", underlining sections of text facilitated cued recall for those sections. However, when the reader was expected to provide an outline of text or a list of solutions to the problem discussed in the text, the readers did not experience any benefits from the signalling. The reader takes into account the purpose for why they are reading when deciding whether to utilise the signalling cues present in the text. If the reader cannot gain anything related to their current task from the cues, it is a better strategy to not dwell on the signalled information if it is not relevant.

If the reader only makes use of signals that are actually useful to the task at hand, the frequency of the signalling is also a factor. If most of the text has some form of signal to cue the importance of the information then the signal might not be as effective or as informative compared to when only the most important text is signalled. This "over-signalling" can reduce the effectiveness of typographical cues. Lorch, Lorch and Klusewitz (1995) asked individuals to read a four-page text where they were tested on memory for specific target sentences. The text either contained no underlining (control), underlining of the target sentences (light signalling) or underlining of the text sentences and half of the non-target sentences (heavy signalling). Performance was improved for the cued recall task when the text had light signalling, but performance was not improved with heavy signalling. This suggests that if the reader can use the signalling to facilitate with the task they are engaged in, but the signalling is not useful for the task, for instance when the signalling is seemingly meaningless, the reader can ignore it. Lorch, Lorch and Klusewitz (1995) then replicated the control and light signalling conditions, but using capitalisation as the signalling instead of underlining. They found that reading was slower for the light signalling condition, but memory recall was improved. Upon further examination they found that the readers slowed down on the signalled content alone and sped up again when reading non-signalled content. This suggests that the reader may have thought the signalled content was important so spent longer on it. These findings are similar to those observed by Lorch and Chen (1986) where readers slowed down their reading speed when encountering text preceded by a number. This numbered text was also shown to be remembered more accurately in recall tasks suggesting that the numbering was signalling the importance of those sections of text.

During reading the reader needs to discriminate important and unimportant information and signals in the text can be used to assist the reader. The writer can specifically mark the important information and/or organise the text into a format that helps simplify decisions about how relevant each section of text is. By not signalling important information the reader may miss out on important points and this may decrease their comprehension of the text. However, it is up to the creator of the text to be responsible for what they signal as important because there is evidence that participants will rely on signals to infer importance, even if it is not appropriate. Lorch, Lemarié and Grant (2011) asked participant to select the sentences in a text that they considered to express the most important text topics. Asterisks were present either between major structural boundaries or between minor structural boundaries. There was also a control condition where there was no demarcation. Participants were either expected to use the asterisks as a signal for where the important information might lie in the text or the participant may ignore the demarcation and simply rely on the semantic content of the text and make their own judgements of importance. Lorch, Lemarié and Grant (2011) found that participants relied more on the demarcation than the semantic content of the text. In the control condition (with no demarcation signals), participants selected sentences from paragraphs that introduced major text topics. However, when there were asterisks before major text topic boundaries participants selected the sentences with asterisks significantly more often than in the control group. A similar effect was observed for the participants who had asterisks before minor text topic boundaries. These participants significantly favoured the minor text topic sentences that had the asterisk demarcation. The asterisks seemed to override the appropriate interpretation of the text. This suggests that signals can override semantic cues from the text and that participants are willing to believe the signalling rather than make their own objective judgements.

Lorch, Lemarié and Grant (2011) suggest that signalling cues can have limits to their usefulness. Another problem can occur such that when the scheme for classifying the typographical cues is very complex, it can prove a hindrance to the reader. For example, if four or five categories of text content are signalled by different combinations of cues (colour variation, typeface variation, and underlining). This suggests that typographical cues may be beneficial only if the nature of the cueing is simple enough for the reader to readily understand the relationship between the signal cues and the text content.

Finally, typographical cues can be expected to influence text search processes because the typographical cues are salient and it is easier to visually distinguish the signalled text from the rest of the text content. Frase and Schwartz (1979) demonstrate that when readers search for information within a technical text, their search is speeded up by the use of spatial cues (segmentation and indentation).

In terms of reading on the Web, hyperlinks may serve as typographical signals as well as organisational signals. In the previous section on organisational cues it was discussed that hyperlinks could serve as a title or as a topic heading for other webpages that might help the reader build up the structure of the text they are reading. However, hyperlinks could also be said to be a typographical signal due to the fact that hyperlinks are a single word or short phrase that is salient from the rest of the text. Hyperlinked words could be considered as important pieces of information compared to the rest of the non-linked text and the presence of the hyperlink may add emphasis to that section of text.

SARA – A model for signalling

A recent model of signalling was published by Lorch, Lemarié and Grant (2011). They suggest a theory for signalling and its effect on text processing. The theory is called SARA which stands for "Signalling, Available, Relevant, Accessible". Basically, the *signalling* devices used must make *available* information that is both *relevant* to the reader's goals and *accessible* to cognitive processing by the reader. If any of these points are invalid then the signalling may not be successful at assisting the reader with their processing of the text.

Lorch, Lemarié and Grant (2011) argue that signals are defined as "metastatements" that the author anchors to sections of the text. The signals work in a similar method to punctuation. We all know that when we see a question mark that the text preceding the question mark is in the form of a question. Similar metastatements exist for the various different types of signalling. For example, a heading lets us know that the following piece of text is about the topic in the heading and boldface used on a word in the main text lets us know that the author wants that word/phrase to stand out from the rest of the text.

The main claim from SARA theory is that the signals communicate at least one of the following seven types of information:

1. Signals may demarcate underlying structural boundaries in the text.
2. Signals may communicate the hierarchical organisation of sections of text.
3. Signals may communicate the sequential organisation of sections of text.
4. Signals may label a part of the text.
5. Signals may identify the topic of a part of the text.
6. Signals may identify the function of part of the text.
7. Signals may emphasise parts of the text.

Some signals may only have one function from the list above, such as boldface text emphasising the bolded words in the text. Other signalling devices may have multiple functions, such as a heading. A heading serves as a structural boundary and helps label the text following it and identifies a topic for that following text.

Lorch, Lemarié and Grant (2011) are careful to heed the warning that research on different types of signalling device should not be grouped together by the type of signalling device. It should instead be approached from what function a signalling device serves. Therefore, instead of asking questions such as “how do headings affect text processing?” we should be instead asking “what function will the heading serve and how will the reader utilise it?” Not all headings have the same function and the function served can be dependent on the task of the reader. These are important factors to consider when exploring how people read, especially on the Web.

As a final comment Lorch, Lemarié and Grant (2011) mention the impact of maintaining topic and integrating information when in a hyper text environment. Hypertext allows the reader the ability to leave a webpage and go to another webpage or even another website. This means the reader may have more challenges integrating information together when compared to a linear document with only one author. Websites can have a number of different authors, all with their own approaches to signalling content. Lorch, Lemarié and Grant (2011, pp.140) argue that a “signalling device must make available information that is both relevant to the reader's goals and accessible to cognitive processing by the reader." However, if the signals vary and change across different webpages this might prove a greater challenge to the reader than just searching through information from a single source. Most sentences within a subsection can be understood from the context immediately preceding it (Kintsch, 1998). However, it may be more difficult to understand the first sentence of a new subsection from the final sentence of the preceding topic (Hyönä & Lorch, 2004). This problem could also be an issue for reading hypertext where the preceding topic and preceding sentence are no longer available on the screen after following a hyperlink. It is far from certain that the previous sentence on a different webpage will fit into the context of the webpage where the reader ends up after following a hyperlink.

Task effects and reading strategies

The readers' goal has an influence on inference generation and memory for expository text (van den Broek, Lorch, Linderholm, & Gustafson, 2001). Participants were asked to read either for study or for entertainment. They had to engage in a think-aloud procedure during reading and they also had to complete offline memory tests afterwards. Those asked to read for study had more coherence-building inferences where they sought explanations and often paraphrased sections of text to assist in creating a deeper understanding of the text. Those participants that were given the task to read for entertainment showed a different strategy where they made more evaluations and associations focused on connecting the text to their own personal experiences. van Den Broek, Lorch, Linderholm and Gustafson (2001) suggest that their pattern of results is due to the readers employing *standards of coherence*, which they use to guide their inferential activity and memory of the text. This concept suggests that as readers proceed through the text they try to maintain a standard of coherence as a criterion for comprehension. Dependent on the task at hand the reader may maintain a level of comprehension that might not be suitable for another task. For example, those reading for entertainment might not require the same level of comprehension compared to someone required to sit an exam on the topic they are reading. The readers' goal of either reading for study or for entertainment influences the standards of coherence. Those participants with the study goal employed stricter standards and focused more on the intratextual relations and on understanding and coherence. They also spent longer on the texts, reading more slowly and wanting to understand the connections between topics. Conversely, those participants who read for entertainment were less concerned with constructing a coherent representation of the text, but rather focused on how the text linked to the reader's personal experiences and making more evaluative comments, serving as a commentator on the text. This concept of standards of coherence is quite similar to that proposed by foraging theory (see section: Skim reading and foraging for information). Foraging theory assumes that an individual is sensitive to their information gain and thus can change their information search strategies based on how much information gain they are happy to receive.

There also seems to be an influence of familiarity on reading strategy. Kim (2001) asked novice and expert Web users to find the answers to questions in webpages. The novices visited a greater number of pages and spent longer trying to find the information that was required for the task. Kim suggests that the novices lacked an ability to differentiate pertinent elements from others. The novices could not locate the important information or use cues to the same extent as the experts for suggesting where the information might be. The novices also tended to use the "Home" button more often than the expert users, suggesting that they were disorientated more often and would go back to the "Home" screen to re-orientate themselves in the Web environment. When exploring how many layers the novices and experts consecutively travelled they found that the novices tended to travel more than two layers whereas the experts tended to only travel one or two layers. It appears that experts were aware that the risk of disorientation increased the further they travel from a start point and therefore chose not to travel as far. Previous research has shown that novices depended more on signalling contained in the text then experts (Dee-Lucas & Larkin, 1988), but in the Kim (2001) study it seems that the novices are not able to utilise the signalling cues available to them. The novices' inability to utilise the signalling cues could be because the novices do not know what the cues mean. Conversely, the novices may have found the scheme for what the cues represent too complex, thus proving a hindrance to the reader (Lorch et al., 2011).

There is also evidence of different reading tasks and goals having an impact on eye movement behaviour. Kaakinen and Hyönä (2010) explored task effects on eye movements during reading by comparing proof-reading and reading for comprehension. They found that the readers' goal was important for deciding how to read and comprehend the text. The proof-reading task increased fixation times, shortened saccades, and landing positions were positioned more to the left of the word, in other words the task influenced both when and where the reader moved their eyes.

Signalling and eye tracking

There has been surprisingly little research exploring signalling with the methodology of eye movements. However, the research that has been conducted has clearly demonstrated the importance of signalling cues on the readers' behaviour.

Previously, eye-tracking has been used to study the influence of signals on text processing. It was found that readers use headings to identify new topics and to guide the processing of the text structure (Hyönä & Lorch, 2004). This is observed by a pattern of regressions from the end of text sections back to the sentence-initiating headings. In the experiment topic shifts were either signalled by topic headings or the headings were absent. The presence of headings resulted in a greater number of topics being recalled in text summaries created after reading. Also, a facilitation of text processing was observed in the eye movement measures for the text that contained headings. Headings speeded up the initial processing of the topic sentences, particularly the very first sentence introducing the new topic. Hyönä and Lorch (2004) suggest that this is evidence that topic headings are powerful signals that facilitate online comprehension processes and improve memory for the topic.

There also seems to be evidence for individual differences in reading strategies. Hyönä, Lorch and Kaakinen (2002) asked participants to read text and then summarise what they had read. Eye movements were tracked during reading and were used to identify the reading strategies and individual differences between the participants. Hyönä, Lorch and Kaakinen (2002) discovered four unique strategies being used by the readers in the way they processed the text. Fast readers tended not to look back at previous content. Slow readers made more forward fixations and engaged in more re-reading of each sentence before moving forwards to the next. Nonselective reviewers frequently looked back at previous sentences. Topic structure processors spent longer on headings and performed the best at creating accurate summaries of the text. This finding suggests that there are individual differences in the reading strategies that the readers use. It also suggests that those using signals (the topic structure processors) performed the best at the task, demonstrating that signals can be useful to the reader, but only if they use them.

There is a great deal of evidence suggesting that using signalling can assist the reader with comprehending the text. However, we need to also explore at a global level how the reader actually comprehends passages of text. In the next section I will explore cognitive models of reading theories and how the reader builds up a representation of the text as a whole in order to understand what they have read.

Cognitive models of reading theories

As people read they build a hierarchically structured mental representation of the information. This build-up of information is often described as involving processing at different levels (Kintsch, 1998). One framework for comprehension is the Construction-Intergration Model (Kintsch, 1988, 1998). First, at a linguistic level, when processing the particular words in the text the reader needs to decode the symbols of the text. So the letters "g-e-e-s-e" are identified as the word "geese". Perceptual processing is involved and also word recognition and parsing. Next, the semantics of what the words mean must be used to form the ideas within the text. So the reader forms an idea of "geese" and what they are. As the readers read they link the ideas (propositions) expressed by the text into a developing hierarchical representation. In the sentence "Geese crossed the horizon as wind shuffled the clouds," there are two propositions "geese crossed the horizon" and "wind shuffled the clouds". The microstructure is constructed by forming the relationship between propositions (Traxler, 2011). However, whole sections of text can be related to each other, with the microstructure itself organised at an even higher level. For example, for the following propositions "Mary went to the airport. She took a taxi. She checked in for the flight to Paris. She went to the gate. She got on the plane. The plane took off. The plane landed. She got off the plane.", the macroproposition would be "Mary flew to Paris". This global structure of the text is called the *macrostructure*. The macrostructure consists of global topics and their interrelationships (Van Dijk, 1980). Together, the microstructure and macrostructure are called the *textbase*. Finally, the textbase represents the meaning of the text, but if the text is only read at a surface level and not engaged with then the reader will only have a shallow representation of the text and not a deeper understanding. For a deeper understanding the text content needs to be constructed into a situation model. A situation model is a mental model of the situation described by the text. This requires the integration of the text content with prior knowledge that the reader holds. With this situational model the reader can make inferences from the text content, for example, if the text reads "Bill is older than Bob, and Bob is older than Frank.", we can infer that Bill is older than Frank, even though the text does not specifically say that. With our past knowledge of logic and experience we can infer other information from the text that is not explicit.

Another model of comprehension comes from Gernsbacher, Varner and Faust (1990). They propose that comprehension consists of three processes: laying a foundation for the text; mapping information onto that foundation; and shifting to new structures when the new information does not fit onto existing structures. Although this has a similar approach to the Construction-Integration Model, they differ on what contributes to comprehension skill. The Construction-Integration Model suggests that prior knowledge and the building of a coherent model results in good comprehension of text. However, the suggestion from Gernsbacher et al (1990) states that comprehension skill is guided by the ability to suppress irrelevant information. To support this theory Gernsbacher et al (1990) presented skilled and unskilled readers with a sentence which was followed by a target word. The participants had to decide if the target word matched the meaning of the sentence they had just read or not. The last word in the sentence was also manipulated to either be ambiguous or unambiguous. For example, "He dug with the *spade*." vs "He dug with the *shovel*.", when the target word presented was "ACE". Due to the fact that "spade" has an ambiguous meaning, participants were expected to be slower to respond to the target word "ACE" compared to when the last word was the unambiguous word "shovel". Gernsbacher et al (1990) found exactly this suggesting that the readers had a higher activation for the ambiguous words' inappropriate meaning.

The structure of the text can also affect the reader's comprehension. The comprehension of text content is dependent on the reader developing an understanding for the text they read. As the concepts and ideas within the text build up into a linked structure through text reuse, elaborating on previous sections of text and developing new ideas, these arguments then link together to form a cohesive representation of the text. If the text is structured to be cohesive then the text will be easier for the reader to create a well-structured and meaningful mental representation of the information in the text (Eylon & Reif, 1984).

It is more difficult to construct a good representation of the information in the text if the text is disjoined or disorganised. This could be a potential problem for hypertext documents where the reader can choose to navigate away from the text and could cause confusion if the information does not link together well across webpages (Charney, 1994). When a hypertext reader loses track of where they are (and where they have been) in the network of information, they may end up processing a large amount of material that is irrelevant to their task at hand (Whiteside, Jones, Levy, & Wixon, 1985). This is in line with signalling research where readers show improved memory and comprehension of text that contains organisational signalling because the organisational signals help the reader develop a knowledge structure for organising the text into (Lorch, 1989).

When we are exposed to a large amount of text to read we cannot always read it all in a timely fashion and comprehend it all to the same degree. This is especially important to consider when reading on the Web where there is such a large amount of content. Therefore, a strategy for reading quickly, but without sacrificing comprehension of the text is necessary. This is explored in the next section which looks at skim reading of passages of text.

Skim reading and foraging for information

In typical reading and eye movement experiments (for reviews, see Rayner, 1998, 2009), researchers want to ensure that the participants are reading and processing the sentences fully. This is implemented by asking the participants to answer comprehension questions after reading (usually a portion of) the sentences and encouraging them to read carefully without ‘skimming’. However, the reality is often quite different, as readers may skim-read text, for example, when they do not need full comprehension of that text or are only searching for a specific piece of information. The current literature suggests that reading on the Web is more likely to involve skim reading (Liu, 2005; Morkes & Nielsen, 1997). Therefore, in this section, skim reading is explored and explanations for skim reading in terms of theories of foraging behaviour.

One of the first experiments exploring skim reading behaviour used eye movement methodology to investigate the differences in how people read when they are reading normally or reading quickly. This research was first noted in a lab report (Just, Carpenter, & Masson, 1982) and is reported in Just and Carpenter (1987). Just and Carpenter (1987) studied three different types of reading: normal reading; skim reading; and speed reading (using participants who had graduated from a speed reading course). Just and Carpenter (1987) suggested that readers increase their speed by sacrificing the amount they understand from the text, thereby exhibiting a trade-off of greater speed at the cost of reduced comprehension. They found that speed readers were three times faster than normal readers and the skimmers were two and a half times faster than the normal readers in reading through the text presented to them. The eye movement analyses showed that the skimmers and speed readers fixated fewer words than the normal readers and the normal readers had longer fixations when they fixated a word. Speed readers and skimmers were also more likely to skip over multiple words compared to the normal readers.

In terms of which words were fixated, Just and Carpenter (1987) found that the normal readers fixated twice as many content words when compared to function words during normal reading. Reasoning that this may have been due to differences in word length between content and function words, they explored their data, but found that readers were more likely to fixate three letter content words than three letter function words, which is consistent with the standard pattern seen in word skipping (Brysbaert et al., 2005). However, a slightly different result was observed for the speed readers and skimmers. They were also more likely to fixate long words compared with short words, but they did not discriminate between short content words and short function words, both were skipped as often as each other suggesting that word length is an even more important factor for speed readers and skimmers compared to normal readers when planning where to move the eyes. Also, because the speed readers and skimmers skipped more words, the readers are fixating words far into their peripheral vision and therefore cannot gain any useful information other than discriminating word boundaries (such as whether a word is a function or a content word) due to the reduced acuity in the periphery.

Just and Carpenter (1987) also examined gaze durations. The gaze durations were shorter for the speed readers and skimmers, who spent on average 100 ms (around one–third) less time on each fixation. However, even with this reduction in fixation times the speed readers and skimmers still showed effects of frequency (low frequency words had longer fixation times compared to high frequency words) and word length (longer words had longer fixation times compared to shorter words) similar to those seen in normal readers, but the sizes of the effects were much smaller. All three groups showed changes in their reading speeds dependent on the sub-section of text they were reading. These changes in speed were roughly parallel across the groups suggesting that they slowed down and speed up their reading rate on the same sub-sections of text. The speed readers and skimmers tended to make more fixations rather than make longer fixations when they spent longer on a section of text. Just and Carpenter (1987) suggested that the reading rate varied depending on the section of the text because of "local variables that are idiosyncratic to the text", meaning that some sections of text have denser levels of information, or contain more difficult information compared to other sections of the same text. They suggest that if the reader encountered a difficult to parse phrase they may need to sample more densely in order to understand the text.

Finally, Just and Carpenter (1987) found that in terms of comprehension the normal readers had better comprehension than the others two groups. When comparing the speed readers to the skimmers, the speed readers answered more questions correctly (but, it was mostly restricted to general questions rather than those concerning specific details), in spite of reading on average 100 words per minute faster than the skimmers. This is interesting because it would seem that the speed reading training has assisted the speed readers comprehension and reduced the speed-accuracy trade off compared to the skimmers.

Research into skim reading and speed reading has often shown the presence of a speed/comprehension trade off. Issues with reading could be due to landing errors when reading. This is especially important to consider when skim reading because very little re-reading tends to take place so the errors may not be corrected for. A common error occurs when readers make return sweeps when reading multiline text, they often have some error in the landing site and an additional fixation is required to correct this (Just & Carpenter, 1980). However, this correction tends not to be noticed by the reader and does not disrupt reading processes to a large degree. Word identification ability strongly determined reading speed (Kuperman & Van Dyke, 2011). This suggests that reading speed is linked to language processing abilities, rather than eye movement control issues.

We know from research exploring how people read in an RSVP task that readers can recognise words and understand the meaning of a sentence very rapidly. In an RSVP task a single word is presented at a time for around 100 ms. The time of presentation of each word is set and out of controlled of the participant. Masson (1986) presented text via RSVP at 100 ms per word. The last word in each sentence appeared in uppercase letters and remained on the screen until the reader made a response (either a lexical decision task or name the word verbally). For some sentences, the final word was highly predictable from the preceding sentence context (e.g., “He mailed the letter without a STAMP”) and in other sentences the final word was unrelated (e.g., “He mailed the letter without a DRILL”). Masson found that readers were able to identify the final word more quickly when the sentence context made it predictable. We also know from previous research that it helps if the sentence is meaningful. If a participant is shown words in a random meaningless order and then recall them they have difficulty. However, when the words form a meaningful sentence the participants perform well at the task (Potter, Kroll, & Harris, 1980). This study only used individual sentences, when the text to be read is longer the reader has greater difficulty. The reader has to not only identify words and understand a sentence, they also need to build a representation of the entire text. When the text was presented at a similar speed to the reading speed of plain text displayed as a paragraph, the readers’ comprehension and memory of the text was similar in both RSVP and plain text. However, when the RSVP task was displayed in a faster rate the comprehension and memory of the text decreased (Juola, Ward, & McNamara, 1982; Potter et al., 1980). Masson (1983) showed that pauses between sentences are necessary for comprehension success when reading in an RSVP format. Masson (1986) suggests that scaling up to passages reading from single sentence reading comes at a substantial cost.

Other researchers have also shown a reduction in comprehension when reading rate increased. In order to explore skim reading behaviour, Carver (1984) displayed passages of text to participants and gave them varying amounts of time to read the passages. When testing the participants with comprehension questions, those who had the shortest time to read the text performed the worst. This suggests that by increasing reading speed, comprehension is reduced. However, it is difficult to see if comprehension is reduced globally across the text. Comprehension may only be reduced on certain parts of the text and not on all parts, some parts might be considered important and the reader may focus on those parts of the text. Other researchers have tried to explore this by having independent participants rate sentences within passages of text and exploring if skim reading is used to skim over the unimportant pieces of text rather than just skim read all of the text. For example, Masson (1982) manipulated the time participants had to read passages of text and tested their recognition memory for the text in the passages. The recognition rates decreased when the participants' time to read the text decreased. Also, the faster the text was read, the longer the reaction times were to respond to the recognition questions. However, this was only true for those sentences that were rated as 'unimportant' (as judged by a different set of independent participants). The sentences judged as 'important' did not show the comparatively longer reaction times. Masson (1982) suggested that this was due to participants focusing more on relevant and important information in the passages to enable faster processing of the text, and paying less attention to the unimportant sentences in order to read more efficiently.

Conversely, Dyson and Haselgrove (2000) asked participants to read short passages at either normal reading speed or as a self-paced faster rate (about twice as fast). Participants completed a multiple-choice comprehension task and performed worse when they read faster. The information they did recall was more general in nature: specific details were not remembered as well. There was no interaction between reading speed and the type of information the participants remembered.

In order for skimming to be an effective method of gaining information quickly, without sacrificing comprehension the reader needs to focus on the important information and not waste time on the unimportant information. Skim reading can be thought of as efficient reading, an attempt to increase reading speed while trying to keep comprehension high. Effective skimmers scan a text for headings or key words to locate relevant information that could be important. The reader can then skim over unimportant regions of text and focus on reading the sections that might be important more carefully. Hyönä, Lorch and Kaakinen (2002) found that the readers who looked at headings wrote the most accurate text summaries. The reader does not have to read the same way for every task. Some reading goals involve searching for a single fact and do not require full comprehension of the text as a whole. In this scenario it is more efficient to skim read the text.

Reader and Payne (2007) researched skim reading and found evidence of the 'adaptive allocation of attention' in skim reading tasks. The participants were given four texts of varying difficulty on the topic of the human heart. Participants spent more time on the easier texts, despite the absence of any cues to the difficulty of the text. Reader and Payne (2007) suggest this is a *satisficing* strategy.

The concept of a satisficing strategy comes from information foraging research into browsing where it is assumed that the readers are sensitive to their 'information gain' (how much useful information they are getting over time) and use this as a basis for what to read and when to stop reading. A reader would be monitoring their information gain and have a threshold of how much information they are happy to get in a certain amount of time. If that information gain drops below that threshold the reader will then stop reading that particular piece of text and move on to a new patch where they might gain more information in the same amount of time. Pirolli and Card (1999) use a metaphor of a bird foraging for berries as an example of information foraging. A bird is foraging for berries in patches of bushes. The bird must decide how long to spend on one patch before expending time moving onto a new patch to forage for berries. The problem is at what point does the bird decide to move from the one patch to a new one? The most efficient time to leave for a new patch is when the expected future gains from foraging in the current patch decrease to such a level that it is better to expend time moving to a new patch where the future gains may be greater.

Reader and Payne (2007) suggest that this information foraging approach of satisficing can be applied to skim reading if we assume that the 'patches' are patches of text or paragraphs, and the reader has a threshold for their information gain influenced by the amount of time they have to read the text. If the reader has a short amount of time to read the text, they will want to have a lower threshold for information gain. If they are not receiving enough information from a patch they will want to realise this quickly and move on to a patch that has a higher information gain to make the most efficient use of the limited time. If this is true then the readers will focus on the most important information patches and leave the patches with less important information if their time is limited.

Duggan and Payne (2009) conducted several experiments to test if participants focused on the more important information in the text when skim reading. They found that readers asked to skim read had better memory performance for important details from the text, but not for the unimportant details. Where previous studies that have explored skim reading have shown a decline in comprehension performance, Duggan and Payne (2009) found an improvement in comprehension and found higher scores in comprehension questions for sentences rated independently as 'important'. This suggests that skim reading is an adaptive satisficing strategy. By leaving text before it is processed in depth and when information gain begins to drop, readers can efficiently move through the text at an increased speed, while trying to keep comprehension high. Skim reading is a trade-off whereby the reader is trading depth of comprehension for speed, but while trying to minimise the loss of comprehension by using an effective strategy to move through the text quickly without losing comprehension. They also found that readers spent more time reading text earlier in the paragraph and towards the top of the page. This supports the F-shaped pattern observed by Nielsen (2006) where readers focused more at the top of the page before skimming down to access the rest of the content on a page (see section: Task effects).

Rayner, Schotter, Masson, Potter and Treiman (2015) compiled a review of speed reading research and in summary comment that there is a speed/comprehension trade-off. If the reader wants a full understanding and comprehension of the text then speed reading is not a suitable strategy. However, they comment that having a high language skill (such as increased vocabulary) is the only way a reader can be fast and maintain comprehension.

Reading is not the only task that users on the Web engage in when reading a webpage. They need to decide what hyperlinks to follow and when to follow them. They also need to understand what previous pages they have been to and how all the pages they visit link together. All these extra decisions that have to be made in addition to reading the content make reading on the Web different to reading plain text. In the next section I will explore the research that has investigated how readers navigate through and read webpages and the issues that surround it.

Navigation: Link selection, comprehension and cognitive load

So far in this thesis, the issues of how the visual characteristics of hyperlinks may influence reading behaviour, how signals affect reading behaviour and how the goals of the reader may also influence reading behaviour on the Web have been discussed. Here, we consider how the interconnected nature of Web browsing and navigation may influence reading behaviour on the Web. When reading non-Web forms of plain text, such as novels or magazine articles, the text is written in such a way that it is best comprehended when read from the beginning to end in a linear fashion. webpages, on the other hand, are an interconnected network of text passages that do not necessarily need to be read in a specific order, nor do they tend to have a specified beginning and end point (Slatin, 1990).

There are a number of consequences of the fact that webpages are interconnected and linked to each other. First, when reading any individual webpage, an observer is likely to be navigating their way through a series of pages for a particular purpose (e.g., understanding a topic, purchasing goods online, information retrieval). As a result, their reading of a single page needs to be considered as part of a chain of visits to different webpages in accordance with their current set of goals. Secondly, we have already seen from the previous section on foraging that the observer needs to make decisions regarding when to leave a particular page and click a given link to move to a different page. This decision-making process is likely to require additional processing resources and so some have argued that it leads to an increase in cognitive load.

In the sections below, the issue of finding information when navigating through a webpage versus linear text is explored. Also, how having to make decisions about what hyperlinks to follow could increase cognitive load is discussed. From a theoretical perspective, examining these factors in Web reading is of value because in typical eye movement and reading studies, participants only read a single line of text that is not related in any way to the previous sentences that have been read (so no navigation is required), and furthermore, participants do not need to make a decision regarding where to click to move to next and navigate.

Navigating through a non-linear space

One of the main differences between reading plain text and hypertext is the fact that hypertext is non-linear and has no specific beginning and end point and no strict route through the information. Slatin (1990) argued that linear reading requires linear thinking, piecing the story together from beginning to end, while reading hypertext requires associative thinking which is more discontinuous, linking associated topics together, that do not necessarily follow each other. However, as Conklin (1987) suggests, with a large website containing many interlinked pages, the reader could easily become 'lost in hyperspace' when trying to navigate through a website. Therefore, this section will examine research that has explored how people navigate through non-linear spaces such as the Web.

As discussed previously (see Schema theory), individuals build mental schemas based on past experiences to make assumptions about what will happen in similar environments in the future (Bartlett, 1932). Previous research on cognitive maps explores a similar premise. Tolman (1948) suggests we can build cognitive maps in our minds which are a cognitive representation of the layout of an environment. McKnight, Dillon and Richardson (1990) explored time spent on index/contents sections (navigation and organisation sections that contain the list of content and page numbers/number of pages) versus time spent on content sections (sections containing the actual content of the text) of hypertext and linear documents. The participants had to complete an information retrieval task where they had to find the answers to a set of questions. They found that participants reading hypertext spent more time in the index/contents sections of the documents and noted that participants often jumped from content back to the index. Participants had poorer scores for information retrieval in the hypertext documents in comparison to the linear documents and seemed to jump between the index and exploring a page before choosing another page, to try and find the answers they needed. Participants preferred to return to the index instead of using the nodes/links within the text to browse between pages. In other words, the participants did not seem to develop a map of the structure of the information and simply went back to the index to try a different page if they did not find the content they were looking for. In comparison, the participants looking through linear text were confident to browse through the content to find the information they needed. McKnight et al. (1990) suggested that the unknown scope of the hypertext document could lead to incorrect assumptions about the scope of the documents content and result in a poor reading strategy. In a linear text it is much easier to see the scope of the document and browse through the content. Dillon, McKnight and Richardson (1993) argue that if the user does not know how the information is organised it makes it difficult to find specific information. Paper-based documents such as books tend to have a convention for how the information is organised such as an index page and a contents page to help find a topic and see the overall organisation of the whole text. This links back to research on signalling cues where index/contents pages can serve as organisational cues that can be used to assist the reader in processing the text (see Signalling).

McDonald and Stevenson (1996) asked participants to answer ten questions from a text that was either displayed in a linear form or as hypertext documents which were either hierarchically structured or in non-linear hypertext topologies. The main difference between the hierarchical and non-linear hypertext structure is that the hierarchically structured document has more of a framework that the information is organised into. Whereas the non-linear hypertext structure has no specific structure and the participants are free to move through the link structure as they please. Participants performed better at the task with the linear text than the non-linear, but the hierarchical document fell in the middle between the two. This suggests that the participants' performance was improved when there was some structure in place. Elm and Woods (1985) suggest that users may be overwhelmed and disorientated by the sheer amount of choice offered by hypertext. The users may not understand the structure of the system and how it all fits together and what potentially exists in the hypertext document. McDonald and Stevenson (1996) argue that although a large linear text can also be confusing for a reader, there are a number of discourse cues such as page numbers, contents listings and headings that the reader can take advantage of. The non-linear hypertext lacks a lot of these types of cues. The same text presented as a linear document versus as hypertext can lead to navigational problems. Due to the non-linear nature of hypertext, this can cause

problems when trying to understand the structure of the information and build a representation of the information. However, when the hypertext document is hierarchically structure this removed some of the confusion, but the readers still did not perform to the same level as those who only experienced the linear text.
When navigating on the Web decisions have to be made about which hyperlinks to follow. In comparison to reading plain linear text these decisions are an extra task that needs to be completed alongside the task of reading. An important question to consider is whether this extra task of deciding which hyperlinks to follow adds extra processing that could hinder the reader. The next section explores whether the decisions the reader has to make while reading hypertext has a negative effect on their reading behaviour.

Decision making and cognitive load

There is an on-going debate about whether in-text hyperlinks hinder reading. Carr (2010) suggested that hyperlinks within the text are a distraction and therefore hinder comprehension of the text. He argued that having to evaluate hyperlinks and navigating a path through them is demanding and is an extraneous task to the act of reading itself. Carr goes on to suggest that deciphering hypertext substantially increases readers' cognitive load and weakens their ability to comprehend and retain what they are reading (Carr 2010, pp.126).

Carr's argument is based on research investigating the cognitive load of hypertext on users (e.g. Miall & Dobson, 2001) which suggests that comprehension was increased when participants read plain text compared to when they read hyperlinked text. However, as noted by Miall and Dobson, their study is somewhat limited in being able to generalise to other forms of hypertext, including reading Web pages. The text used by Miall and Dobson (2001) was a piece of literary fiction that had been converted to hypertext and hyperlinks were added to it, making the hyperlinked document quite artificial. This artificial hypertext document may be the reason for the increase in cognitive load, which makes it difficult to generalise these results to 'real' hypertext documents. This being said, other research does corroborate with Carr's (2010) suggestion that the extra demands of having to make decisions about whether to follow hyperlinks or not could impair reading.

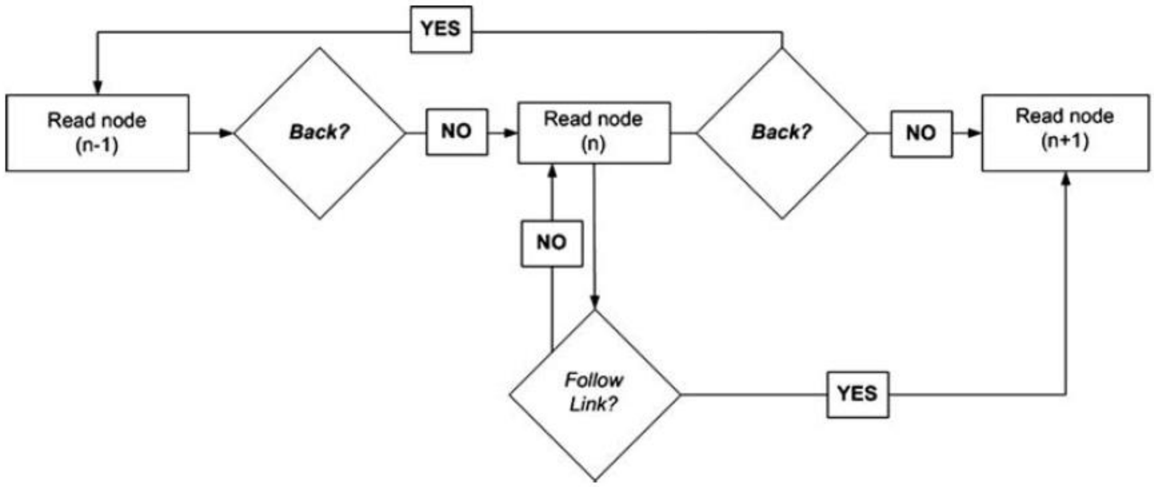


Figure 1.8 A process model for hypertext reading (DeStefano & LeFevre, 2007). When a node is read there are two main decisions to be made. 1) Should a hyperlink in the text be clicked and followed? 2) Should the reader go back to previous content?

There is also research that explores working memory and the concept of cognitive load and its impact on reading hypertext. DeStefano and LeFevre (2007) conducted a review of cognitive load in relation to reading hypertext. They argued that the extra task demands of reading hypertext causes an increased cognitive load to the readers in comparison to linear text. Because the readers have to make decisions about which hyperlinks to follow this puts demands on working memory. Working memory is used to encode, activate, store and manipulate information while a person performs cognitive tasks (Baddeley, 2003). Working memory models suggest that working memory is limited and only a finite amount of information can be stored at any one time (Baddeley & Logie, 1999). DeStefano and LeFevre (2007) remark that this aspect of the working memory models supports the assumption that increased cognitive load from reading hypertext may result in reduced performance for reading hypertext. In linear text there are no hyperlinks so the reader is not required to make any decisions about whether to follow a hyperlink or not. DeStefano and LeFevre (2007) argue that the decision to follow a hyperlink or not requires cognitive resources and so therefore increases cognitive load (See Figure 1.8 for a process model for hypertext reading from DeStefano & LeFevre, 2007). Recently, Scharinger, Kammerer and Gerjets (2015) measured both the EEG and pupil size of readers engaging in a task that closely simulated hypertext reading and link selection. They found evidence of increased load on executive functions when the reader had to perform hyperlink-like selection in the form of a significant increase in pupil size as well as a significant decrease of alpha frequency band power. This suggests that there is additional cognitive load when having to make hyperlink-like selections when reading hypertext which might hamper reading and comprehension when reading hypertext. Scharinger, Kammerer and Gerjets (2015) also ran further experiments to rule out confounding valuables and confirmed that this additional load was caused by the act of having to make the decision rather than the colour of the target word or the motor response (the act of making a mouse click) causing the increase cognitive load.

It is not just the decision of whether or not to click a hyperlink that could increase cognitive load. Whenever a reader chooses to follow a hyperlink and explore different content, this could interrupt the on-going comprehension processes. Comprehension involves the creation and development of situation models, which are complex mental representations that the reader forms to integrate statements in the text they are reading into their knowledge (Kintsch, 1988). This is also observed by Dee-Lucas and Larkin (1995) who found that hyperlinks in text distract users by interrupting information processing. While reading, users may stop to click on hyperlinks in the middle of text content, thus interrupting their cognitive processing and leaving the reader with a fragmented representation of the text content. Because of the nonlinear nature of hypertext, when a reader is reading text on one topic on a webpage, if they choose to click a hyperlink, it takes them to another webpage. This new webpage may contain content which is unrelated to the content they have just come from on the previous webpage. This could cause disruption to the readers developing a situation model.

Discussion: Synthesis and Direction for the Present Thesis

Reading on the Web is a complex task that is not well understood. However, previous reading and eye movement research can help research in this field. The current thesis takes the theories from eye movements and reading research and takes a stepwise approach to move from the controlled experiments on which these theories are based to a more ecologically valid approach towards exploring reading on the Web. The main issue is that there is very little empirical research exploring reading on the Web, so we start at the very basics of how reading on the Web might differ to reading plain text. This means that the experiments in this thesis are, and especially the ones at the beginning of the thesis, closer to traditional, very controlled, eye movements and reading research. However, with each experiment the move towards a more ecologically valid Web reading environment is made and an understanding of the different factors involved in reading on the Web is explored.

There are two major reasons why researching reading on the Web could have a novel contribution and impact. First, we can explore interesting theoretical questions by using the Web. We can explore not just the addition of relatively higher level sources of information compared to reading single sentences for comprehension (higher level information as in information not contained within the lexical information of a word, such as the fact that a word is hyperlinked and as such linked to other content), but also use the Web as an environment for asking novel questions. Second, there is an applied benefit in that it can help inform and develop a better Web, more suited to the characteristics of the human information processing system. A major criticism of the previous work referred to in the literature review, such as the more traditional eye movement and reading experiments, is that although there is an extensive amount of research into eye movements and reading, there is very little exploring the specific effects the Web has on reading. Much of the research exploring eye movements and reading is lacking in ecological validity because it tends to explore single line reading. However, this research offers the building blocks for understanding how people read and can serve as a stepping stone to understanding more complicated reading tasks. This thesis serves as such a stepping stone towards understanding how people read on the Web. There is a great deal of research exploring how we use the Web from a computer science or marketing perspective, but none of this research explores how we actually read hypertext on the Web. We have already seen that eye movement patterns are dependent on the task at hand (Yarbus, 1967), but there has been very little research exploring the various tasks that a user engages in when reading on the Web.

This thesis aims to understand the impact of reading hypertext on reading behaviour. The empirical work in this thesis will begin with understanding the basic impact of having a salient word within text (Chapter 2: Experiment 1A and 1B). This is to pull apart the impact of a hyperlink in the text from the impact of just having a salient word in the text. After exploring the role of saliency in this context we can then explore the impact of salient hyperlinks on reading on the Web (Chapter 2: Experiment 1C). This will begin by only focusing on participants reading static webpages where they will not yet be able to click on hyperlinks and navigate the Web. By removing the ability to click and focusing on the reading behaviour we can tease apart the differences between reading on the Web and traditional reading and explore the impact of these differences.

Chapter 3 focuses on the impact of having hyperlinks in the text by examining the influence it has on the importance the readers put on sentences in the text (Chapter 3). We explore how the properties of the text on a webpage affect the perceived importance of the content. This includes looking at the presence and number of links and exploring the impact on importance ratings.

The data from Chapter 3 then feeds into Chapter 4 where we consider task effects during reading. It is not always possible to read for comprehension, especially when there is a large amount of information to read through. With this in mind we explore the differences between reading for comprehension and skim reading on the Web and the impact of hyperlinks on these tasks (Chapter 4).

Finally, the previous experiments are set in a controlled environment that looks like the Web, but participants are not allowed to click and navigate. In the final experiment in this thesis we want to explore reading behaviour in a realistic setting where the participants were able to click and navigate. This is not an easy task to create and the methodology for this experiment is explained in Chapter 5. In Chapter 6, the experiment mimics an interactive Web environment where the users can read, click and navigate. From this experiment we can observe what additional impact being able to navigate has on reading behaviour and comprehension (Chapter 6).

By building up from basic experiments exploring the saliency of words during reading all the way towards an interactive, controlled Web environment we can observe the individual impact of the different factors involved in reading on Web.

Chapter 2

Do Hyperlinks have a Negative Influence on Reading When Reading for Comprehension?

Introduction

As noted in the literature review, the majority of reading research tends to have a narrow focus on how people read for comprehension. Moreover, theoretical and computational models of reading focus almost exclusively on accounting for data obtained from reading a single line rather than passage reading. Additionally, they also tend to focus primarily on lexical processing factors, such as word frequency, rather than high level factors such as task effects associated with different forms of reading (for example, proofreading or reading on the Web). Contrary to this traditionally narrow focus, in this chapter we examine the reading of paragraphs for comprehension within the context of reading on the Web.

One of the main differences between reading on and off the Web is that the materials that are being read on the Web contain hyperlinks embedded in the text. Two differences of hyperlinks compared to plain text that are explored in this chapter are that hyperlinks are coloured and salient compared to the rest of the text. This is from a cognitive psychology point of view a low-level influence. Another difference is that there is high-level information attached to a hyperlink. A hyperlink links one piece of information to another, perhaps on a separate page of the same website, or a different website all together. Thus, a hyperlink represents more than just a salient word in a passage of text.

This chapter consists of three experiments that attempt to pull apart the influence of saliency on the processing of hyperlinks and the influence of the hyperlinks linking to additional content.

Saliency

One of the main differences between reading plain text and reading online is that online text contains hyperlinks, which are salient in comparison to the surrounding text. Salient items are those that stand out from the rest of the items in some perspective. Visual salience is a stimulus-driven signal that announces to us that a certain item or location is different to the rest of the visual field and is worthy of attention. For example, a lone red item in a field of green items will stand out to us and be salient compared to the rest of the items and draw our attention (Treisman & Gelade, 1980).

Hyperlinks are often denoted in blue and in a different colour from the rest of the text, which is often black. Nielsen (1999) claimed that 'the mother of bad design conventions is the decision to make hypertext links blue' because only 2% of the cones on the retina are sensitive to blue making it a poor choice in terms of usability (Galitz, 1997). However, Nielsen admits that the convention of the blue hyperlink should remain because users know that blue text denotes a hyperlink, making it easier for users to recognise hyperlinks more rapidly. This is supported by research on automatic attention which suggests that when a user consistently searches the same environment for the same information which is consistently represented in the same way, the processing becomes automatic (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). This could also be true for hyperlinks because blue text in a webpage context almost always represents a hyperlink. This effect was observed by Campbell and Maglio (1999) where they found that participants were quicker when searching for a target word in a webpage that was blue and underlined than target words that were black and underlined, even though the black text should be easier to read in terms of visual discriminability.

However, there is very little research that explores the saliency of words when reading for comprehension. Simola, Kuisma, Oorni, Uusitalo and Hyönä (2011) explored reading in a Web environment and found that salient advertisements can distract attention and disrupt reading. If salient adverts can distract readers, it is conceivable that salient words may as well. White and Filik (2009) examined bold words in passages of text. They found that bold text had shorter fixation durations suggesting that saliency in text can affect information processing. They suggest their finding reflects the improved visual discrimination of the text, making it easier to identify. There is also evidence suggesting that saliency can affect not just when we move our eyes, but also where we move them. Leyland, Kirkby, Juhasz, Pollatsek and Liversedge (2013) in an eye tracking study during reading looked at fully or partially shaded words within the text and found that when they shaded a word it had an effect on saccadic targeting, influencing where the eyes move to. If only the first half of the word was shaded, the targeting was closer to the beginning of the word. Furthermore, partially shaded words were fixated for longer than fully shaded words, or non-shaded words, suggesting that visual non-uniformity (in the respect that it interferes with perceiving word boundaries) also affects when we move our eyes.

Recently, and building on the work presented in this chapter, Gagl (2016) asked participants to read text that featured target words that were highlighted either by being coloured in blue or by being underlined. Gagl found that highlighting a word by colouring it or underlining it had no negative or positive impact on reading during first passage. However, in total viewing times there was an effect of whether the target was highlighted. The un-highlighted black words showed a reduced viewing time in contrast to all other conditions. This suggests that highlighting with colour or underlining increased re-reading. Gagl suggests that this means having hyperlinks coloured in blue is a good choice because it does not disrupt first pass reading, but attention is drawn to the highlighted words in re-reading so it serves the functions of highlighting important information.

There has also been research into learning from electronic texts that suggest that hyperlinks do attract attention to them and that it does not disrupt reading, but actually assists in the retention of the hyperlinked word. It was suggested that the saliency of the hyperlinked words ensured better acquisition and retention (Nikolova, 2004). This is supported by the observation of a classic phenomenon called the Von Restorff effect (1933), where items that 'stand out' are more likely to be remembered.

High Level Processing

High-level processes within the current context are those processes based on information not contained within the lexical representation of a word. Hyperlinks denote a connection to other content somewhere else on the Web. Carr (2010) suggested that hyperlinks within the text are a distraction and therefore hinder comprehension of the text. Having to evaluate hyperlinks and navigating a path through them is demanding and is an extraneous task to the act of reading itself. This means that having a hyperlink in the text that links to other content renders the act of reading more laborious and we expect this higher-level process to be reflected in the eye movement measures.

In terms of a prediction for a high-level factor on reading, the current models of eye movements during reading do not make direct predictions for the impact of hyperlinks but the closest to a prediction that can be derived from one originates from the E-Z Reader model of eye movements during reading (Reichle et al., 2009) which suggests that higher-level processes intervene in eye movement control only when "something is wrong" and either send a signal to stop moving forward or to execute a regression. This is exclusively seen to impact the later eye movement measures so based on this model, so on the basis of that model we hypothesise seeing effects of reading hyperlinks exclusively in the later eye movement measures.

Signalling

Written text contains many different types of signals that the reader can use to assist in their comprehension of the text such as highlighting important words in the text. Signals are a diverse set of writing devices (Lemarié et al., 2006) reflected in a variety of means to emphasise and help organise the text. For this chapter we will focus on typographical cues such as boldface or underline that can be used to highlight single words or sections of text.

Typographical cues have been shown to improve memory for the signalled content (Cashen & Leicht, 1970; Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch et al., 1995). However, simply bolding or underlining the text does not automatically mean it will be remembered, the signal needs to be useful to the reader. Golding and Fowler (1992) found that when the reader expected "questions on specific details", underlining sections of text facilitated cued recall for those sections. The important information that the reader needed for the task was highlighted and this helped them find the information easily. However, when the reader was expected to provide an outline of text or a list of solutions to the problem discussed in the text, the readers did not experience any benefits from the signalling as it wasn't useful for the task at hand. Thus, signals need to be relevant to the reader to assist them in their task.

There is also an issue that even if some signals are useful, will the addition of (even) more signals be more useful for the reader? If most of the text has some form of signal to cue the importance of the information, then the signal might not be as effective or as informative compared to when only the most important text is signalled. This "over-signalling" can reduce the effectiveness of typographical cues. For example, Lorch, Lorch and Klusewitz (1995) asked individuals to read a four-page text after which they were tested on memory for specific target sentences. The text either contained no underlining (control), underlining of the target sentences (light signalling) or underlining of the target sentences and half of the non-target sentences (heavy signalling). Recall was improved when the text had light signalling, but performance was not different to the control condition when there was heavy signalling. If the signalling is not useful for the task, for instance when the signalling is seemingly meaningless, the reader will ignore it. Lorch, Lorch and Klusewitz (1995) went on to replicate the control and light signalling conditions, but using capitalisation as the signalling tool instead of underlining. They found that reading was slower for the light signalling condition, but memory recall was improved. Upon further examination they found that the readers slowed down on the signalled content alone and speeded up again when reading non-signalled content. This suggests that the reader may have thought the signalled content was important so decided to spend more time on it. During reading the reader needs to discriminate important and unimportant information and signals in the text can be used to assist the reader.

In terms of reading on the Web, hyperlinks could be said to be a typographical signal due to the fact that hyperlinks are a single word or short phrase that is salient from the rest of the text. Hyperlinked words could also be considered important and the presence of the hyperlink may add emphasis to that section of text.

Research Questions: How Does Colour/Hyperlinks Affect Reading Behaviour When Reading for Comprehension?

The experiments in this chapter focus on how we read linked text and whether links influence reading behaviour. In order to examine how links affect reading behaviour, we first need to find out if any disruption of reading occurs exclusively due to the link being a salient colour, before examining whether this is due to the link being perceived as important due to the additional information that it can link to.

Three experiments were conducted to explore this issue. The first experiment, Experiment 1A, explored whether a salient, coloured word negatively impacts reading behaviour outside of a hypertext context. Experiment 1A only used a single coloured word in a single-line sentence to explore the impact of saliency. A follow-on experiment (Experiment 1B), was conducted, exploring whether multiple coloured words had an impact on reading and we also include a word frequency manipulation to see if the difficulty of the word interacts with the fact that the word is coloured. This manipulation allowed us to examine whether there would be an additional cost of the colouring for words that are more difficult. The third experiment, Experiment 1C, explored whether perceiving the words as links impairs reading behaviour by presenting the coloured words in text that can be perceived as hypertext. We also implemented a word frequency manipulation in this experiment in order to explore whether common lexical effects are present in hyperlinked text and to investigate if they are modulated by the word being hyperlinked. Together, these experiments assessed whether there is a difference between reading coloured words (embedded in words of a different colour) and reading links and how this affects reading behaviour. Experiment 1A and 1B will help us to separate whether any observed effects seen in Experiment 1C are exclusively due to the saliency of a blue word or due to the fact that the blue words are in a hypertext environment.

Experiment 1A: Does Displaying a Word in a Different Colour to the Rest of the Text Change Reading Behaviour?

Experiments 1A and 1B explore whether salient, coloured words impact reading behaviour. Before examining what the impact is of hyperlinks coloured in blue in a hypertext environment, it is important to examine the impact of reading a word in blue in a sentence outside of a hypertext context. If any disruption to reading occurred during the reading of a blue coloured word embedded in a sentence with black coloured words, then it is the saliency of a coloured word that disrupted reading. Therefore, blue hyperlinks in text would also disrupt reading due to the saliency of the blue word by itself, and not just because they are links. In the current experiment, we manipulate the colour of a single target word to explore whether a salient word effects reading and if the colour chosen makes a difference to this. As previously mentioned only 2% of the cones on the retina are sensitive to blue making it supposedly a poor choice in terms of usability (Galitz, 1997), and this could impact on reading behaviour as well if it is more difficult to read text in a certain colour. In line with previous research which has suggested that hyperlinks disrupt reading behaviour (Carr, 2010; Nielsen, 1999), we predicted that the coloured words would be fixated for longer because of the saliency of the coloured word. Several colours were used for the target word to investigate whether blue was indeed particularly disruptive and we also predicted that for grey target words that they would be fixated for longer due to their reduced contrast, thereby making them visually more difficult to process (e.g. Drieghe, 2008).

Experiment 1A: Methodology

Participants

Thirty native English speakers (2 male, 28 female) with an average age of 19.80 years participated in exchange for course credits. All had normal or corrected-to-normal vision and no known reading disabilities.

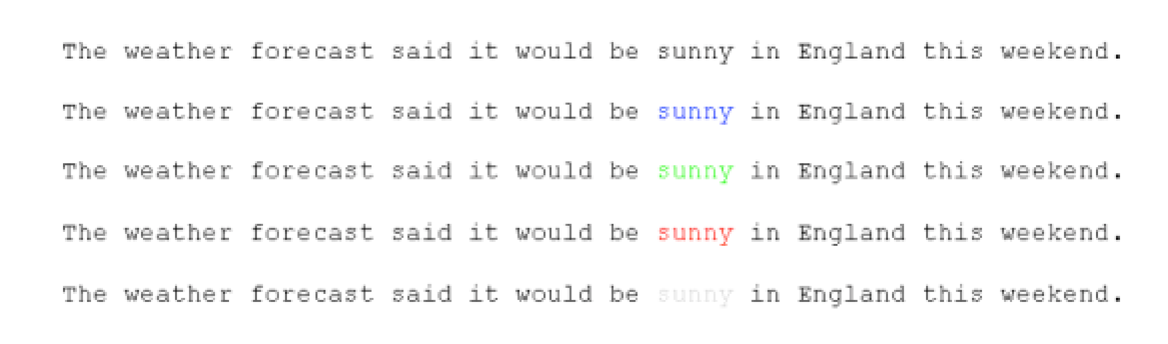


Figure 2.1 Example stimulus from Experiment 1A. This is an example of the five different versions of a single stimulus.

Apparatus

Eye movements were measured with an SR-Research Eyelink 1000 eye tracker operating at 1000 Hz (1 sample every millisecond). Participants viewed the stimuli binocularly, but only the right eye was tracked. Words were presented in 14pt mono-spaced Courier font. The participant's eye was 73 cm from the display; at this distance three characters equalled 1° of visual angle.

Stimuli

Thirty sentences were used and a single target word in each sentence would appear in one of five colours, which correspond to the five experimental conditions (black, blue, green, red or grey; see Figure 2.1). The rest of the sentence was always rendered in black. A counterbalanced design was used in which each participant read one version of each of the thirty sentences with an equal number from each condition. Participants were instructed to read for comprehension and told that they would occasionally have to answer comprehension questions about the sentences. Comprehension questions were presented randomly after 25% of trials, they were simple yes/no questions and the accuracy of answering these was high (97.5% accuracy), indicating that participants were reading the text correctly.

Procedure

Participants were given an information sheet and a verbal description of the experimental procedure and informed that they would be reading sentences on a monitor while their eyes were being tracked. They were told to read for comprehension and that they were to respond to comprehension questions presented after the sentences. The comprehension questions were present to ensure the participants were reading the text for comprehension. The participants were seated in front of the monitor and their heads were stabilised using a head and chin rest to reduce head movements and to ensure reliable eye tracking data. The initial calibration required approximately five minutes before the actual experiment began. At the beginning of each trial a fixation point was presented on the screen where the beginning of the text was set to appear. The participants were required to fixate this point before the trial began to ensure that the first fixation fell on the first word of the sentence. When the participant had finished reading the text on the screen, they pressed a button to move onto the next trial. Each participant first read three practice trials to become familiar with the procedure. The experiment lasted approximately 15 minutes.

Experiment 1A: Results

Eye trackers record a large amount of data (one sample every millisecond) and these data can contain erroneous fixations that are not representative of the dataset. In some cases, these erroneous or outlier fixations will be caused by errors in the calibration eye-tracker (detected by the algorithms used to detect saccades) and in other cases, the participant may have had a lapse of concentration, leading to very long fixation durations. Regardless of the cause, and because we were interested only in those instances when the participants were paying attention and reading the text, we cleaned the dataset before conducting our statistical analyses. In the current experiments we followed the procedures for cleaning our data that have been adopted by the reading research community.

Trials where there was tracking loss were removed prior to the analyses. Fixations shorter than 80ms that were within one character of the previous or following fixation were merged and all fixations shorter 80ms or longer than 800ms were removed to eliminate outliers (resulting in the removal of 4.87% of the total dataset). Finally, when calculating the eye movement measures, data that were more than 2.5 standard deviations from the mean for a participant within a specific condition were removed (<1% of dataset). Data loss affected all conditions similarly.

Around each target word an *interest area* was drawn. The interest area is the size of the target word including the space preceding it. The analyses below are conducted using the fixations that are within the interest area drawn around the target word.

Several eye-movement measures were calculated based on the target word. As mentioned in the Introduction, skipping probability is the probability that a target word does not receive a direct fixation during the first-pass, first-fixation duration is the duration of the initial first-pass fixation on the target word, single fixation duration is the duration of a fixation if the reader only made one first-pass fixation on the target word, gaze duration is the sum of all first-pass fixations on the target word, go past time is the time between first fixating the word and moving past it, to the right and total time is the total amount of time spent on the target word during the whole trial, including any re-reading that might occur.

We ran linear mixed effects models (LMMs) using the lme4 package in R (2009) to explore the impact of the colour of the target words. In reading research, LMMs provide considerable advantages over traditional statistical methods such as ANOVAs. Firstly, LMM's are better able to deal with missing data and imbalanced datasets without losing statistical power (Pinheiro & Bates, 2000). This is particularly important in eye tracking studies as data loss due to rejection of trials due to poor tracking occurs. Likewise, word skipping of a target word will result in missing data for the fixation duration analyses. Secondly, LMMs allow the use of participants and items (individual stimuli) as crossed random effects within one model in comparison to ANOVAs where this requires separate F1 and F2 tests for participants and words (Forster & Dickinson, 1976).

The condition (the colour of the target word) was included as a fixed factor, with treatment contrasts specifying black as the baseline in order to be able to compare the other colours to reading the target word embedded in a plain sentence without a salient target word. Participants and items were included as random effects variables in a so-called maximal random model (Barr, Levy, Scheepers, & Tily, 2013). If a model did not converge, it was reduced by first removing the random effect correlations and then the interactions between the slopes and then by successively removing the random effects explaining the least variance until the maximal converging model was identified. For participants both intercepts and slopes for the effect of colour were allowed to vary. For items only intercepts were allowed to vary.

All the patterns observed in the models were identical whether they were run on log-transformed or untransformed fixation durations, allowing us to present the data run on the untransformed fixation durations in order to increase transparency.

The means for all of the eye movement measures for Experiment 1A are listed in Table 2.1. Participants were significantly less likely to skip a target word when it was not in black (see Table 2.2 for the LMM output). This suggests that the saliency of the coloured target word draws attention to it, making it more likely that participants will fixate it.

However, there were no statistically significant differences in fixation time across the conditions except when the target word was grey. The reduced contrast of the target word increased the fixation time on that word both in early and late eye movement measures compared to any other condition because it was more difficult to process and read (e.g., Drieghe et al, 2008). Also there was a significant difference in the later eye movement measures for when the target word was shown in green.

Participants spent longer on the green target word in total reading time and were more likely to regress back to re-read the preceding text, as shown by the increased go past times. This suggests that the participants also found the green text a bit more difficult to process and we suggest this is also due to the reduced contrast of the green text compared to the other colours used (see Figure 2.1). To verify this, we looked at the luminance of the colours used on the screen during the experiment. Luminance is a measure of the luminous intensity per unit area of light travelling in a given direction. Luminance is measured in candela per square metre (cd/m2). If we look at the luminance of each of the colours used we notice that the grey and green are similar in luminance (grey:80.0 cd/m2; green: 73.2 cd/m2) and closer to the luminance of the white background (103.0 cd/m2) than any of the other colours; blue (10.2 cd/m2), red (18.6 cd/m2) and black (0.7 cd/m2). However, it is not clear why this effect of luminance would manifest itself only in later eye movement measures for the green colour.

These results suggest that salient words are less likely to be skipped because the saliency draws attention to the word making it more likely to be fixated. It could also be that colouring a word provides a signal in the text that suggests that the coloured word is important in some way as suggested previously by Lorch, Lorch and Klusewitz (1995). There were no differences in fixation times across the different

Table 2.1

Means of Eye Movement Measures for Experiment 1A. Standard deviation in parentheses.

Target Word Colour	Skipping Probability (%)	First Fixation Duration (ms)	Single Fixation Duration (ms)	Gaze Duration (ms)	Go Past Time (ms)	Total Reading Time (ms)
Black	27 (19)	222 (36)	229 (54)	242 (47)	282 (106)	284 (66)
Blue	14 (15)	219 (49)	225 (80)	244 (55)	280 (95)	319 (87)
Green	11 (14)	234 (41)	243 (46)	260 (57)	339 (115)	364 (110)
Red	8 (10)	209 (36)	210 (44	229 (47)	279 (59)	305 (91)
Grey	13 (13)	244 (40)	247 (51)	279 (50)	337 (96)	354 (94)

Table 2.2

Fixed Effect Estimates for all Eye Movement Measures for Experiment 1A.

	Skipping Probability			First Fixation Duration (ms)		
	Estimate	Std Error	*z* value	Estimate	Std Error	*t* value
Intercept	-1.28	0.30	-4.34 *	222.63	7.86	28.34 *
Blue	-0.97	0.35	-2.78 *	-4.22	9.30	-0.45
Green	-1.40	0.46	-3.03 *	10.82	8.38	1.29
Red	-1.53	0.37	-4.14 *	-13.98	8.30	-1.69
Grey	-0.98	0.39	-2.54 *	23.14	8.03	2.88 *
	Single Fixation Duration (ms)			**Gaze Duration (ms)**		
	Estimate	Std Error	*t* value	Estimate	Std Error	*t* value
Intercept	221.40	9.38	23.62 *	244.50	10.79	22.67 *
Blue	4.53	14.68	0.31	-3.56	11.15	-0.32
Green	14.67	9.86	1.49	15.07	11.08	1.36
Red	-8.15	10.95	-0.74	-16.91	11.01	-1.54
Grey	34.37	11.80	2.91 *	35.57	11.14	3.19 *
	Go Past Time (ms)			**Total Reading Time (ms)**		
	Estimate	Std Error	*t* value	Estimate	Std Error	*t* value
Intercept	279.04	19.11	14.6 *	282.65	18.55	15.24 *
Blue	0.22	24.90	0.01	32.52	17.74	1.83
Green	57.83	26.56	2.18 *	76.59	17.64	4.34 *
Red	-3.18	23.73	-0.13	19.91	17.52	1.14
Grey	57.75	20.69	2.79 *	73.03	17.74	4.12 *

$^{*}z > |1.96|$

$^{*}t > |1.96|$

colours except from when the target word had reduced contrast (when it was grey or green) which suggests that having a salient coloured word does not affect reading behaviour in terms of processing the coloured word, unless the contrast is reduced making it visually more difficult to process. To reiterate, when the target word was black, blue or red there was no difference in fixation times suggesting that colouring a word does not hinder or help the reading of that word.

Experiment 1A: Discussion

Experiment 1A demonstrated that a coloured word is less likely to be skipped, perhaps because the reader thought the colour serves as a signal that it might be important in some way (Golding & Fowler, 1992; Lorch et al., 1995). Or it could simply be because the coloured word was salient against the rest of the text and attracted the readers' eye (Leyland et al., 2013; Simola et al., 2011; White & Filik, 2009). There was no negative impact on reading behaviour in terms of fixation times when a word was coloured unless the colour was associated with reduced contrast making it more difficult to read as seen when the target word was grey or green.

Experiment 1B: Does Displaying Multiple Words in a Different Colour to the Rest of the Text Change Reading Behaviour?

Experiment 1B follows on from Experiment 1A by including multiple coloured words in a sentence to investigate if the extra salient words have an impact on reading behaviour. If a single coloured word causes a reduction in word skipping for that word, will it occur for all salient words in a sentence when there are multiple? Or will an effect of "over-signalling" occur where the signal of importance is reduced when words are coloured seemingly randomly (Golding & Fowler, 1992; Lorch et al., 1995). Note that on web sites the presence of multiple hyperlinks across the screen will be the default as opposed to a single coloured word.

In Experiment 1A we found no differences between the different colours, all showed the same reduction word skipping probability compared to the word in black and no effect on fixation durations except when the target word had reduced contrast, where fixation durations were increased due to the difficulty of reading a low contrast word. Therefore, in Experiment 1B we only use the colour blue or black for our target words to feature in our coloured word condition. We choose to only use blue or black to represent the colours most often used in a Web environment, where black text tends to represent the unlinked text and the blue text represents the hyperlinked text. We also include a word frequency manipulation to explore whether word difficulty interacts with whether a word is presented in a salient colour (compared to the rest of the text) or not. A reader typically spends more time fixating a difficult or low frequency word than an easy or high frequency word (Henderson & Ferreira, 1990; Inhoff & Rayner, 1986; Rayner & Fischer, 1996). We want to explore if a reader will spend even longer processing a more difficult word if it is also coloured/salient.

Experiment 1B Methodology

Participants

36 native English speakers (17 male, 19 female) with an average age of 25.25 years participated in exchange for course credits. All had normal or corrected-to-normal vision and no known reading disabilities. These participants did not take part in either Experiment 1A or 1C.

Apparatus

The apparatus was identical to the one used in Experiment 1A.

Stimuli

Seventy-two sentences were used and there were six conditions with each participant seeing twelve sentences in each condition. Each sentence contained a target word of which we manipulated the word frequency and how many coloured words were present in the sentence, being either no coloured words, one word (the target word) or three words (the target plus two other words chosen randomly). For the word frequency manipulation where the target word is either high or low frequent the word length was controlled for each pair of stimuli and was either 4 or 5

Lucy put on her dress before going to work in the garden.

Lucy put on her dress before going to work in the garden.

Lucy put on her dress before going to work in the garden.

Lucy put on her smock before going to work in the garden.

Lucy put on her smock before going to work in the garden.

Lucy put on her smock before going to work in the garden.

Figure 2.2 Example stimulus from Experiment 1B. This is an example of the six different versions of a single stimuli. The first three represent the high frequency condition with the target word "dress" and it is displayed with either no, one or three coloured words.

characters in length (on average 4.63 characters). The word frequencies were taken from the Hyperspace Analogue to Language (HAL) corpus (Burgess & Livesay, 1998), which consists of approximately 131 million words gathered across 3,000 Usenet newsgroups. The frequency norms were used to extract both high and low frequency words to create the experimental stimuli. The high frequency words had an average log transformed HAL frequency of 10.16 and the low frequency words has an average log transformed HAL frequency of 6.39 according to the norms collected in the HAL corpus. There was a significant difference between the word frequency in the low and high frequency stimuli $t(71)=28.93$, $p<0.001$.

In total there were six conditions (see Figure 2.2). A counterbalanced design was used in which each participant read all seventy-two sentences with an equal number from each condition. Participants were instructed to read for comprehension and told that they would occasionally have to answer comprehension questions about the sentences.

Comprehension questions were presented after 25% of trials and the accuracy of answering these was high (95.50% accuracy), indicating that participants were reading the text for comprehension.

Procedure

Participants were given an information sheet and a verbal description of the experimental procedure and informed that they would be reading sentences on a monitor while their eyes were being tracked. They were told to read for comprehension and that they were to respond to comprehension questions presented after the sentences. The procedure was identical to Experiment 1A. The experiment lasted approximately 30 minutes.

Experiment 1B: Results

The data cleaning procedure and eye movement measures used were identical to those used in Experiment 1A (resulting in the removal of 4.46% of the total dataset). Finally, when calculating the eye movement measures, data that were more than 2.5 standard deviations from the mean for a participant within a specific condition were removed (<1% of dataset). Data loss affected all conditions similarly.

As in Experiment 1A, around each target word an *interest area* was drawn and we calculated the same eye-movement measures that were used in Experiment 1A: Skipping probability, first fixation duration, single fixation duration, gaze duration, go past times and total reading times.

We ran linear mixed effects models using the lme4 package in R (2009) to explore the impact of word frequency and the impact of the number of salient coloured words in the sentence. Word frequency and number of coloured words were included as fixed factors, with high frequency target word and having no coloured words present as the baseline in order to be able to compare reading a plain sentence with no salient target words to a sentence with salient words and/or a sentence with a low frequency target word. Participants and items were included as random effects variables. A maximal random model was initially specified for the random factors (Barr et al., 2013). If a model did not converge, it was pruned by first removing the random effect correlations, then the interactions between slopes and finally by successively removing the random effects explaining the least variance until the maximal converging model was identified. Model comparisons were also utilised to compare models for the best fitting model for the data for each measure. For most of the measures the best fitting model required the removal of the interaction between word frequency and the number of coloured words. The only exception is the go past time measure where including the interaction gave a better fit for the data. For word skipping the participants and items intercepts were allowed to vary. For the measures of first fixation durations, single fixation durations, gaze durations and total time the intercept for the items variable was allowed to vary, whilst the participant variable included both the intercept and the slope obtained for the additive effects of word frequency and number of coloured words. Finally, for go past times the items variable intercept was allowed to vary, whilst the participant variable included both the intercept and the slope obtained for the interaction between word frequency and number of coloured words, but no correlations between random effects was included. All the patterns observed in the models were identical whether they were run on log-transformed or untransformed fixation durations, allowing us to present the data run on the untransformed fixation durations in order to increase transparency.

The means for all of the eye movement measures for Experiment 1B are listed in Table 2.3. Successive differences contrasts were used such that the intercept corresponds to the grand mean and the fixed factor estimate for a categorical factor can be interpreted as the difference between the two conditions (see Table 2.4 for the LMM statistics). We observe a significant effect in skipping probability and fixation durations for the target word frequency where the low frequency word is less likely to be skipped and has longer fixation durations when it is fixated, replicating previous research (Henderson & Ferreira, 1990; Inhoff & Rayner, 1986). The frequency effect does not interact with whether the target word was coloured or not. We replicate the reduction in skipping observed in Experiment 1A, with a significant difference between whether the target word was coloured or not and less skipping when there was a coloured word (see Table 2.4). However, this effect is only present when there is only a single coloured word. When there are three coloured words the reduction in skipping does not happen and the skipping rates are in line with reading when there are no coloured words in the text. In terms of fixation durations, there was no effect of the number of coloured words in the sentences on fixation duration measures (except in total time) suggesting salient/coloured words have very little impact on reading behaviour. There is a main effect of number of coloured words in total time spent on the target word where significantly more time (38 ms on average) is spent

Table 2.3

Means of Eye Movement Measures for Experiment 1A. Standard deviation in parentheses.

Word Frequency	Number of Coloured Words	Skipping Probability (%)	First Fixation Duration (ms)	Single Fixation Duration (ms)	Gaze Duration (ms)	Go Past Time (ms)	Total Reading Time (ms)
High	0	21 (15)	207 (27)	210 (27)	223 (33)	256 (56)	264 (63)
High	1	17 (14)	203 (30)	202 (36)	215 (37)	253 (82)	268 (72)
High	3	19 (14)	202 (27)	205 (35)	217 (37)	255 (83)	267 (78)
Low	0	15 (12)	219 (29)	229 (40)	245 (41)	282 (69)	293 (74)
Low	1	13 (12)	217 (31)	223 (36)	243 (47)	303 (81)	315 (99)
Low	3	14 (15)	214 (34)	221 (37)	238 (49)	292 (101)	296 (92)

Table 2.4

Fixed Effect Estimates for all Eye Movement Measures for Experiment 1B.

	Skipping Probability			First Fixation Duration (ms)		
	Estimate	Std error	*z* value	Estimate	Std error	*t* value
Intercept	-1.51	0.15	-9.91 *	206.47	3.96	52.13 *
Frequency	-0.40	0.10	-3.79 *	12.96	3.18	4.07 *
No of Links 1	-0.25	0.13	-1.97 *	-2.85	3.75	-0.76
No of Links 3	-0.16	0.13	-1.30	-4.57	3.41	-1.34
Contrast – No of links 1 vs No of Links 3	0.09	0.13	0.67	-1.72	3.22	-0.53
Frequency * No of Links 1						
Frequency * No of Links 3						

	Single Fixation Duration (ms)			Gaze Duration (ms)		
	Estimate	Std error	*t* value	Estimate	Std error	*t* value
Intercept	210.22	4.62	45.55 *	222.56	5.21	42.7 *
Frequency	17.43	3.64	4.78 *	23.71	4.76	4.99 *
No of Links 1	-5.16	4.61	-1.12	-5.28	4.46	-1.18
No of Links 3	-6.44	3.97	-1.63	-6.86	4.53	-1.51
Contrast – No of links 1 vs No of Links 3	-1.29	4.04	-0.32	-1.57	4.08	-0.39
Frequency * No of Links 1						
Frequency * No of Links 3						

	Go Past Time (ms)			Total Reading Time (ms)		
	Estimate	Std error	*t* value	Estimate	Std error	*t* value
Intercept	258.73	9.91	26.10 *	260.57	10.73	24.28 *
Frequency	23.89	10.68	2.24 *	37.32	7.50	4.976 *
No of Links 1	-5.74	12.42	-0.46	13.60	6.80	2.001 *
No of Links 3	-3.58	11.24	-0.32	3.06	6.82	0.45
Contrast – No of links 1 vs No of Links 3	2.16	11.65	0.19	-10.54	6.74	-1.56
Frequency * No of Links 1	28.27	17.36	1.63			
Frequency * No of Links 3	12.93	14.55	0.89			

*$z > |1.96|$

*$t > |1.96|$

on the target words when there is only one coloured word present in the sentence. This suggests that the readers are going back and rereading the single coloured word as we do not see this effect in any of the first-pass reading measures.

Experiment 1B: Discussion

Experiment 1B demonstrated that a single coloured target word is less likely to be skipped and more likely to be revisited in the form of rereading. However, when there are multiple coloured words we do not observe the same reduction in skipping, nor any impact of fixation times. In Experiment 1A and 1B, we thus observe a reduction in skipping but only from a single coloured target word. This is presumably because a single coloured word in a sentence works as a signal of importance to that particular word (Lorch et al., 1995). When there are multiple words being highlighted in the sentence this results in "over-signalling" which can reduce the effectiveness of typographical cues. The reader does not perceive the signal of the coloured text as important if it is not useful to them. When one word is coloured it might suggest that that particular word is important in the sentence. However, when three words in the sentence are coloured at random the signal seems meaningless and serves no use to the reader so is therefore ignored.

An alternative, not mutually exclusive, explanation is that by including multiple coloured words we are reducing the saliency of the coloured target word (Treisman & Gelade, 1980). A single target word is presumably very salient in a single line of text, but less salient when there are other coloured words in that same line of text. Therefore, when the saliency is reduced due to the presence of multiple coloured words in Experiment 1B, we would no longer observe the reduction in skipping of the a single coloured target word observed in Experiment 1A.

Experiment 1C: Does a Hyperlinked Word Change Reading Behaviour? Does Hyperlinking Interact with Lexical Processing?

Experiment 1C explored whether perceiving the words as hyperlinks impairs reading behaviour and if they are processed differently to coloured words in plain text. Previously it has been suggested that the hyperlinked words would be fixated for longer due to the saliency of the blue words (Carr, 2010; Nielsen, 1999). However, because blue hyperlinks are so commonplace in webpages, the processing may become automatic (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). This could mean that in the case of hyperlinks, blue text is automatically processed as being a hyperlink without an additional cost because blue text tends to always be a hyperlink. We also implemented a word frequency manipulation in order to explore whether common lexical effects are present in hyperlinked text and to investigate if they are modulated by the word being hyperlinked. For word skipping, we predict that high frequency words are skipped more often than low frequency words (Inhoff & Rayner, 1986), and that the saliency of a hyperlinked word can draw attention to the word regardless of the frequency and as such increase the chance of fixating the hyperlinked word, reducing skipping for these words.

Hyperlinks indicate that the word links to other content, a signal which can be considered high-level information. The prediction for how this additional high-level information will affect eye movement behaviour comes from the E-Z Reader model of eye movement control (Reichle et al, 2009). It suggests that higher-level processes intervene in eye movement control only when “something is wrong” and either send a signal to stop moving forward or a signal to execute a regression. This is exclusively seen to impact the later eye movement measures so we expect to see hyperlinks impacting fixation durations in the later eye movement measures even though the signal that there is a link present in a word does not necessarily means “something is wrong”.

Experiments 1A and 1B explored the impact of colour/saliency to see if simply colouring a single word or multiple words has an impact on reading behaviour, because hyperlinks are of course firstly coloured/salient words in the text on webpages. However, hyperlinks also represent much more, they are the links to other webpages and are used as a form of navigation. In Experiment 1C, we move forward from Experiment 1A and 1B in that we create trials that resemble a real Web environment, but we do not (yet) allow the readers to navigate, only to read the content presented to them. This will help us tease apart the impact of the visual experience of reading hypertext from the impact of reading and navigating a Web environment with all the navigational decisions that go along with it. The present chapter wishes to explore the impact of the visual experience of hyperlinks before the rest of the thesis builds on these results in more complicated scenarios where the clicking and navigating of hyperlinks adds to the complexity. Reading on the Web is a much more complex task than reading plain text, therefore we need to break the task down and understand the differences to fully develop our understanding of how people read on the Web.

Experiment 1C: Methodology

Participants

Thirty-two native English speakers (3 male, 29 female) with an average age of 19.72 years participated in exchange for course credits. All had normal or corrected-to-normal vision and no known reading disabilities. All these participants also took part in Experiment 1A first.

Apparatus

Identical apparatus to Experiment 1A.

Stimuli

The stimuli in Experiment 1C consisted of twenty edited Wikipedia articles (see Figure 2.3). Eighty target words were embedded in carrier sentences (one target word per sentence) and four carrier sentences were inserted into each Wikipedia article. The font was a mono-spaced font and the line spacing was approximately three-line spacing to improve the discrimination of fixations between lines. The target words within these articles were either displayed in blue or black to denote if the word was a hyperlink or not, respectively. There was also a word frequency manipulation where the target word is either high or low frequent. The word frequencies were taken from the Hyperspace Analogue to Language (HAL) corpus (Burgess & Livesay, 1998), the frequency norms were used to extract both high and low frequency words to create the experimental stimuli. The high frequency words had an average log transformed HAL frequency of 9.91 and the low frequency words has an average log transformed HAL frequency of 5.75. There was a significant difference between the high and low word frequency stimuli, $t(79)=24.61$, $p<0.001$. The target words were matched on word length and were between 4 and 7 characters

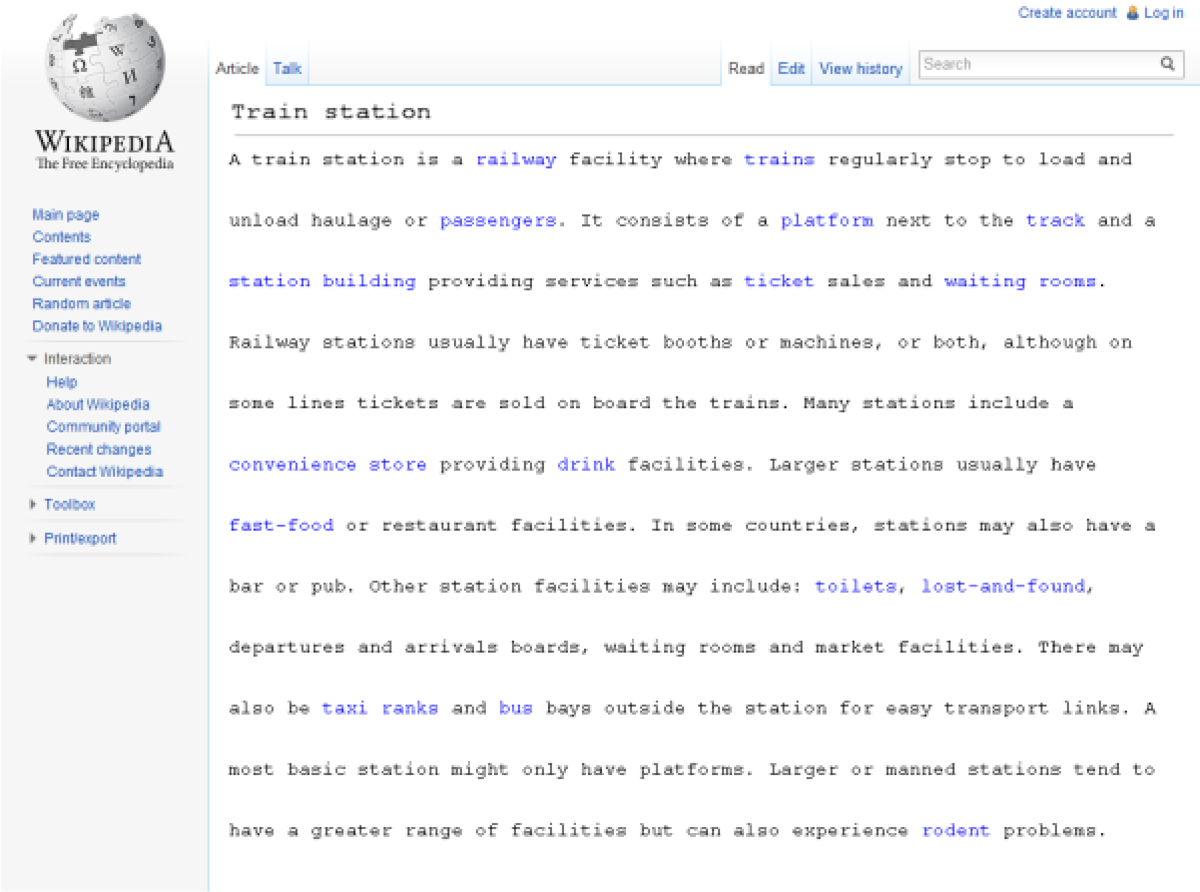


Figure 2.3 Example stimulus from Experiment 1C.

in length (5.24 characters on average). In total there are 4 conditions in a 2 x 2 design within participants. Participants were again instructed to read for comprehension and told that they would be have to answer comprehension questions about the sentences, these appeared after all trials. The experiment lasted approximately 30 minutes.

Procedure

Participants were given an information sheet and a verbal description of the experimental procedure and informed that they would be reading passages on a monitor while their eyes were being tracked. They were told to read for comprehension and that they were to respond to comprehension questions presented after the sentences. The comprehension questions were simple and required a yes or no response. The comprehension question where present to ensure the participants were reading the text displayed to them for comprehension. The total accuracy for the comprehension questions was 95.31%. The participants were seated in front of the monitor and their heads were stabilised using a head and chin rest to reduce head movements and ensure reliable eye tracking data. The initial calibration required approximately five minutes before the actual experiment began. At the beginning of each trial a fixation point was presented on the screen where the beginning of the text was set to appear. The participants were required to fixate this point before the trial began to ensure that the first fixation fell on the first word of the sentence. When the participant had finished reading the text on the screen, they pressed a button to move onto the next trial. Each participant first read two practice trials to become familiar with the procedure. Comprehension questions were presented on every trial to ensure each article was read in full and the experiment lasted approximately 45 minutes.

Experiment 1C: Results

The data cleaning procedure and eye movement measures used were identical to that used in Experiment 1A and 1B (resulting in the removal of 4.47% of the total dataset). Finally, when calculating the eye movement measures data that were more than 2.5 standard deviations from the mean for a participant within a specific condition were removed (<1% of dataset). Data loss affected all conditions similarly.

As in Experiment 1A and 1B, around each target word an *interest area* was drawn. We calculated the same eye-movement measures that were used in Experiment 1A and Experiment 1B: Skipping probability, first fixation duration, single fixation duration, gaze duration, go past times and total reading times.

We ran linear mixed effects models using the lme4 package in R (2009) to explore the impact of word frequency and the target word being displayed as a linked or unlinked word. Word Frequency and Word Type (whether the word was linked or unlinked) was included as fixed factors. Participants and items were included as random effects variables. A maximal random model was initially specified for the random factors (Barr, Levy, Scheepers, & Tily, 2013). If a model did not converge, it was reduced by first removing the random effect correlations, the interactions between the slopes and finally by successively removing the random effects explaining the least variance until the maximal converging model was identified. For most of the measures the intercept for the items variable was allowed to vary and the participant variable included both the intercept and the slope obtained for the interactive effects of Word Frequency and Word Type. The only exceptions were skipping proportion and gaze duration where the intercept for the items variable was allowed to vary and the participant variable included both the intercept and the slope for the effect of Word Frequency. The effect of Word Type had to be remove to allow the models to converge. All the patterns observed in the models were identical whether they were run on log-transformed or untransformed fixation durations, allowing us to present the data run on the untransformed fixation durations in order to increase transparency. Successive differences contrasts were used such that the intercept corresponds to the grand mean. The means for all of the eye movement measures for Experiment 1C are listed in Table 2.5 and the LMM statistics in Table 2.6.

The effect of Word Frequency was present in all fixation-based eye movement measures. However, for word skipping, the high frequency words were numerically skipped more often than the low frequency words but this effect did not reach statistical significance. The high frequency words also had significantly shorter fixation times than the low frequency word (16 ms shorter) when they were fixated. This mostly replicates previous experiments that have demonstrated that low frequency words are skipped less often and have longer fixations times because they are more difficult to process than highly frequent words (Inhoff & Rayner, 1986). However, in go-past times and total time the effect of Word Frequency was qualified by an interaction with Word Type (Figure 2.4) which we will discuss in detail below.

Replicating the results of Experiment 1A and 1B for coloured words, there were no significant differences between whether the target word was linked or not in the early fixation duration measures (first fixation durations, single fixation durations and gaze durations). This suggests that having a word linked does not make it any more difficult to process in first-pass reading. However, there was also no difference in the amount of skipping observed in Experiment 1C between when the target word was linked or not. This is especially interesting because in Experiment 1A and Experiment 1B when a word was coloured it was less likely to be skipped. However, in Experiment 1B when there was more than one word coloured in a sentence, the reduction in skipping observed for single coloured words did not exist. If the reduction in skipping of a coloured word is limited to when there is only one word

Table 2.5

Means of Eye Movement Measures for Experiment 1C. Standard deviation in parentheses.

Word Frequency/Word Type	Skipping Probability (%)	First Fixation Duration (ms)	Single Fixation Duration (ms)	Gaze Duration (ms)	Go Past Time (ms)	Total Reading Time (ms)
High/Linked	43 (23)	216 (34)	214 (35)	227 (36)	298 (123)	261 (54)
High/Unlinked	46 (24)	215 (32)	219 (40)	228 (38)	291 (84)	266 (67)
Low/Linked	40 (24)	232 (45)	249 (47)	258 (47)	364 (112)	320 (68)
Low/Unlinked	43 (24)	231 (38)	238 (44)	251 (47)	306 (88)	297 (60)

Table 2.6

Fixed Effect Estimates for all Eye Movement Measures for Experiment 1C.

	Skipping Probability			First Fixation Duration (ms)		
	Estimate	Std error	*z* value	Estimate	Std error	*t* value
Intercept	-0.33	0.22	-1.52	222.98	5.46	40.88 *
Word Frequency	-0.15	0.09	-1.67	18.71	5.12	3.65 *
Word Type	-0.12	0.09	-1.36	1.18	5.12	0.23
Word Frequency * Word Type	0.03	0.18	0.16	4.20	8.48	0.50
	Single Fixation Duration (ms)			**Gaze Duration (ms)**		
	Estimate	Std error	*t* value	Estimate	Std error	*t* value
Intercept	229.36	6.16	37.23 *	239.74	6.67	35.95 *
Word Frequency	28.77	5.99	4.8 *	32.45	4.49	7.22 *
Word Type	3.55	4.99	0.71	1.75	5.53	0.32
Word Frequency * Word Type	14.02	8.47	1.66	13.50	8.97	1.51
	Go Past Time (ms)			**Total Reading Time (ms)**		
	Estimate	Std error	*t* value	Estimate	Std error	*t* value
Intercept	311.22	12.93	24.06 *	286.58	9.96	28.78 *
Word Frequency	40.25	14.92	2.70 *	46.06	9.11	5.06 *
Word Type	20.14	14.77	1.36	7.31	8.67	0.84
Word Frequency * Word Type	49.47	23.31	2.12 *	31.94	14.03	2.28 *

*$z > |1.96|$

*$t > |1.96|$

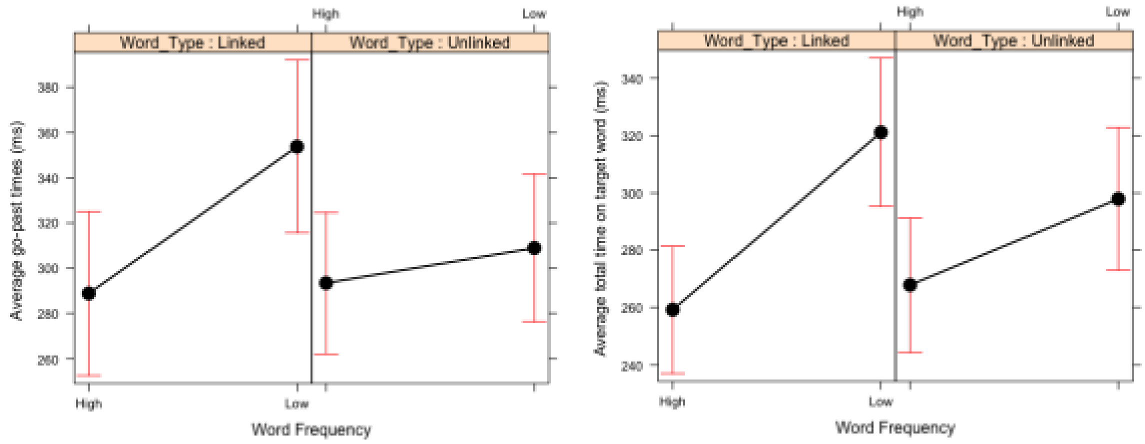


Figure 2.4 Two-way interaction between Word Frequency x Word Type interaction in Experiment 1C. Means and standard error bars for go-past times and total reading times.

coloured, this would explain why we do not observe the reduction of skipping in Experiment 1C where there are multiple coloured/linked words, similar to how we did not observe the effect in Experiment 1B which featured trials that had multiple coloured words.

In the later eye movement measures of go-past times and total reading times we observed a significant interaction between Word Frequency and Word Type. As Figure 2.4 shows, the low frequency linked words had significantly longer go-past times and total reading times on the target word compared to the other three conditions. This interaction was only present in the late eye movement measures which suggest that the low frequency linked word causes regressive eye movements due to difficulty in processing, in other words the participants are reading the low frequency linked words and rereading the preceding content to re-evaluate it.

Experiment 1C: Discussion

Experiment 1C demonstrated that coloured words in passages of text in an environment resembling the Web do not have a negative effect on the early measurements of reading behaviour. In Experiment 1C, we observed a main effect of word frequency where low frequent words had longer fixation times than high frequent words. Importantly, this frequency effect was also qualified by an interaction with whether the target word was hyperlinked or not in the later eye movement measures. There were significantly longer fixation times in go-past times and total reading times but only when the target word was hyperlinked and low frequent (see Figure 2.4). This finding is compatible with the E-Z Reader model of eye movement control (Reichle et al, 2009) that suggests that high-level processes only intervene in eye movement control when the processing is difficult and this is exclusively observed in the later eye movement measures, as observed in Experiment 1C. This might also explain why there are no differences in the fixation times when the target word is hyperlinked and high frequent, only when the target word was hyperlinked and the word itself difficult (i.e. low frequent) was the processing difficulty sizeable enough to trigger re-reading.

General Discussion

By conducting these three experiments together we have examined the potential differences between reading text with and without a coloured word embedded in the sentence and reading text with a hyperlinked word embedded. Experiment 1A demonstrated that a coloured word is less likely to be skipped when it is the only coloured word in the sentence, but that making a word coloured does not negatively impact reading behaviour as measured in fixation durations unless the colour has reduced contrast making it difficult to read, as seen when the target word was grey or green.

Experiment 1B replicated Experiment 1A and showed a reduction in skipping of the target word when it was coloured. However, this reduction of skipping was not observed when there were multiple coloured words within the sentences displayed. There was also a standard frequency effect showing less skipping of and longer fixations on low frequency words (Henderson & Ferreira, 1990; Inhoff & Rayner, 1986), but this did not interact with whether the word was coloured. There are two possible suggestions to why there was a reduction in word skipping for a single coloured word, but not when there are multiple coloured words in the text. One suggestion is there could be an effect of over-signalling, where the importance of the coloured words gets reduced when there are too many words coloured randomly and the reader could not utilise the signal as it was meaningless to them (Golding & Fowler, 1992; Lorch et al., 1995). The alternative suggestion is that the saliency of the target word is reduced when there are multiple coloured items present in the sentence (Treisman & Gelade, 1980). A single target word is presumably very salient in a single line of text, but less salient when there are other coloured words in that same line of text. If a single word is salient it may draw the reader's attention, however, reduce that saliency and the effect disappears.

Experiment 1C demonstrated that there is a difference between processing a word that is coloured during reading and a hyperlinked word. We observed a significant difference in go-past times and total reading times between whether the target word was hyperlinked or not in the reading of the Wikipedia pages, and this effect was qualified by an interaction with frequency. The hyperlinked, low frequency words had longer fixation times in these measures which indicated that the reader had difficulty integrating and processing the low frequency word when it was hyperlinked. As a result, participants were more likely to reread the preceding content to re-evaluate it. A hyperlink on a word indicates that that word is important and implies there is more information behind the hyperlink regarding the word or topic that the hyperlinked word corresponds to. When the hyperlinked word is a low frequency word the reader may wonder why that word is hyperlinked and want to re-evaluate the preceding content presumably to make sure that they understood it, or try to decide why it is important.

The longer reading time in the later eye movement measures on low frequency, hyperlinked words can also be discussed from a signalling perspective. We found in Experiment 1B that if most of the text has some form of signal to cue the importance of the information, then the signal might not be as effective or as informative compared to when only the most important text is signalled (Lorch et al., 1995). The signal needs to be useful to the reader in order for them to try and utilise it (Golding & Fowler, 1992). In Experiment 1B the signalling served no purpose, but in a hypertext-like environment, like in Experiment 1C, it could serve a purpose. For example, the hyperlinked words could be suggesting that those pieces of text are important parts of the text and the reader may want to pay attention to them. When the target word was low frequency and hyperlinked they may have wondered why that particular word was linked and want to reassess the information preceding it to evaluate it.

In Experiment 1A a single coloured word was significantly less likely to be skipped than black words, this is replicated in Experiment 1B where there are single coloured words in the text. However, in Experiment 1B when there were multiple coloured words in the text, there was no similar reduction in skipping rates. Similarly, in Experiment 1C there were also no differences in skipping rates when the word was hyperlinked/coloured versus when it was unlinked. This suggests that the reduced skipping of the coloured target words observed in Experiment 1A is likely restricted to when there is only one word that is coloured. When multiple words are coloured (even when this is in other trials but in the same experiment) the coloured word will presumably be perceived as less attention grabbing. Moreover, in Experiment 1C the colouring did serve a purpose (it indicated a hyperlink) although our current design does not allow is to tease apart whether the absence of a decreased skipping of coloured words in Experiment 1C compared to 1A is due to the presence of multiple coloured words reducing the coloured words saliency and/or because the colouring was meaningful as a signal.

There is also the issue to consider that Experiment 1A and Experiment 1B were run on single sentences, whereas Experiment 1C was run on passages of text. One study found that total viewing times for words were greater for passages of text, however, first pass viewing times were shorter (Radach, Huestegge, & Reilly, 2008; Wochna & Juhasz, 2013). This suggests that when the text is displayed as a passage the reader will make a quick first pass over the text and then engage in re-reading, a behavioural pattern not observed or at least not to the same extent when the participants read the text as single lines. Note however, that this difference between sentence and paragraph reading will not influence our findings of the influence of hyperlinks in Experiment 1C.

These experiments have shown that coloured text in itself does not hinder reading, but also that coloured hyperlinks can cause the participant to reread previous content if the word is a low frequent/difficult word in order to re-evaluate the content. Although the participants could not click and navigate the hyperlinks, this experiment at least shows that having hyperlinks shown as salient blue text does not negatively impact on reading behaving. It does increase re-reading when a reader reaches a low frequency, hyperlinked word, but this is not necessarily a negative behaviour. Because there is no effect on earlier reading measures this suggests that having coloured words in the text does not make reading the text difficult. However, the hyperlinks do signal that the hyperlinked words are special/important in some way. Gagl (2016) has already replicated our findings in a follow-on study from this research (Fitzsimmons, Weal, & Drieghe, 2013). Gagl conducted a boundary paradigm experiment where there was no preview effect of colour. However, in re-reading measures, a reduced fixation time for black words compared to blue words. This is similar to the finding in Experiment 1C. This suggests there is no detectable perceptual disadvantage of coloured words, but increased attraction of attention resources, after first pass reading. Blue text does not hinder the readers' ability to easily perceive the text, yet at the same time, the blue text highlights the word so that it can easily be revisited.

In terms of Web design and layouts, the present results highlight the importance of carefully considering which words are hyperlinked in webpages. The key lesson for Web designers that we have found here is that coloured words do not have any negative impact on reading behaviour. This is the case no matter the colour, unless the contrast between the text colour and the background colour is low, as seen in the longer fixation times on the low-contrast grey/green words in Experiment 1A. Therefore, efforts made in Web development to avoid using blue as the hyperlink colour and instead using a different colour may have no positive influence for the reader reading the text, but instead make it more difficult for the reader to know what is a hyperlink when they are expecting it to conform to the convention of hyperlinks being denoted in blue.

The present set of experiments represent the first steps in understanding how we read hyperlinked text. A hyperlink is not just a salient word in a passage of text, it denotes that more information that may be relevant lies behind that hyperlink. Enriching hypertext documents with large numbers of links that are automatically generated (as seen in Milne & Written, 2008) may cause disruption to reading behaviour. Hyperlinks do not necessarily cause a disruption to reading, but our research here has shown that if you hyperlink a low frequent/difficult word when there is not a strong reason to do so there is re-reading of the content to assess why that particular word is hyperlinked.

Even though in the current experiments participants only engaged in reading behaviour and did not have to make decisions and click any hyperlinks, there was still a significant difference between reading a coloured word and a coloured word in a Web-like environment. The suggestion of a blue word representing a hyperlink is enough, in a Wikipedia/Web environment, to influence eye movement behaviour in all likelihood associated with the processing of a hyperlink even without the ability to click the hyperlinks. However, we will explore the effects of a clickable, navigable Web environment in Chapter 6.

The current chapter suggests that readers are placing importance on the hyperlinked words. In the next chapter, the importance placed on text in a hypertext environment is explored. If hyperlinks within the text suggest where important information might be in the text, then the reader would rate sentences with hyperlinks as more important than text without it. This is explored in depth in the next chapter.

Chapter 3

Rating the Importance of Sentences containing Hyperlinks while Reading on the Web

Introduction

Current literature suggests that reading on the Web is very likely to involve skim reading (Liu, 2005; Morkes & Nielsen, 1997). Liu (2005, pp.700) suggests that screen-based reading behaviour is characterised by 'more time spent browsing and scanning, keyword spotting, one-time reading, non-linear reading, and reading more selectively, while less time is spent on in-depth reading, and concentrated reading.'. Chapter 3 explores how aspects of reading on the Web, such as the presence of hyperlinks in the text, might influence how we read text and judge which information is important.

There is a great deal of evidence suggesting that during skim reading some comprehension may be lost compared to more in-depth reading (Carver, 1984; Dyson & Haselgrove, 2000; Just & Carpenter, 1987; Masson, 1982). However, this loss in comprehension is not consistent across all of the text being read. There appears to be a difference between information regarded as important versus unimportant. The important information (as rated by independent participants) appears not to receive the same loss of comprehension that is observed for the unimportant information (Duggan & Payne, 2009; Masson, 1982; Reader & Payne, 2007). To explain these findings, it has been suggested that the reader engages in an adaptive strategy in order to gain as much information from the text as possible, in a reduced time. But in order to do this the reader must judge which pieces of information are important and which pieces are not. This chapter focuses on how the reader might use cues in the text, such as hyperlinks, to suggest where important information might lie in the text. This suggestion will now be discussed in detail in the next section.

Foraging and Satisficing Strategy

When reading text, it may be difficult to judge how important something is until you have read and understood it. However, there might be specific cues that the reader can use to predict importance, especially when reading on the Web in a hypertext environment. More specifically, hyperlinks are salient and may highlight important sections of text. Pirolli and Card (1999) use a metaphor of a bird foraging for berries as an example of information foraging. A bird is foraging for berries in patches of bushes. The bird must decide how long to spend on one patch of bushes before moving onto a new patch to spend time finding berries. The most efficient time to leave for a new patch is when the expected future gains from foraging in the current patch decreases to such a level that it is better to spend time moving to a new patch, where the future gains may be greater.

In terms of information foraging while reading on the Web, we could consider that the important information is the goal such that important information represents the number of berries. The hyperlinks on a page would be a predictor of whether there are enough berries, in other words whether there is enough important information present and as such, the reader could use the links to find important information. Once a section of text has been exhausted, the new ‘patch’ could represent a new line, new paragraph or new webpage that needs to be foraged.

The current experiment explores what kind of cues the reader uses to judge the importance of the text and whether hyperlinks would be used to help find the important information. If this is so, hyperlinks could potentially be used as a typographical cue to signal important information.

Signalling

Typographical cues, such as boldface or underline serve to highlight a word or small section of text, while organisational cues, such as headings or numbering serve to organise the text into an easy to read structure. Research has shown that making a keyword or phrase distinct in the text results in the reader paying more attention to the emphasised content when reading (Lorch, Lorch & Klusewitz, 1995) and often results in better memory for those emphasised pieces of text (Cashen & Leicht, 1970;

Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch et al, 1995). Hyperlinks are salient words or phrases in the text and we have already demonstrated that readers treat hyperlinks differently to the other words in the text. We observed this in Experiment 1C, where the readers showed a significantly longer fixation times in later eye movement measures when the target word was hyperlinked and low frequent. This suggests that the hyperlinked words could be indicating that those pieces of text are the most important and the reader may want to pay attention to them. When the target word was low frequency and hyperlinked the reader may have wondered why that particular word was linked and want to reassess the information preceding it to re-evaluate it. Perhaps this is because the hyperlinks serve as a signal to where the important information might lie in a text. If important sentences contain more hyperlinks and the hyperlinks are salient enough to be used as a signal when reading, then hyperlinks may prove a very useful typographical cue for the readers. Firstly, the signalling may help highlight the important sections of text (Lorch, 1989). Secondly, the signalling may aid the memory of these important sections (Cashen & Leicht, 1970; Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch et al., 1995). Indeed, we also need to consider the evidence that instead of exclusively making their own judgments from the semantic content of the text, participants will rely on signals to communicate importance (Lorch et al., 2011). Lorch, Lemarié and Grant (2011) found that participants relied more on the demarcation than the semantic content of the text. In the control condition (with no demarcation signals), participants selected sentences from paragraphs that introduced major text topics. However, when there were asterisks before major text topic boundaries participants selected the sentences with asterisks significantly more often than in the control group. A similar effect was observed for the participants who had asterisks before minor text topic boundaries. These participants significantly favoured the minor text topic sentences that had the asterisk demarcation. The asterisks seemed to override the appropriate interpretation of the text. This suggests that signals can override semantic cues from the text and that participants are willing to believe the signalling rather than make their own objective judgements. We might see a similar effect when considering that hyperlinks in the text are salient and may be used as a signal of importance.

Experiment 2: How Do the Properties of the Text on a Webpage Affect the Importance Rating of the Content?

Experiment 2 explored how readers allocate importance to the different sentences within passages of text (in this case, edited Wikipedia articles). One of the factors this experiment examined was whether the importance rating is modulated by the presence of hyperlinks or whether importance is solely guided by the semantic content of the text. We might assume that hyperlink signals could be viewed in a similar way and render the hyperlinked text more influential to the participant than it would be based on the semantic content of the text (Lorch et al., 2011).

In order to measure the impact of hyperlinks we conducted the experiment on two groups of participants. In the first group participants viewed passages of text in a Wikipedia environment. Links were present in the text and the participants had to rate each sentence for its overall importance in that passage. Selected target words were either hyperlinked or not. In the second group, another group of participants saw the same passages of text in a Wikipedia environment, but all of the embedded links throughout all the passages of text were removed. Again, the participants had to judge the importance of each of the sentences, but in this experiment the participants would not be influenced by any links in the text. By doing this, we were able to separate the impact of hyperlinks from the importance of other textual and content factors.

Two additional factors besides the presence or absence of hyperlinks that we wanted to take into consideration in this experiment were the length of the sentence and the sentence position on the page. We predicted that the length of the sentence would have an impact on the ratings for each sentence, with longer sentences being rated as more important because of a bias very similar to the so-called information bias. Information bias is when someone believes that the more information that can be acquired to make a decision, the better, even if that extra information is irrelevant for the decision (e.g Baron, Beattie, & Hershey, 1988). So regardless of the semantic content of a sentence, if the sentence is longer the participant will be biased to think it must be important because there is more information. As for the sentences position on the page, it is a common mass media writing style to write articles with the most important information at the top and the least important at the bottom. This is called the "inverted pyramid" structure (Scanlan, 2003). Essential information such as the Who, What, When, Where, and Why is included in the lead paragraph and communicates the essential facts. Additional details, background, or other information are added to the article in order of importance, such that the least important items are at the bottom. The inverted pyramid originates from old media technology such as the telegraph. In the past news outlets would telegraph information over the wires, it made sense to use the inverted pyramid because the most vital information in the story was transmitted first (Scanlan, 2003). This has been a very common media writing style and the majority of readers is familiar with it (even though often not consciously) and knows what to expect when reading articles. Therefore, we predict that readers will rate information at the top of the page as more important than the information at the bottom of the page because that is the common writing style that is often adopted.

In terms of hyperlinks, we predict that when there are no hyperlinks present in the text the participants will exclusively use the semantics within the text to judge the importance of each sentence. As shown by Lorch, Lemarié and Grant (2011), when signalling is absent the reader instead has to rely on the semantic information in the text to suggest importance. For the conditions where the hyperlinks are present, we predict that the number of hyperlinks may have an impact on the participants' rating judgements. Specifically, they may believe that the sentence is more important as it contains more hyperlinks. Signalling research suggests that if a typographical signal is present that highlights a small portion of text this can result in the reader paying more attention to the emphasised content when reading. However, the number of links may have a limit on how important it can make a sentence. If something is 'oversignalled' it can reduce the effectiveness of typographical cues (Lorch, Lorch & Klusewitz, 1995). The reader might not perceive the linked text as important if a large proportion is linked. So when just a few words are linked, it might suggest that that particular word is important in the sentence. However, when many words in the sentence are linked, the signal importance could be reduced.

Methodology

Participants

Fifty native English speakers (Linked Experiment: 3 male, 29 female, with an average age of 20.22 years; Unlinked Experiment: 2 male, 16 female, with an average age of 20.33 years) participated in exchange for course credits or payment (£6). All had no known reading difficulties.

Apparatus

Participants were seated in an experimental cubicle in front of a desktop computer monitor and a laptop computer. On the desktop computer the participants were presented with each edited Wikipedia article in its entirety. Meanwhile, the laptop screen displayed each sentence from the edited Wikipedia article individually. The sentences appeared one at a time on the laptop screen in the same order as the edited Wikipedia article shown on the desktop computer monitor. The reason for this setup was because in order to rate the importance of each sentence the reader needs to judge it based on the semantic content of the entire article. Therefore, each participant was instructed to read the whole Wikipedia article on the desktop display and then rate each sentence for that article on the laptop. The experiment was self-paced and they could take as long as they needed to read the articles and then rate the sentences.

Stimuli

The stimuli consisted of forty edited Wikipedia articles. Twenty were reused from Experiment 1C and an additional twenty were created using the same approach for generating them as for the original stimulus set. In total one hundred and sixty target words were embedded in sentences (one target word per sentence) and four carrier sentences were inserted into each Wikipedia article. The target words had a 2 x 2 design whereby Word Frequency and Word Type (whether it was linked or unlinked) were implemented as within-subject factors in a Latin square design. However, in this experiment the importance ratings of the sentences were not analysed as a function of the manipulations on the target word. This is because in this experiment we wanted to analyse all the sentences (also the ones without target words) as a function of number of links, position on the screen and length of the sentence. We did include the target word manipulations in the experiment because we used the obtained importance ratings for analysing eye movement data on these target words in Chapter 4. For completeness, we did check whether the manipulations on the target words significantly influenced the importance ratings of the entire sentences, but this was not the case (all *F*'s < 1, n.s.). So for the sentences that did feature a target word the data are presented collapsed across target word conditions. The font was a mono-spaced font (Courier New) and the line spacing was approximately three-line spacing. These stimuli were also used in a subsequent eye tracking experiment (Experiment 3), therefore the mono-spaced font and increased line spacing are implemented to assist analysis and discriminating fixations between lines. For one group of participants the links were included in the stimuli; for the other group, no hyperlinks were included (see Figure 3.1). This was chosen to be a between participant's design because the effect of seeing hyperlinks in some articles and not others may affect the rating given by the participants. To be more precise, we did not want the ratings given to the unlinked sentences to be influenced by the presence of hyperlinks in other sentences during the same experiment. Indeed, if hyperlinks do indeed indicate important sentences, then it is not inconceivable that adding them to certain sentences on the screen, would actually decrease the perceived importance of the sentences that do not feature hyperlinks. Indeed, this setup will allow us to test this hypothesis.

For completeness we will describe the properties of the target words. The target words within these articles were either displayed in blue or black to denote if the word was respectively a hyperlink or not. There was also a word frequency manipulation where the target word was either high or low frequency. The word frequencies were taken from the Hyperspace Analogue to Language (HAL) corpus. The frequency norms were used to extract both high and low frequency words to create the experimental stimuli. The high frequency words had an average log transformed HAL frequency of 9.94 and the low frequency words had an average log

Figure 3.1 Example stimulus from Experiment 2. 1) Stimulus containing hyperlinks, 2) Stimulus with hyperlinks removed.

transformed HAL frequency of 5.81 (according to the norms collected in the HAL corpus, Burgess & Livesay, 1998). There was a significant difference between the high and low word frequency stimuli, $t(159)=29.66$, $p<0.001$. All target words were 4-7 characters in length with an average of 5.60 characters. Word length was matched for each of the high/low frequency pairs, but differed between experimental sentences. The various versions of each stimulus were controlled with a Latin square design, meaning every participant saw only one version of each edited Wikipedia article. To be clear, the overall presence of hyperlinks was implemented between-participants, the presence of hyperlinks on the target words was manipulated within-subjects but only for the group of participants that were presented with hyperlinks in the experiment.

Procedure

This experiment did not involve eye tracking. Participants were given an information sheet and a verbal description of the experimental procedure and informed that they would be reading Wikipedia articles on the desktop computer screen. The participants were instructed to read through the entire Wikipedia article on the desktop screen and then move to the laptop screen to rate the importance of every sentence from that article. They were instructed to rate each sentence on how important it was to the general meaning of the article as a whole and respond using the buttons 1 (Not important) – 5 (Very important) on the laptop keyboard. The experiment was self-paced and lasted approximately 60 minutes.

Results

The results are broken down along three main research questions. The main aim was to explore the impact of the presence of hyperlinks on the importance of that text, but we are also aware that other factors that could affect the ratings, such as sentence length and position on page.

First, as mentioned in the Introduction, we wanted to explore if the presence of any hyperlinks in the body of texts had any impact on the importance rating of the unlinked sentences. To do this we compared the unlinked sentences in the Linked Experiment versus the unlinked sentences in the Unlinked Experiment. These sentences are identical and should be rated the same if the participant was using the same sentences' qualities to rate them. However, if the participant is using the presence of links in a sentence to suggest importance, then the unlinked sentences may actually be rated lower in comparison. In this case, we should see the unlinked sentences in the Linked Experiment rated lower than when they are in a scenario without any links throughout the Unlinked Experiment.

Our second research question was whether the presence of links in a sentence had any direct impact on the importance rating. We explored the Linked Experiment and compared whether the presence or absence of links affected the importance rating. If the participants were using hyperlinks to suggest importance we would see higher ratings for the sentence containing links compared to those without links.

Finally, we wanted to investigate whether once links are present, the number of links has any effect of how important a sentence is rated. We explore the Linked Experiment and only look at the sentences with one or more links and evaluate if the number of links has an impact on the importance ratings. So for this investigation number of links was examined as a continuous variable after removing the ratings for the sentences that did not feature any links.

In all of these analyses we also took into consideration factors such as the length of the sentence and its position on the page, therefore we included these in all of our analyses. Because we did not have strong predictions about potential interactions between the three factors we explored (linked/unlinked, position on the screen, length of the sentence), model comparisons will be carried out to see whether the interactions add to the fit of the statistical models and the most parsimonious model will be reported.

How does the Presence of Links Affect the Importance of Unlinked Text?

First, we ran linear mixed models using the lme4 package in R (2009) to explore the differences between the ratings for the unlinked sentences for the group of participants which had passages presented that contained links versus the group that had those links removed. The Experiment Type (whether it was the Linked Experiment or the Unlinked Experiment), Length of Sentence in characters (which was centered) and Position on the Page (which was also centered) were included as fixed factors. Participants and items were included as random effects variables. A 'maximal' structure was initially specified for the random variables (Barr et al., 2013), but the model failed to converge. If a model did not converge, it was reduced by first removing the random effect correlations and then by removing the interactions between slopes and finally by successively removing the random slopes explaining the least variance until the maximal converging model was identified. In the reported analyses the intercept was allowed to vary for items, and for participants the intercept plus the slope obtained for Position on Page. For the fixed factors successive differences contrasts were used such that the intercept corresponds to the grand mean and the fixed factor estimate for a categorical factor can be interpreted as the difference between the two conditions.

There was a main effect of Experiment Type where unlinked sentences in the Unlinked Experiment were rated higher than those in the Linked Experiment (Unlinked: 3.23; Linked: 3.10, see Table 3.1. for the LMM's). This suggests that when an unlinked sentence is presented with other sentences that have links, the unlinked sentence is rated lower compared to when it is featured in a context without any hyperlinks. There is also a main effect of Length of Sentence where the longer sentences are rated higher and a main effect of Position on Page where sentences closer to the top of the page are rated higher (see Table 3.1 for the signs of the fixed effects estimates).

These main effect are qualified by both a two-way interaction between Length of Sentence x Position on Page (Figure 3.2) and a three-way interaction between Experiment Type x Length of Sentence x Position on Page (Figure 3.3).

The two-way interaction between Length of Sentence x Position on Page (see Figure 3.2) is occurring because as we see in Figure 3.2 for sentences on top of the page long sentences are rated as more important than short sentences whereas for sentences at the end of the page long sentences are rated as less important than short sentences. However, this two-way interaction is qualified by a three-way interaction that also includes whether the experiment included links or not. Figure 3.3 clearly shows that this two-way interaction we just described is far more pronounced for the experiment that contains links than the experiment that does not contain links. Note the slope of the full line (linked) changes more drastically as the position of the page increases than the dashed line (unlinked).

Table 3.1

Fixed Effects Estimates, Standard Error and t value for LMM model for Experiment 2 comparing the unlinked sentences in the Linked and Unlinked versions of the experiment.

	Estimate	Std. Error	*t* value
Intercept	3.09	0.05	61.61 *
Experiment Type	0.16	0.03	4.54 *
Length of Sentence	0.00	0.00	6.69 *
Position on Page	-0.24	0.01	-19.8 *
Experiment Type x Length of Sentence	0.00	0.00	-0.26
Experiment Type x Position on Page	0.02	0.02	1.23
Length of Sentence x Position on Page	0.00	0.00	-5.47 *
Experiment Type x Length of Sentence x Position on Page	0.00	0.00	2.02 *

*$t > |1.96|$

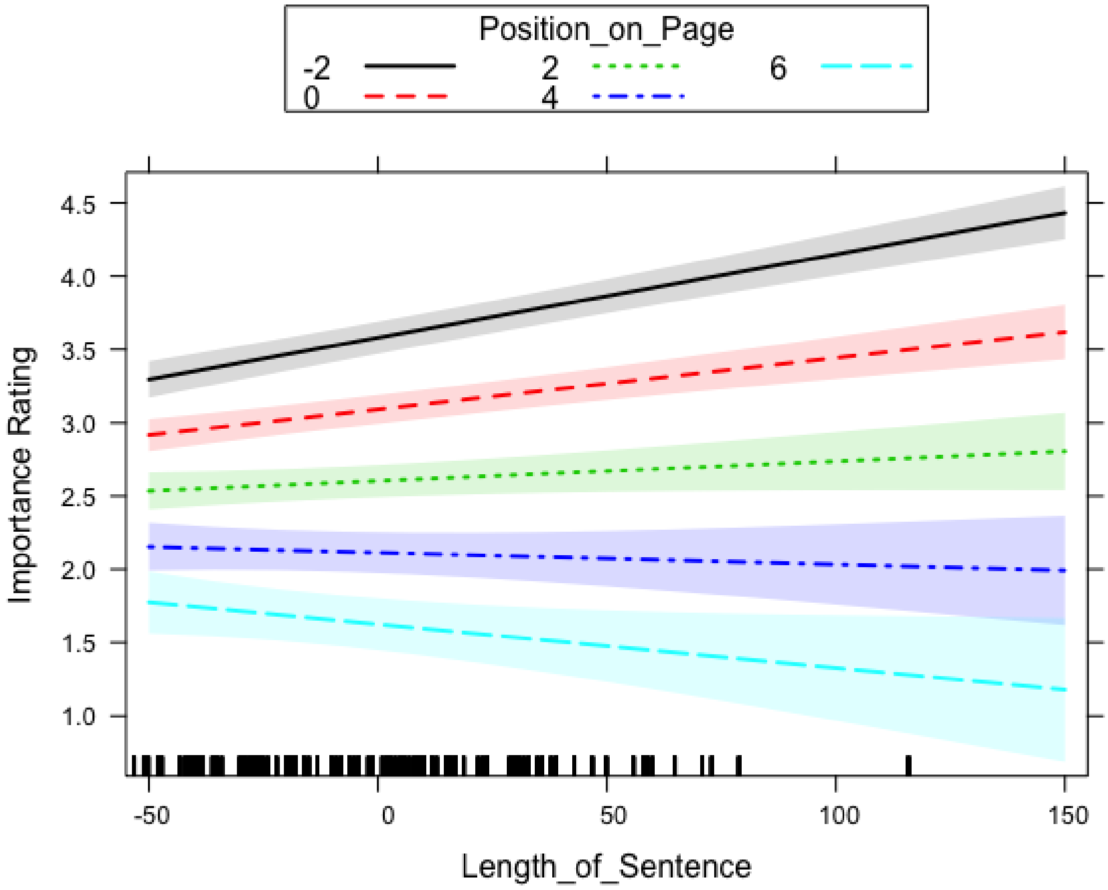


Figure 3.2 Length of Sentence x Position on Page interaction for the Importance Ratings of the unlinked sentences in Experiment 2. The lines on the graph represent the different positions a sentence can sit on the page. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

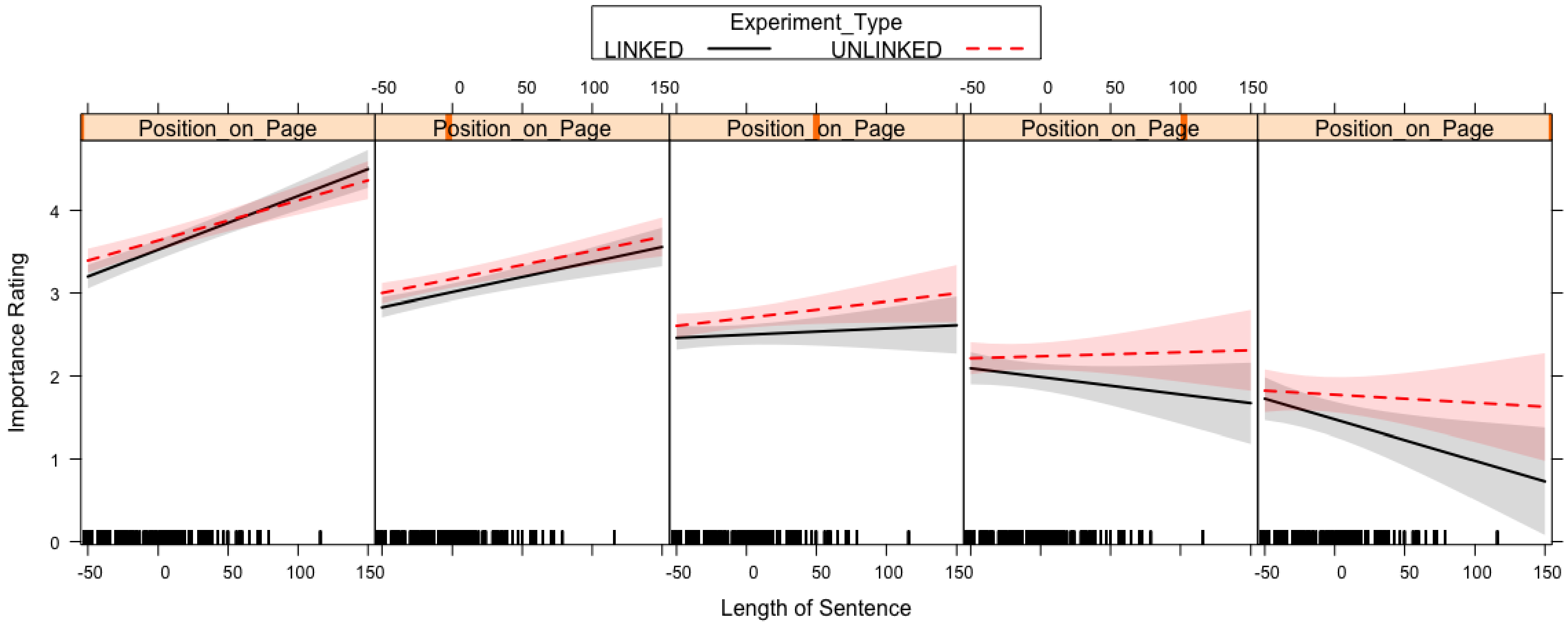


Figure 3.3 Experiment Type x Length of Sentence x Position on Page interaction for the Importance Ratings for the unlinked sentences in Experiment 2. The graphs represent different positions of the sentences on the page, increasing from left (sentences at the top of the page) to right (sentences at the bottom of the page) as indicated by the bar at the top of the graph. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

Are Sentences with Links Rated as More Important than Sentences Without Links?

To answer our second research question we ran linear mixed models using the lme4 package in R (2009) to explore the differences between the ratings when the passages contained links versus the same text when the links were removed. Note that the remainder of the reported analyses is exclusively run on the group of participants that saw paragraphs with hyperlinks present. Whether the Sentence Contained Links, Length of Sentence in characters (which was centered) and Position on the Page (which was centered) were included as fixed factors. Participants and items were included as random effects variables. A 'maximal' structure was initially specified for the random variables (Barr et al., 2013), but the model failed to converge. If a model did not converge, it was pruned by first removing the random effect correlations, then by removing the interactions between slopes and finally by successively removing the slopes for the random effects explaining the least variance until the maximal converging model was identified. For the current set of analyses only the model with intercepts both for participants and item converged. For the fixed factors successive differences contrasts were used such that the intercept corresponds to the grand mean and the fixed factor estimate for a categorical factor can be interpreted as the difference between the two conditions. The models are presented in Table 3.2.

Table 3.2

Fixed Effects Estimates, Standard Error and t value for LMM model for Experiment 2 comparing sentences with and without links.

	Estimate	Std. Error	*t* value
Intercept	3.11	0.09	35.49 *
Sentence Contains Links	0.28	0.03	8.94 *
Length of Sentence	0.00	0.00	9.86 *
Position on Page	-0.24	0.01	-40.15 *
Length of Sentence x Position on Page	0.00	0.00	-9.48 *
Sentence Contains Links x Position on Page	0.04	0.01	3.11 *

$^*t > |1.96|$

There was a main effect of whether the Sentence Contains Links with the sentence with links being rated as more important (average rating of sentence, with links: 3.28; without links: 3.10). There was also a main effect of Length of Sentence with longer sentences being rated higher and a main effect of Position on Page where sentences at the top of the page were rated higher.

These effects were qualified by a two-way interaction between Length of Sentence x Position on Page and another two-way interaction between Sentence Contains Links x Position on Page. The two-way interaction between Length of Sentence x Position on Page (Figure 3.4.) is the same interaction as when we explored only the unlinked sentences in the previous section. For sentences on top of the page long sentences are rated as more important than short sentences whereas for sentences at the end of the page long sentences are rated as less important than short sentences.

The two-way interaction between Sentence Contains Links x Position on Page can be observed in Figure 3.5. Sentences at the bottom of the page are ranked as lower in importance but to a lesser extent when there are hyperlinks present in that sentence.

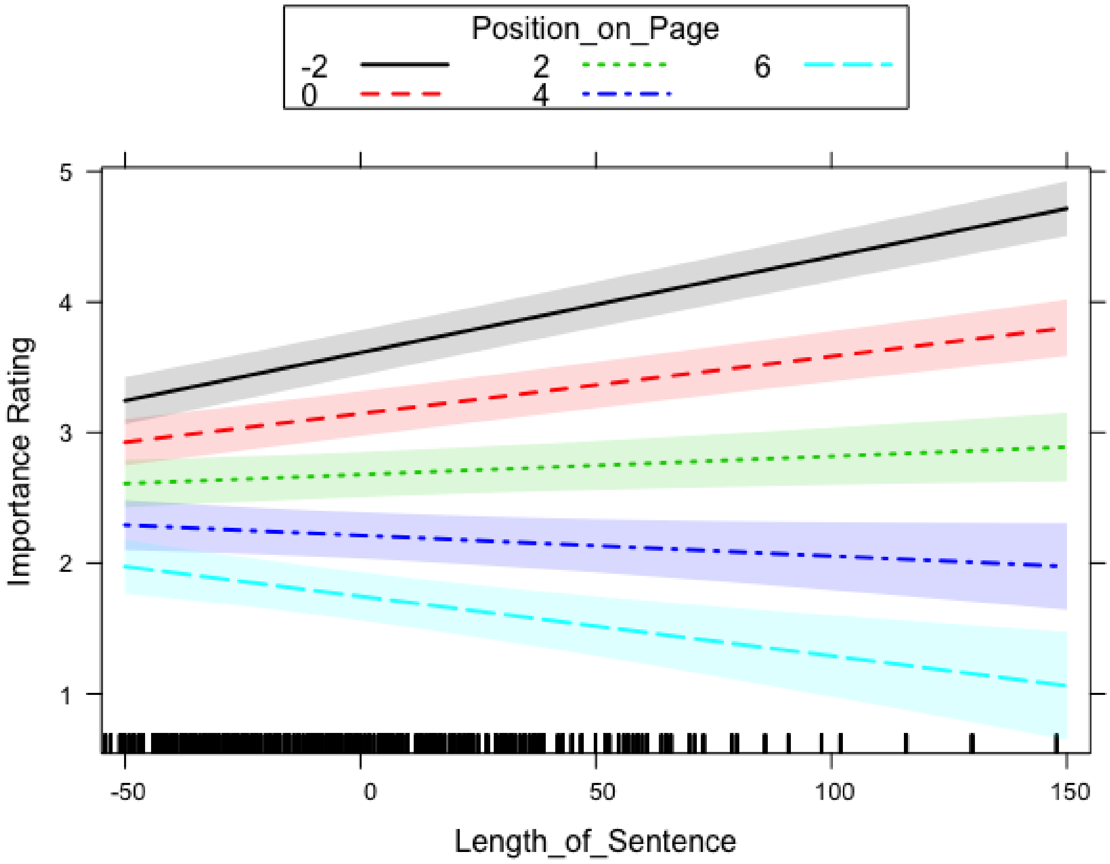


Figure 3.4 Length of Sentence x Position on Page interaction for the Linked Experiment in Experiment 2. The lines on the graph represent the different positions a sentence can sit on the page. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

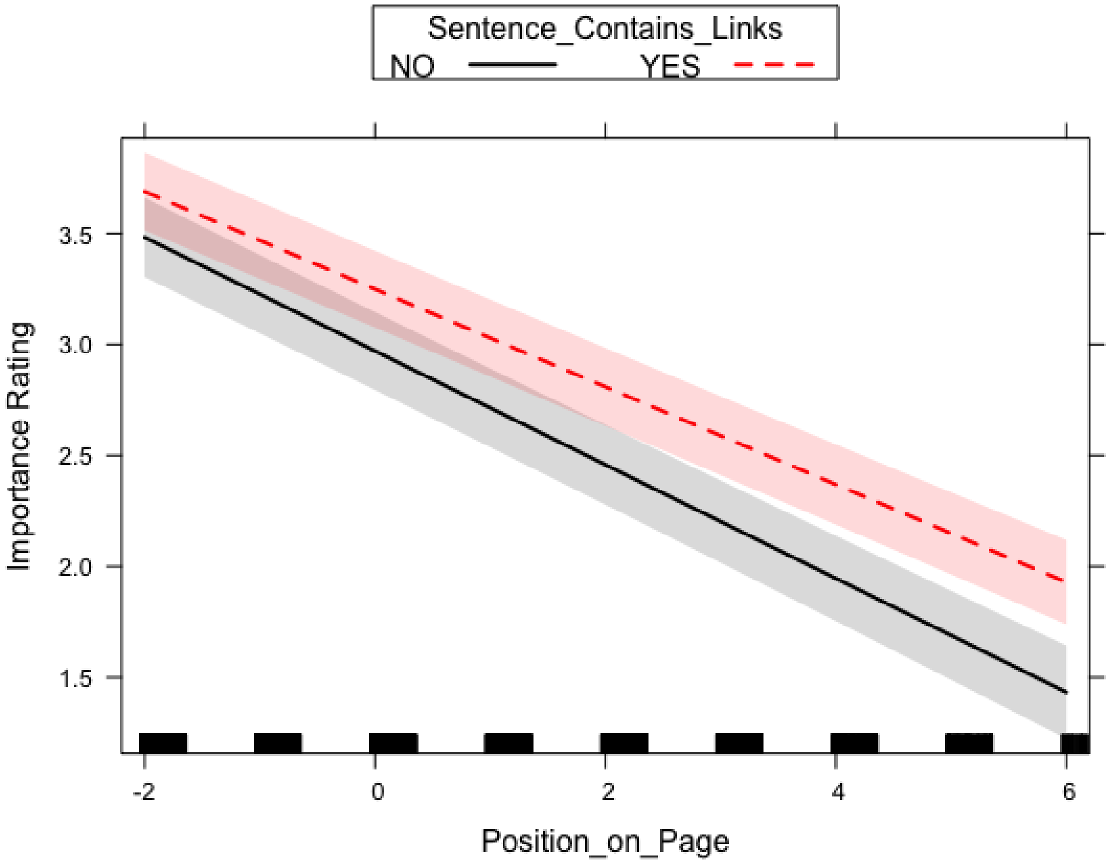


Figure 3.5 Sentence Contains Links x Position on Page interaction for the Linked Experiment in Experiment 2. The lines on the graph whether the sentence contains links (dashed line) or not (solid line). A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

Does Having More Links Increase Importance Rating of the Sentences?

We ran linear mixed models using the lme4 package in R (2009) to explore the impact of the number of links in a sentence and investigate whether once a sentence does have links, whether it will be it rated as more important the more links it has. For this analysis we did not analyse the sentences that did not feature any links. The Number of Links (a as a continuous factor which was centered), Length of Sentence in characters (which was centered) and Position on the Page (which was centered) were all included as fixed factors. Participants and items were included as random effects variables. A 'maximal' structure was initially specified for the random variables (Barr et al., 2013), but the model failed to converge. If a model did not converge, it was reduced by first removing the random effect correlations and then by removing the interactions between slopes and finally by successively removing the slopes for the andom effects explaining the least variance until the maximal converging model was identified. The intercepts were allowed to vary for the items, and both the intercepts and a slope for Number of Links and Position on Page (but not the interaction between the two) were allowed to vary.

There is a main effect of Number of Links, with more links being rated higher than fewer links. There is also a main effect of Length of Sentence where longer sentences are rated higher and an effect of Position on Page where sentences closer to the top of the page are rated higher. However, all these main effects were qualified by interactions.

There is a two-way interaction between Length of Sentence x Position on Page and a two-way interaction between Number of Links x Position on Page and both of them are qualified by a three-way interaction between Number of Links x Length of Sentence x Position on Page. The two-way interaction between Length of Sentence and Position on Page (Figure 3.6) is the same as we have seen before in the previous two analyses and indicates that at the top of the page long sentences are rated as more important than short sentences whereas for sentences at the end of the page long sentences are rates as less important compared to short sentences. The three-way interaction with Number of Links (Figure 3.8) qualifies this interaction in that this two-way interaction is not present for sentences that contain a high amount of links.

Table 3.3

Fixed Effects Estimates, Standard Error and t value for LMM model for Experiment 2 comparing the number of links in the sentences when they contain links.

	Estimate	Std. Error	*t* value
Intercept	3.17	0.09	33.93 *
Number of Links	0.04	0.01	2.68 *
Length of Sentence	0.00	0.00	4.06 *
Position on Page	-0.18	0.01	-12.41 *
Number of Links x Length of Sentence	0.00	0.00	1.79
Number of Links x Position on Page	-0.02	0.00	-4.99 *
Length of Sentence x Position on Page	0.00	0.00	-7.27 *
Number of Links x Length of Sentence x Position on Page	0.00	0.00	3.87 *

$^{*}t > |1.96|$

The two-way interaction between Length of Sentence and Position on Page (Figure 3.6) is the same as we have seen before in the previous two analyses and indicates that at the top of the page long sentences are rated as more important than short sentences whereas for sentences at the end of the page long sentences are rates as less important compared to short sentences. The three-way interaction with Number of Links (Figure 3.8) qualifies this interaction in that this two-way interaction is not present for sentences that contain a high amount of links.

Finally, the two-way interaction between Number of Links and Position on Page (Figure 3.7) indicates that at the top of the page the sentences with the most links are rates as the most important whereas at the bottom of the page the sentences with fewer links are rated as the most important. However, this is again qualified by a three-way interaction that also includes the Length of the Sentences (Figure 3.8) in that the two-way interaction only applies to short sentences but not for long sentences. For long sentences, the sentences with a high amount of links will always be rated the most important regardless of the position on the screen.

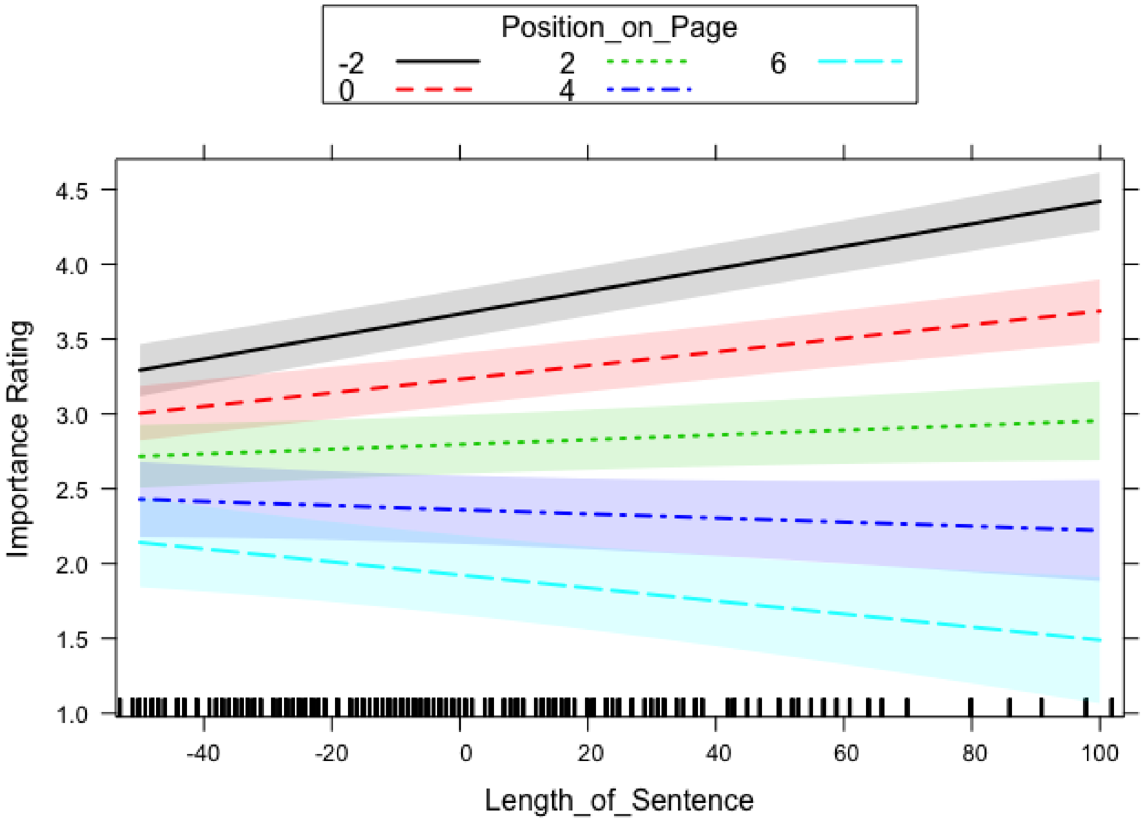


Figure 3.6 Length of Sentence x Position on Page interaction for all sentences that contain links in Experiment 2. The lines on the graph represent the different positions a sentence can sit on the page. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

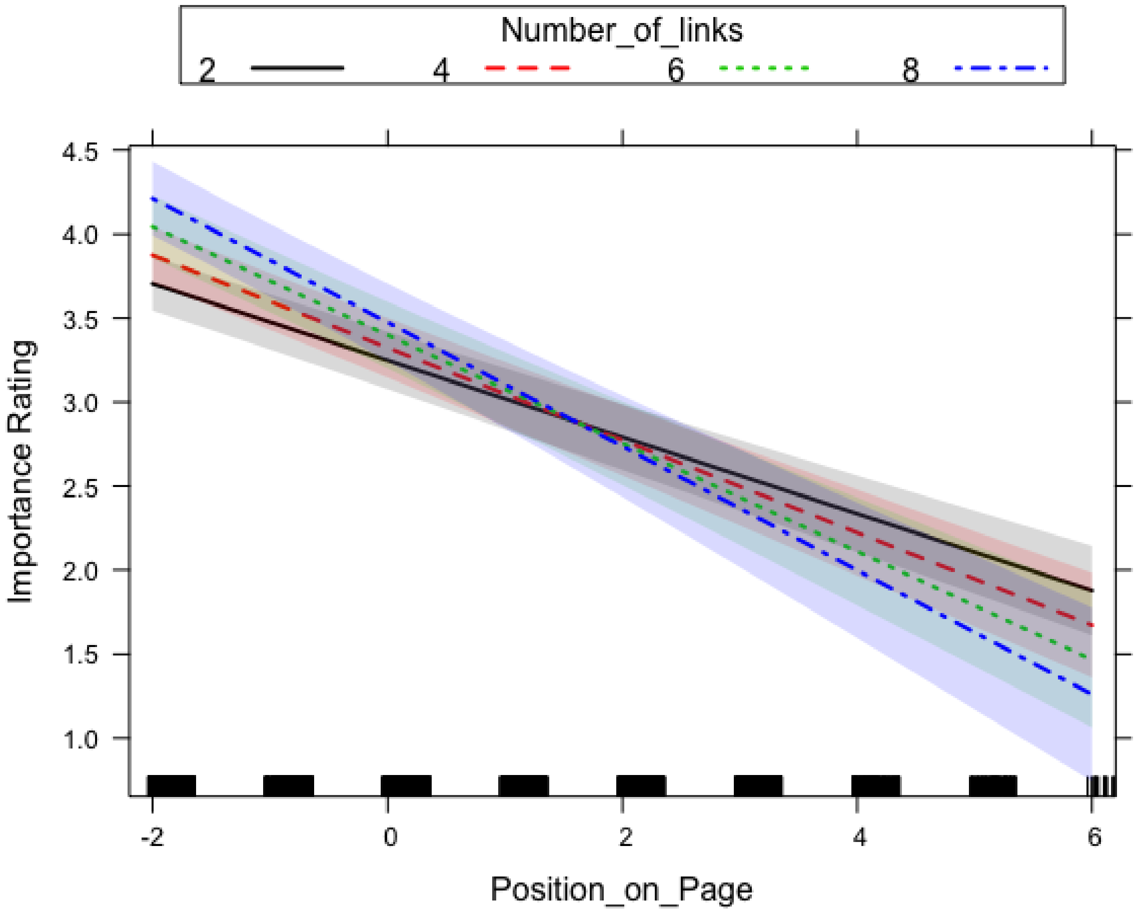


Figure 3.7 Number of Links x Position on Page interaction for all sentences that contain links in Experiment 2. The lines on the graph represent the number of links a sentence contains. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

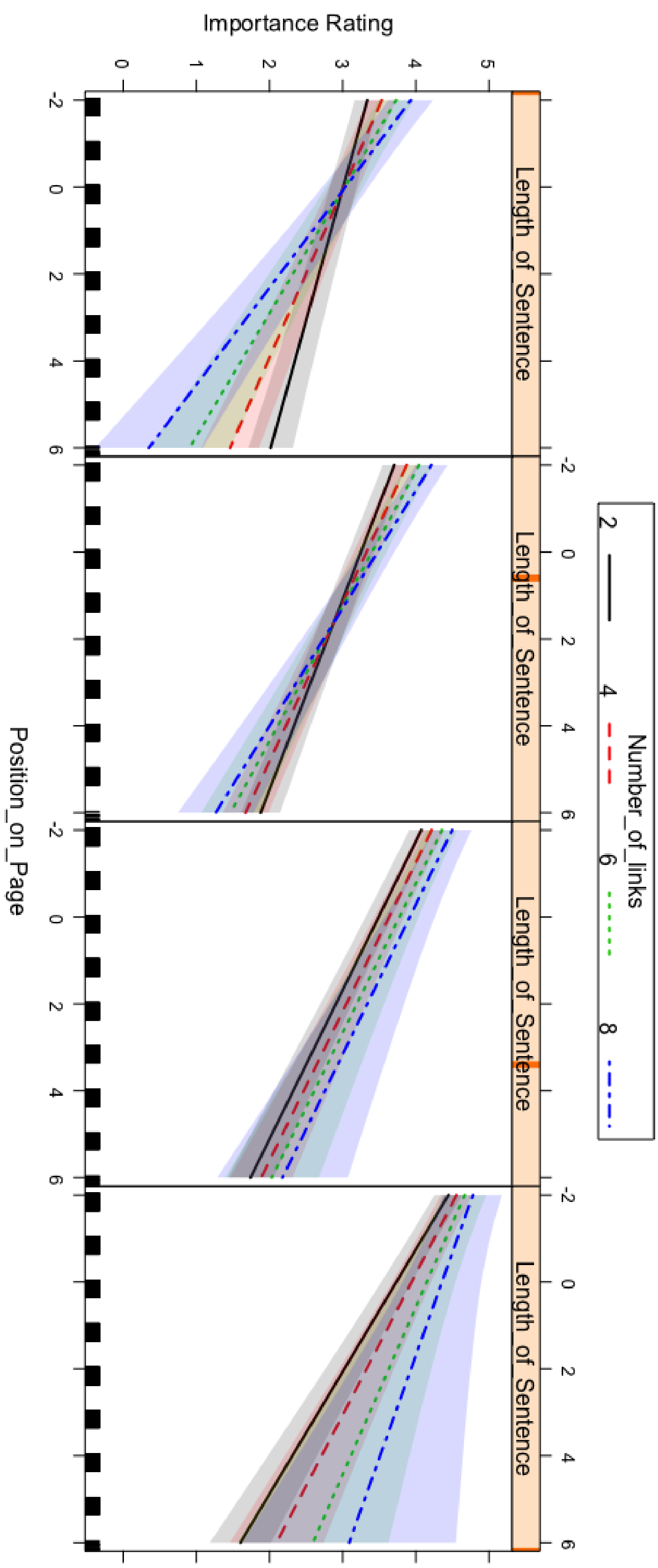


Figure 3.8 Number of Links x Length of Sentence x Position on Page interaction for all sentences that contain links in Experiment 2. The graphs represent the different sentence lengths, increasing from left (short sentences) to right (long sentences) as indicated by the bar at the top of the graph). A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

Discussion

Experiment 2 demonstrated that there are a number of factors that people use to rate the importance of a sentence. From comparing unlinked sentences in both the Linked and Unlinked Experiment we learnt a valuable lesson from comparing the importance ratings of an identical sentence and only changing the visual signals of the sentences around it. When a sentence does not contain links, but it is in a hypertext environment with other sentences that contain links it will be rated lower. The suggestion for this is that the sentences containing links are rated as more important and are therefore, by comparison, taking importance away from the sentences without links. This happens regardless of the content of that sentence as we were directly comparing sentences that were visually identical, the only difference between the unlinked sentences being the presence of hyperlinks in the surrounding sentences. This is important to consider when designing webpages. If there is text without hyperlinks that contains important information, the reader may not register this as important when it is near other sentences that do contain hyperlinks. The reader reduces the perceived importance of the sentences without links if there are other sentences with links available in the text.

When comparing sentences that have links versus those that do not have links (but in an environment that always features some links) we find that those with links are rated higher. Furthermore, this effect is more apparent towards the bottom of the page. This could be because having a link present serves as a boost to importance and when a sentence near the bottom of the page would be rated quite low, if it contains a link it gets a boost of importance. This has less of an impact at the top of the page, where sentences are rated as very important in general.

On exploring the number of links we found that short sentences with more links that were at the top of page were rated higher, but this drops off as the reader moves down the page. However, long sentences with many links were rated higher, no matter the position of the sentence on the page. This suggests that links do make a sentence more important, especially for long sentences.

One interesting finding that runs through all of these results as the fact that Position on Page seems to have a very large impact on the rating of sentences and also influences the impact of other factors. Sentence length interacted with Position on Page in all our analyses when links where present in the experiments such that at the top of the page a long sentence is rated as more important than a short sentence whereas at the end of the page a short sentence is rated as more important than a long sentence. Although this observation was qualified by the experiment that looked at the influence of the amount of hyperlinks which showed that for long sentences, those sentences with a large amount of hyperlinks will always be rated the most important regardless of the position on the screen.

Summarising the results, whereas two-way and three-way interactions sometimes have resulted in rather complex data patterns they are also often of a modest size and as such will definitely merit reinvestigation and further investigation. A clear picture does emerge across the three sets of analyses. Overall short sentences are rated more important. Sentences at the top of the page are usually rated as more important and adding hyperlinks to a sentence increases the perceived importance of that sentence.

Research has shown that important information does not receive the same loss of comprehension that is observed for the unimportant information (Duggan & Payne, 2009; Masson, 1982; Reader & Payne, 2007). Hyperlinks can be used to help signal important information and could be a very clear signal for the reader to help them find the important information. This can be very helpful to readers who do not have the time to fully read and comprehend large pieces of text they encounter. Important sentences containing one or more hyperlinks can be easily recognised due to the saliency of the coloured words and they can prove a very useful typographical cue for the readers. Signalling research has shown that signals can help highlight the important sections of text (Lorch, 1989). Also, signalling may aid the memory of important sections (Cashen & Leicht, 1970; Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch et al., 1995). This type of high saliency signal could be very important for readers engaged in skim reading. The reader can efficiently identify important information in the text and move through the text faster, ignoring unimportant information and using their time to instead focus on the text flagged as important through signals such as sentence length, position on page and if the sentence has hyperlinks.

However, we also need to consider that participants may rely on signals to communicate importance instead of making their own judgements from the semantic content of the text (Lorch et al., 2011). If a sentence has links and does not have much semantic importance, then having links in the text will become an inefficient signal of importance. If a reader picks up on this and realises that the links do not communicate importance, then the reader will probably start to ignore them as a signal of importance. This has been observed previously in research looking at over-signalling. Over-signalling can reduce the effectiveness of typographical cues, if the signalling is not useful for the task, for instance when the signalling is seemingly meaningless, the reader will ignore it (Lorch et al., 1995).

In the present experiment, the number of links is utilised by the reader to assume the importance of a sentence. Other factors such as the sentence length and position of the sentence on the page all have an impact on the importance rating of the sentence and the reader can use these factors to efficiently assume the importance of each sentence. At the top of the page the greater the number of links the higher the rating, however as the reader moved down the page, the sentence is rated higher if it has at least one link making the sentence salient. These findings can have genuine impact on how to signal important information in webpages. There is also the fact that sentences without links in hypertext may suffer a reduced importance rating because other sentences contain links in the text. This is an importance issue to consider when designing webpages.

The rating data from this chapter feeds into Chapter 4 where we explore the impact of the importance ratings of sentence on the eye movements of readers.

Chapter 4

Skim Reading: Using Adaptive Strategies for Reading on the Web

Introduction

When processing written text, we do not always choose to read carefully or for comprehension. Sometimes we skim read text to gain a general impression of the information presented. Indeed, when there is a large amount of information to read, it is not always efficient or necessary to read everything in great detail. This may often be true of reading on the Web, especially considering the large amount of information that can be present for the reader to access. Therefore, adopting a strategy of skim reading may be the most effective way to move through the information quickly. However, there is a concern that during skim reading some comprehension may be lost (Carver, 1984; Dyson & Haselgrove, 2000; Just & Carpenter, 1987; Masson, 1982).

In typical reading studies (for reviews, see Rayner, 1998, 2009), researchers want to ensure that participants read for comprehension and fully process the sentences that are presented. In order to check that participants are indeed reading for comprehension, comprehension questions are typically inserted between trials. These comprehension questions serve as a check that the participants were fully processing the sentences. If the participant has a low accuracy for answering these questions, then the experimenter knows that the participant was not fully engaged with the task. However, when reading outside of the laboratory, people may 'skim' through text and not fully process all aspects of the text that has been presented to them. Current literature suggests that reading on the Web is more likely to involve skim reading (Liu, 2005; Morkes & Nielsen, 1997). Liu (2005, pp.700) suggests that screen-based reading behaviour is characterised by 'more time spent browsing and scanning, keyword spotting, one-time reading, non-linear reading, and reading more selectively, while less time is spent on in-depth reading, and concentrated reading.'

The key differences between reading for comprehension and skim reading hypertext documents will be explored in this chapter. I will address two key questions. Firstly, when reading hypertext on the Web, what is the impact of hyperlinks on skim reading? Secondly, can hyperlinks assist with the comprehension of text while skim reading? Given the concern that skim reading has a negative effect on comprehension and the suggestion that reading on the Web involves skim reading these are key questions.

Skim Reading

To begin, we must explore the differences between reading for comprehension and skim reading. Firstly, in terms of eye movement control, there is evidence that eye movements are influenced by skim reading behaviour compared to reading for comprehension. In an early experiment, (first noted in a lab report, Just et al., 1982), Just and Carpenter (1987) explored skim reading and compared eye movements to when the participants were engaged in reading for comprehension. They found that skim readers were about two and a half times faster than normal readers. Furthermore, the eye movement analyses showed that the skim readers fixated fewer words than the normal readers and that the normal readers had longer fixations when they fixated a word. When examining the gaze durations on words, Just and Carpenter (1987) also found them to be shorter for the skim readers, who spent on average 100 ms (around one–third of the average fixation time during normal reading) less time on each fixation. However, even with this reduction in fixation times the skim readers still showed effects of frequency (low frequency words had longer fixation times compared to high frequency words, Henderson & Ferreira, 1990; Inhoff & Rayner, 1986; Rayner & Fischer, 1996). and word length (longer words had longer fixation times compared to shorter words (Brysbaert et al., 2005; Rayner & McConkie, 1976)), similar to those seen in normal readers, but the sizes of these two effects were much smaller. Although Just and Carpenter (1987) refer to reading for comprehension as normal reading, there is nothing 'abnormal' about skim reading. It is a very common reading behaviour utilised every day.

There is a great deal of evidence suggesting that during skim reading some comprehension may be lost (Carver, 1984; Dyson & Haselgrove, 2000; Just & Carpenter, 1987; Masson, 1982). However, this loss in comprehension is not consistent across all of the text being read. There appears to be a difference between information regarded as important or unimportant. The important information does not receive the same loss of comprehension that is observed for the unimportant information (Duggan & Payne, 2009; Masson, 1982; Reader & Payne, 2007). To explain these findings it has been suggested that the reader engages in an adaptive strategy in order to gain as much information from the text as possible, in a reduced time. This suggestion will now be discussed in detail in the next section.

Foraging and Satisficing Strategy

As we already mentioned, in order to explain information foraging research, Pirolli and Card (1999) use a metaphor of a bird foraging for berries in a patch of bushes, as an example of information foraging. The bird must decide how long to spend on one patch of bushes before spending time moving onto a new patch to find berries. The most efficient time to leave for a new patch is when the expected future gains from foraging in the current patch decreases to such a level that it is better to spend time moving to a new patch.

In terms of information foraging while reading, information foraging research assumes that readers are sensitive to their 'information gain' (how much useful information they are obtaining over time) and they can use this as a basis for what to read and when to stop reading. If the information gain drops below a threshold the reader is happy with they will stop reading that particular piece of text and move on to a new section of text where they might gain more information in the same amount of time. The new 'patch' could represent a new line, new paragraph or new webpage.

Reader and Payne (2007) suggest that this information foraging approach of satisficing can be applied to skim reading if we assume that the 'patches' are patches of text or paragraphs, and the reader has a threshold for their information gain influenced by the amount of time they have to read the text. For example, if the reader has a short amount of time to read the text, they will have a lower threshold for information gain. If the reader is not receiving enough information from a patch they will want to realise this quickly because of the limited amount of time. Therefore, they will want to move on to a patch that has a higher information gain as soon as possible in order to make the most efficient use of the limited time. If this is true then the readers will focus on the most important information patches and leave the patches with less important information as soon as they can in order to spend their time in the most effective and efficient way possible.

It has also been shown that a satisficing strategy can benefit memory of the text. Duggan and Payne (2009) conducted experiments to test if participants focused on the more important information (as rated by independent participants) in the text when skim reading and tested memory performance. They found that readers that were asked to skim read had better memory performance for important details from the text, but not for the unimportant details. This suggests that skim reading is an adaptive satisficing strategy. By leaving text before it is processed in depth and when information gain begins to drop, readers can efficiently move through the text at an increased speed, while trying to maintain high levels of comprehension.

As we observed in Chapter 3, hyperlinks can be used as a typographical cue to signal important information, readers of webpages would be able to use hyperlinks to focus on which sentences/patches of text contain important information, and this may also help with the adaptive strategy. With that in mind, I will now describe research on signalling.

Signalling

Lorch (1989) defines signals as "writing devices that emphasise aspects of a text's content or structure without adding to the content of the text". Lorch (1989) speaks about two different types of signalling: typographical cues and organisational cues. Typographical cues, such as boldface or underline serve to highlight a word or small section of text, while organisational cues, such as headings or numbering serve to organise the text into an easy to read structure. Research has shown that making a keyword or phrase distinct in the text results in the reader paying more attention to the emphasised content when reading (Lorch, Lorch & Klusewitz, 1995) and often results in better memory for those emphasised pieces of text (Cashen & Leicht, 1970; Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch et al, 1995).

Hyperlinks are salient words or phrases in the text and we have already demonstrated that readers treat hyperlinked words differently to non-hyperlinked text in Experiment 1C. We observed that the reader is more likely to re-read the previous text when encountering a low frequency, hyperlinked word. Perhaps this is because the hyperlinks serve as a signal to where the important information might lie in a text. If important sentences contain more hyperlinks and the hyperlinks are salient enough to be used as a signal when reading, then hyperlinks may prove a very useful typographical cue for the readers. Firstly, the signalling may help highlight the important sections of text. Secondly, the signalling may aid the memory of these important sections. However, we also need to consider that, instead of making their own judgements from the semantic content of the text, participants will rely on signals to communicate importance (Lorch et al., 2011).

We observed in Chapter 3 that sentences with hyperlinks were rated as more important than sentences without. Also, when a sentence has no links, but in text that has other sentences with links, the linkless sentences are rated lower than if they were present in the same text without any links. This suggests that the sentences without links in a hypertext environment are reduced in importance. All this leads us to believe that the presence and absence of hyperlinks has an impact on the perceived importance of sentences within a text and it may have an impact on reading behaviour. We will be able to explore this problem in Experiment 3.

Experiment 3: How Does Skim Reading Affect the Reading and Comprehension of Webpages and What is the Influence of Hyperlinks on Skim Reading?

The present experiment focuses on how hyperlinks impact on skim reading behaviour and how individuals sample the text and extract information from it. With the large amount of information online it can be safely assumed that skim reading is a common behaviour (Liu, 2005; Morkes & Nielsen, 1997). Skim reading is an efficient way of gaining as much information as possible in the shortest amount of time, while trying not to sacrifice comprehension. Hyperlinks may be used to assist in the strategy of determining which parts of the text contain important information and should be read to gain the desired amount of comprehension. An experiment was conducted to explore this issue.

Participants were either instructed to read for comprehension or asked to skim read passages of text that resemble a Wikipedia page. Target words within the passages were manipulated to either be black, or blue (resembling a hyperlink), and also their difficulty was manipulated by making the target word either a highly frequent common word (such as *plant*), or a low frequency uncommon word (such as *shrub*). Between each page of text the participant was asked comprehension questions which were either related to important or unimportant sentences in the text (as rated by independent participants not taking part in the main experiment).

The current experiment builds on the findings from Experiment 2 in the previous chapter. Experiment 2 involved participants reading edited Wikipedia articles on a computer screen and then judging each sentence based on how important it was for the general meaning of the content on that particular page. Each sentence was given an importance rating between 1 (Not very important) – 5 (Very important). The results from these importance ratings were used to understand what readers judge as "important" and explore whether the presence of hyperlinks affects the importance ratings of the sentences. The results from Experiment 2 feed into the current experiment using eye-tracking which we will refer to as Experiment 3.

Experiment 3 explores the differences between reading for comprehension and skim reading on reading behaviour by recording the participants' eye movements. The ratings from Experiment 2 were used to examine the impact of the importance that was given to each sentence. The two most important and two least important sentences from each edited Wikipedia article were selected based on the ratings obtained in Experiment 2 and comprehension questions were created based on these sentences. From the comprehension accuracy variable, we can then calculate the level of comprehension for the important and unimportant sentences.
Essentially, the intention is to replicate and extend the work of Duggan and Payne (2009). By using an eye tracker with a very high temporal and spatial resolution and a more stringent control of the stimuli set, we can merge together the research typically conducted in reading research with this previous research exploring reading on the Web.

From previous research we predicted that readers would read faster when asked to skim read, but would have reduced comprehension (Just & Carpenter, 1987). We predict that we would observe these effects by analysing both local (target word based) and global (sentence based) eye movement measures, and also comprehension accuracy for important and unimportant sentences. We predict that we will observe shorter fixation times and more word skipping in the skim reading condition, but that the linked target words will attract the attention of the reader because they are salient resulting in less skipping of linked words.

In terms of global eye movement measures we expect that sentences rated as important will have longer total reading time in comparison to the unimportant sentences as the reader will spend longer on the important sentences in the passages. From Chapter 3 we learnt that sentences with more links are rated as more important. Because links are salient in the text they could easily be used as an efficient strategy for selecting the important sections of text. We predict that when skim reading the reader may use the links as markers to where the important information lies in the page and the reader will spend more time on sentences rated as more important.

As for whether there would be a difference in the comprehension of important and unimportant information, previous research suggests that during skim reading some comprehension may be lost (Carver, 1984; Dyson & Haselgrove, 2000; Just & Carpenter, 1987; Masson, 1982). However, this loss in comprehension is not consistent, there appears to be a difference between information regarded as important or unimportant. The important information does not receive the same loss of comprehension that is observed for the unimportant information (Duggan & Payne, 2009; Masson, 1982; Reader & Payne, 2007). If skim reading is an efficient strategy to read through text the fastest way possible while minimising comprehension loss then we would expect that the skim readers will perform more poorly on comprehension question about the unimportant information. However, if there were a general reduction in comprehension across both important and unimportant sentences, then there would be a straightforward speed-accuracy trade-off for skim readers.

Methodology

Participants

Thirty-two native English speakers (2 male, 30 female) with an average age of 20.00 years participated in exchange for course credits or payment (£9). All had normal or corrected-to-normal vision and no known reading disabilities. None of the participants took part in Experiment 2.

Apparatus

Identical apparatus to Experiment 1A.

Stimuli

The stimuli in Experiment 3 consisted of the same forty edited Wikipedia articles used in Experiment 2 (see Figure 4.1). One-hundred and sixty target words were embedded in sentences (one target word per sentence) and four sentences were inserted into each Wikipedia article. The rest of the text was edited from Wikipedia articles and all words that were links in the original articles were retained for the experiment text. In total there were 8 conditions in a 2 (Task Type: Comprehension, Skimming) x 2 (Word Type: Linked, Unlinked) x 2 (Word Frequency: High, Low) within participants design. At a target word level, the target words within these articles were either displayed in blue or black to denote if the word was a hyperlink or not. There was also a word frequency manipulation where the frequency of the target word was either high or low frequency. The word frequencies were taken from the Hyperspace Analogue to Language (HAL) corpus. The frequency norms were used to extract both high and low frequency words to create the experimental stimuli. The high frequency words had an average log transformed HAL frequency of 9.94 and the low frequency words had an average log transformed HAL frequency of 5.81 (according to the norms collected in the HAL corpus, Burgess & Livesay, 1998). There was a significant difference between the high and low word frequency stimuli, $t(159)=29.66$, $p<0.001$. All target words were 4-7 characters in length with an average of 5.60 characters and the high/low frequency pairs were matched on word length. For the conditions that feature linked words, the various versions of each stimulus were controlled with a Latin square design, meaning every participant saw only one version of each edited Wikipedia article.

The display was 73 cm from the participant's eye and at this distance three characters equal about 1° of visual angle.

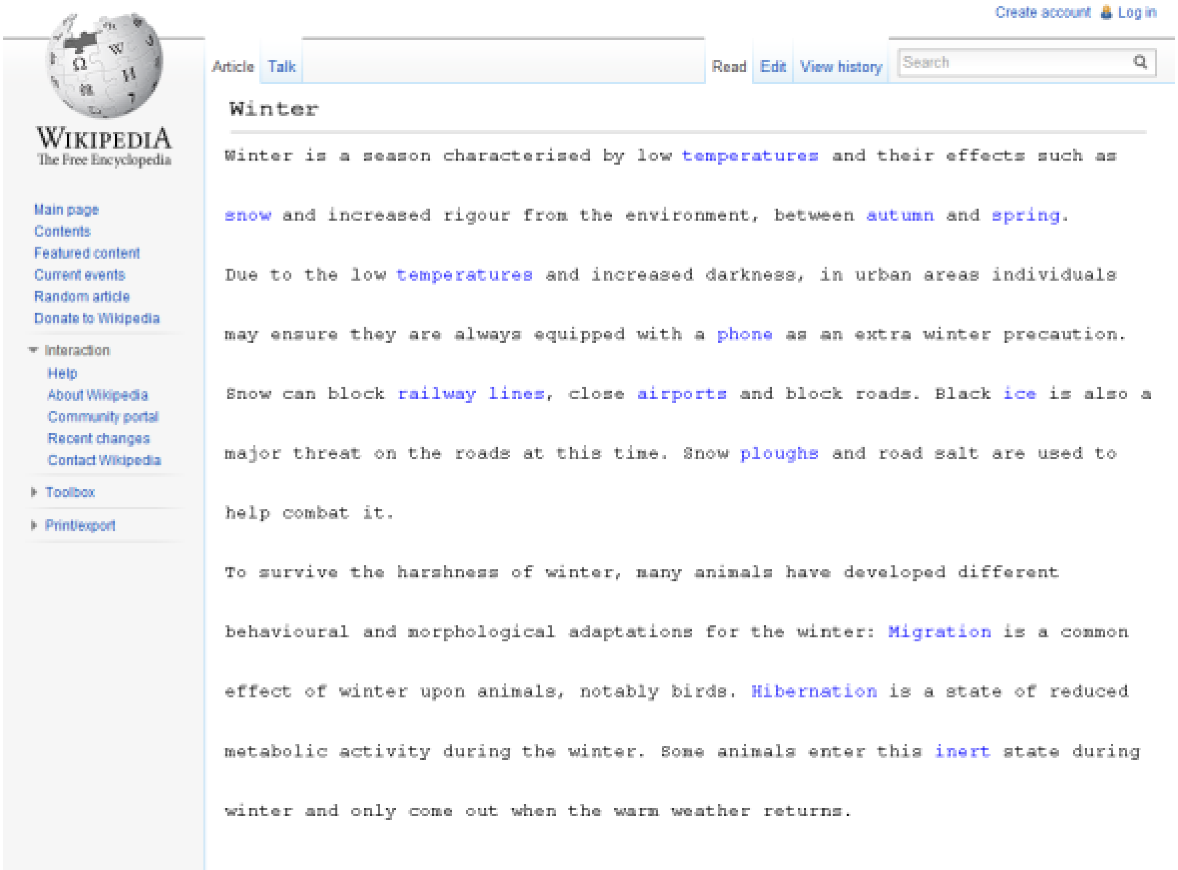


Figure 4.1 Example stimulus from Experiment 3.

Procedure

Participants were given an information sheet and a verbal description of the experimental procedure and informed that they would be reading passages on a monitor while their eyes were being tracked. The text on the screen gave the instructions to read either for comprehension or to be skim read. This was blocked such that the first twenty stimuli were to be read for comprehension and the second twenty to be skim read. We did not counterbalance the Task Type because the comprehension reading blocks might have been influenced by first having to skim read. Participants were not told they were going to be skim reading until just before that half of the experiment was due to begin, so as not to influence the first part of the experiment which was to be read for comprehension. If participants are first asked to skim read, it may become difficult to slow down and read "normally" afterwards and this would affect our data as we would not observe normal reading behaviour. The participants were informed that they were to respond to comprehension questions presented after each trial. The participants' head was stabilised in a head/chin rest to reduce head movements that could adversely affect the quality of the calibration of the eye tracker. At the beginning of each trial the participant had to look at a fixation point on the screen. When the eye tracker registered a stable fixation on the fixation point, the sentence was displayed ensuring that the first fixation fell at the beginning of the text. When participants finished reading they confirmed they had finished by pressing a button on the response box in front of them. After each trial, four comprehension questions were presented to the participants, one at a time. Two of the questions were related to sentences within the passage rated as the most important by independent participants and the other two questions were related to the sentences rated as the least important by independent participants. These ratings were gathered from the hyperlinked version of Experiment 2. In Experiment 2, each sentence was rated for its importance in each article. The top two rated sentences and the bottom two rated sentences of each article were used to create the comprehension questions. The comprehension questions were simple, and required a yes or no response. The comprehension questions were present to ensure the participants were definitely reading and comprehending the text displayed to them and also to measure the level of comprehension across both the sentences rated as important and unimportant. Participants responded to the questions by pressing the appropriate button on a response box. After the questions the next trial would appear. The experiment lasted approximately 90 minutes.

Results

Data cleaning was identical to that used in Experiments 1A, 1B and 1C (resulting in the removal of 5.43% of the total dataset). Additionally, when calculating the eye movement measures, data that were more than 2.5 standard deviations from the mean for a participant within a specific condition were removed (<1% of dataset). Data loss affected all conditions similarly.

For the local target word analyses an interest area was drawn around each target word. The interest area is the size of the target word including the space preceding it. The local analyses below are conducted using the fixations that landed on the target word, within the interest area drawn around it.

For the global sentence-based analyses, interest areas were drawn around each sentence. Each sentence had its own interest area and all fixations that landed within the sentence were used in the global analyses conducted below. The means for all of the eye movement measures for Experiment 3 are listed in Table 4.1.

How does Skim Reading affect the Way we Read Hypertext?

We focused our analysis on three key eye movement measures: Skipping probability, single fixation duration and go-past times. Skipping probability is the probability that the target word is skipped in first-pass reading. Skipping rates are used to show the ease of processing a word. If a word is easy to process, then it may be processed to a high extent prior to fixating it and therefore skipped completely in first-pass reading. Single fixation durations are when the reader makes only one single fixation on the target word in first-pass reading. It is used as a measure to describe how easy a word is to process (Rayner, 1998). Because this measure only includes times where the target word was only fixated once it is one of the cleanest measures to use to represent how difficult a word is to process.

Also, when the target word was fixated, in 93.91% of the cases it received a single fixation. Therefore, we limited the fixation duration analyses to when there was a single fixation on the target word. Go-past times are the accumulated time from when a reader fixates the target word until the reader passes to the right after the target word. This measure is often used to explore if a reader has had trouble integrating the target word because it includes the regressive (backward-directed) fixations when a reader has to reread preceding content.

We ran LMMs using the lme4 package in R (2009) to explore the impact of three variables. The three independent variables were included as fixed factors: Task Type (Comprehension, Skimming), Word Type (Linked, Unlinked) and Word Frequency (High, Low). Participants and items were included as random effects variables. A maximal random model was initially specified for the random factors (Barr, Levy, Scheepers, & Tily, 2013). If a model did not converge, it was reduced by first removing the correlations in the random structure, then removing interactions and finally by successively removing the slopes for random effects explaining the least variance until the maximal converging model was identified.

For skipping probability and go-past times the intercept for the items variable was allowed to vary and the participant variable included both the intercept and the slope for the effect of Word Frequency. The interaction between Word Frequency and Word Type was removed from the model because it did not contribute significantly to the fit of the data. For single fixation durations the intercept for the items variable was allowed to vary and the participant variable included both the intercept and the slope obtained for the effects of Word Frequency and Word Type and their interaction.

All the patterns observed in the models were identical whether they were run on log-transformed or untransformed fixation durations, allowing us to present the data run on the untransformed fixation durations in order to increase transparency. The only exception is for go-past times measures where the logged model was used to represent the data. This was due to the data needing to be normalised because it was skewed and resulted in qualitatively different models for logged versus unlogged go-past times. All fixed effects estimates are shown in Table 4.2 and were calculated using successive differences contrasts so that all main effects were compared to a grand mean rather than a baseline intercept.

Word Skipping.

There was a main effect of Word Frequency in skipping probability. The high frequency words were skipped significantly more often than the low frequency words and also had significantly shorter fixation times when they were fixated. This replicates previous experiments that have demonstrated that low frequency words are skipped less often because they are more difficult to process than highly frequent words (Inhoff & Rayner, 1986). There was also a consistent main effect of Task Type where there is more skipping when the reader is skim reading compared to when they are reading for comprehension. This replicates the research conducted by Just and Carpenter (1987) who found similar results in skim reading.

In skipping probability, as well as the significant main effects of Word Frequency and Task Type, there was also a main effect of Word Type. When the target word is unlinked it is more likely to be skipped compared to when it is linked. This effect of Word Type is qualified by an interaction with Task Type (see Figure 4.2). Subsequent contrasts showed that there was no difference in skipping probability

Table 4.1

Means of Eye Movement Measures for Experiment 3. Standard deviation in parentheses.

Task Type	Word Type / Word Frequency	Skipping Probability (%)	Single Fixation Duration (ms)	Go-Past Time (ms)
Comprehension	Linked/High	52 (20)	221 (44)	378 (223)
	Linked/Low	48 (22)	233 (37)	370 (164)
	Unlinked/High	54 (19)	212 (37)	322 (116)
	Unlinked/Low	51 (20)	246 (45)	375 (140)
Skimming	Linked/High	52 (23)	201 (27)	295 (128)
	Linked/Low	48 (22)	221 (35)	284 (76)
	Unlinked/High	68 (18)	204 (41)	263 (94)
	Unlinked/Low	63 (18)	205 (31)	250 (50)

Table 4.2

Fixed Effect Estimates for Skipping Probability Percentage of the Target Word and the Fixation Times on the Target Word in ms for Experiment 3.

	Skipping Probability			Single Fixation Duration (ms)			Go Past Time (ms)		
	Estimate	Std. error	*z* value	Estimate	Std. error	*t* value	Estimate	Std. error	*t* value
(Intercept)	0.21	0.13	1.66	219.77	3.91	56.25 *	5.60	0.03	164.89 *
Word frequency	-0.19	0.06	-3.05 *	16.03	3.15	5.09 *	0.06	0.03	2.02 *
Word type	0.43	0.06	7.07 *	-1.47	4.30	-0.34	-0.05	0.03	-1.70
Task type	0.30	0.06	4.94 *	-16.62	3.22	-5.16 *	-0.19	0.02	-9.62 *
Word type x Task type	0.57	0.12	4.74 *	5.62	7.10	0.79	-0.05	0.04	-1.28
Word frequency x Task type				-17.66	6.27	-2.81 *	-0.12	0.04	-3.07 *
Word frequency x Word type				14.64	6.35	2.31 *			
Word frequency x Word type x Task type				33.19	12.61	2.63 *			

*$z > |1.96|$

*$t > |1.96|$

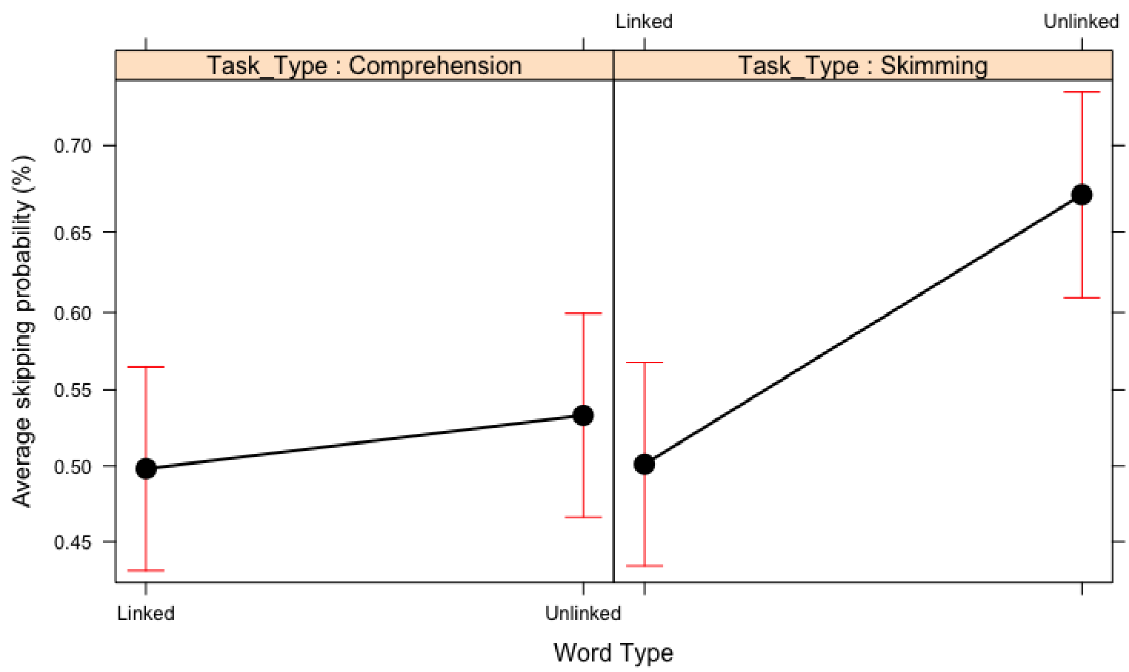


Figure 4.2 Two-way interaction between Word Frequency and Task Type in Experiment 3. Means and standard error bars for skipping probability.

when the target word is linked or unlinked during comprehension reading (z=1.46, SE=0.09, p=0.15), but there was a significant difference in the skim reading condition. Linked words are significantly less likely to be skipped compared to unlinked words in the skim reading condition (z=7.54, SE=0.09, p<.001). This suggests that when the readers are skim reading they are attempting to fixate the linked words. They may be using them as anchor points throughout the passage as the reader thinks the linked words may be the most important words within the passage.

Fixation duration measures.

There was a main effect of Word Frequency in single fixation duration. The low frequency words had significantly longer fixation durations. This replicates previous research where low frequency words are fixated for longer because they are more difficult to process than highly frequent words (Inhoff & Rayner, 1986). There was also a main effect of Task Type in single fixation durations where there are shorter fixation durations when the participant is skim reading. This replicates the research conducted by Just and Carpenter (1987) who found similar results in skim reading.

For single fixation duration there are multiple two-way interactions, and these are qualified by a three-way interaction between Word Frequency, Word Type and Task Type (see Figure 4.3). To explore this interaction, additional contrasts were conducted. Firstly, fixation times were significantly shorter when the passages were skim read compared to when they were read for comprehension (single fixation duration: Estimate=-32.15, SE=6.45, t=-4.99), which supports previous research that shows shorter fixation times when readers are skim reading (Just & Carpenter, 1987). The three-way interaction was caused by the fact that, when participants skim read the passages, a frequency effect only emerged for linked target words (single fixation duration: Estimate=8.71, SE=3.77, t=3.03), and not for unlinked target words (single fixation duration: Estimate=-2.12, SE=3.37, t=-0.63) but when participants read for comprehension, there was a significant frequency effect both for linked target words (single fixation duration: Estimate=9.48, SE=3.23, t=2.94) and unlinked target words (single fixation duration: Estimate=14.80, SE=3.32, t=4.46). This is a particularly interesting result because there is an absence of a frequency effect for the unlinked word in the skim reading condition. This suggests that the readers are not fully lexically processing the unlinked target words they are landing on. The presence of a frequency effect in the linked words in the skim reading condition suggests they are focusing on the linked words and lexically processing them. If, as previous researchers have suggested, readers often skim read when reading on the Web (Liu, 2005; Morkes & Nielsen, 1997), the words we choose to link become very important. Readers might be processing the link at a deeper level, especially if links are used as anchor points to where the important information within the text might lie.

When reading for comprehension, the target word being linked or unlinked had no influence on fixation times when the target word was high frequency (single fixation duration: Estimate=1.27, SE=3.24, t=0.39), and had only a marginal effect in single fixation durations for low frequent words (single fixation duration: Estimate=6.39, SE=3.35, t=1.91). There was also no effect of Word Type in the skim reading condition when the target word was high frequent (single fixation duration: Estimate=4.21, SE=3.36, t=1.29). However, there were significantly longer fixations on the linked target words in the skim reading condition when the target word was low frequent compared to high frequent for single fixation duration and total time (single fixation duration: Estimate=-10.97, SE=3.51, t=-3.13). This is due to the frequency effect being present in the skimming condition for the linked words, but not present for the unlinked words. We see longer fixation times for the low frequent linked words because they are more difficult to process. We do not observe it for the unlinked, low frequency words in the skimming condition because the readers simply were not processing them to the same level, instead focusing on the linked words.

Although the majority of fixations on the target word were single fixations, when the target word was fixated 14.11% of target words had regressions to previous interest areas. For consistency with previous chapters, we will also explore go-past times, however we need to point out that the analysis will not have a high amount of statistical power for examining re-reading as re-reading in itself was obviously rather rare. Go-past times take into account the accumulated time from when the reader first fixated the target word until when they move further to the right, after the target word. All the times where the reader fixated the target word and then made a fixation backward to the preceding text are included in this measure. In go-past times there is a two-way interaction between Word Frequency and Task Type (see Figure 4.4) where the frequency effect is present when reading for comprehension (Estimate=0.13, SE=0.03, t=4.60), but is missing in skim reading (Estimate=-0.01, SE=0.03, t=-0.46). This suggests that the readers were not fully processing the target words when skim reading and not lexically identifying them, which is why there is no frequency effect for skim reading in this case. The lack of a frequency effect has been previously observed in previous research. Rayner and Raney (1996) found that when searching for a target word in a passage of text, the frequency effect seen in comprehension reading, where fixation times are shorter on high frequency words compared to low frequency words (Henderson & Ferreira, 1990; Inhoff & Rayner, 1986; Rayner & Fischer, 1996), was not observed. To engage in a visual search of text you do not need to lexically identify the words as the reader is not reading for comprehension. A similar effect could be seen in skim reading if the reading is simply scanning the text for important information rather than fully engaging and reading the text.

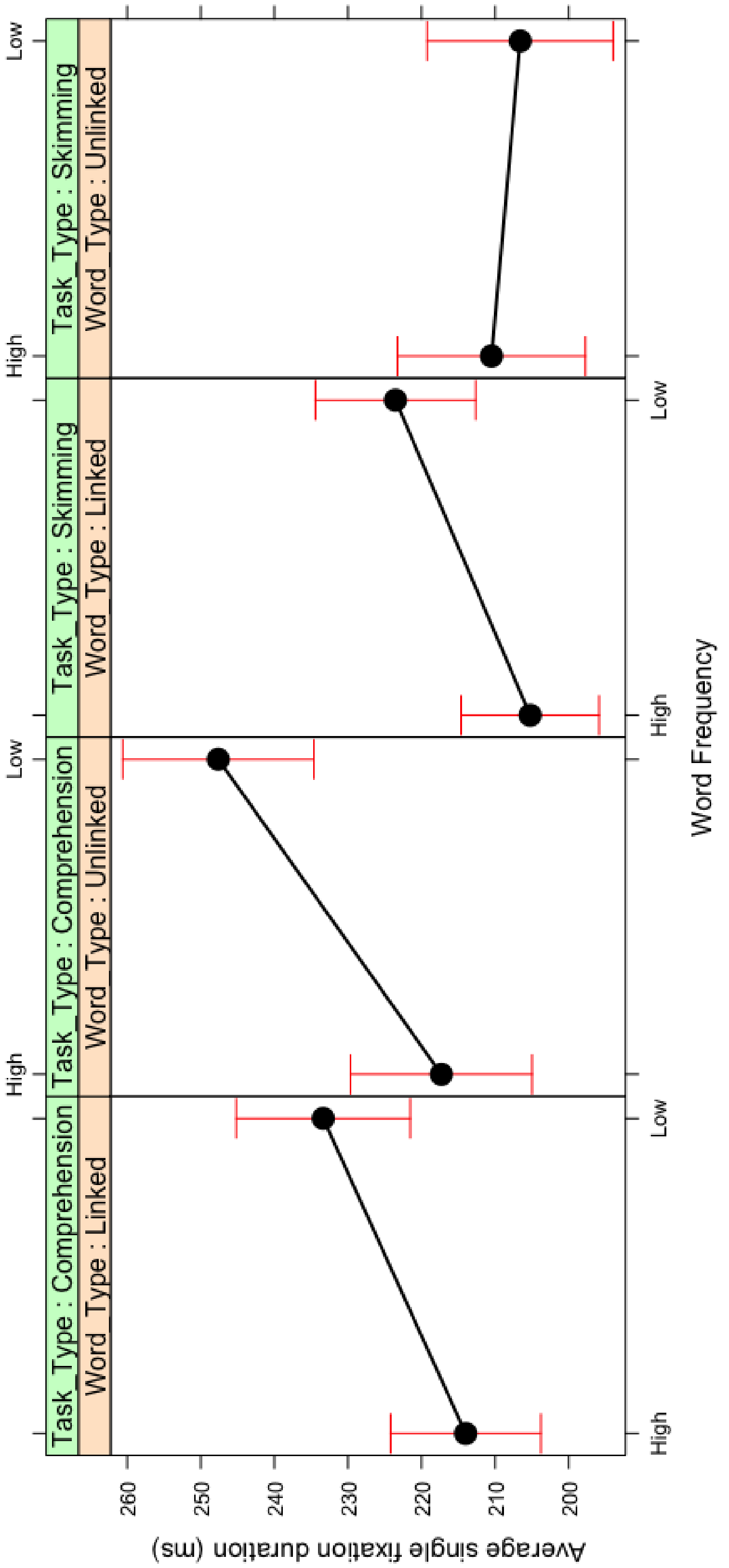


Figure 4.3 Three-way interaction between Word Frequency, Word Type and Task Type in Experiment 3. Means and standard error bars for single fixation durations.

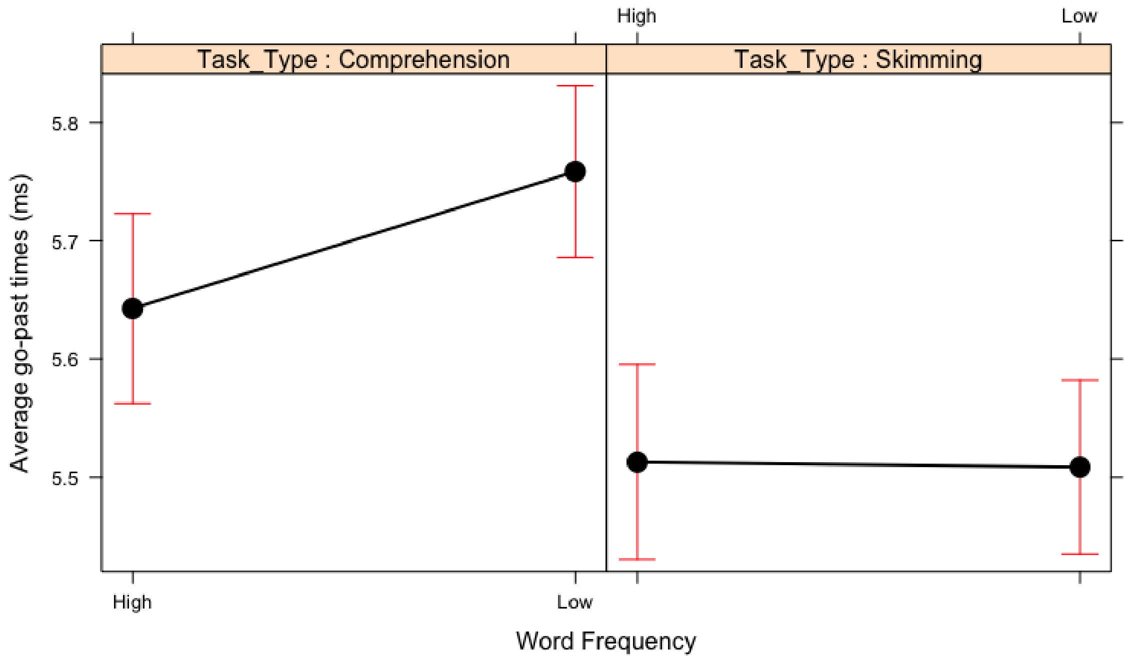


Figure 4.4 Two-way interaction between Word Frequency and Task Type in Experiment 3. Means and standard error bars for go-past times.

How does Skim Reading affect Reading at a Global Level?

Exploring skim reading at a local level can help us understand how the variables we manipulated effected reading, but as a task we need to explore skim reading also at a more global level to understand how the reader is adapting their strategy to efficiently read the information on the page in a reduced time. Therefore, we conducted analyses exploring how the reader could use an adaptive strategy when skim reading and make use of important information only to read the articles in the most efficient way possible. We found in Chapter 3 that a lot of factors are taking into consideration when making importance ratings such as sentence length and position on page and the number of hyperlinks in the sentence. Therefore, in the analyses for the sentence-based measures in this section, we will use the importance rating to enrich the analysis.

Several eye-movement measures were calculated based on each sentence. Sentence gaze duration includes all fixations in the first pass reading of the sentence. 23.72% of fixations involved rereading so we also calculated the total sentence reading time which consisted of all fixations on the sentence including all rereading.

We also calculated wrap up measures for each sentence to measure the integration of the text (Just & Carpenter, 1980; Warren, White, & Reichle, 2009). These are measures that are calculated based on the last word of the sentences (which was fixated 70.44% of the time). An interest area is drawn around the last word of the sentence and the measures of skipping probability, gaze duration, go -past times and total time spent on that word are created and explored. Total time includes all fixations on the final word in the sentence, including all rereading fixations on that target (19.78% of the time the last word was fixated more than once).

We ran LMMs using the lme4 package in R (2009) to explore the impact of two independent variables were included as fixed factors: Task Type (Comprehension, Skimming) and Importance Rating. Importance rating (see Chapter 3) is a score based on a 5-point Likert scale, where independent participants rated each sentences importance between 1 (unimportant) and 5 (important). The importance rating is a continuous variable and was centered. An interaction was included between the fixed factors and for most of the models unless model comparisons proved that the model was a better fix without the interaction term. Participants and items were included as random effects variables. A maximal random model was initially specified for the random factors (Barr, Levy, Scheepers, & Tily, 2013). If a model did not converge, it was reduced by first removing the random effect correlations and then removing interactions and finally, by successively removing the slopes for the random effects explaining the least variance until the maximal converging model was identified. The intercept for the items variable was allowed to vary and for most of the measures the participant variable included both the intercept and the slope obtained for the effects of Task Type and Importance Rating and their interaction. The only exceptions were skipping probability where the only intercept for the items variable was allowed to vary and the participant variable included the intercept and the slope for the effect of Task Type. For skipping probability, the interaction between Task Type and Importance Rating also had to be removed to allow the models to converge. All means and standard deviations are showed in Table 4.3 and all fixed effects estimates are shown in Table 4.4.

For sentence gaze duration we found a main effect of Task Type, where there were longer first pass gaze durations on the sentence when it was read for comprehension compared to skim reading. We also found a main effect of the Importance Ratings, where the higher the rating, the longer the gaze durations. This means that the readers did spend longer on sentences rated as more important than those rated as less important.

We found similar effects for total sentence reading time. We found a main effect of Task Type with longer durations on sentences read for comprehension compared to skim reading. We also found a main effect of Importance Rating with the higher the rating for the sentence the longer the time spent on it. This suggests that Task Type and Importance play a role in both first-pass reading and rereading, but there was no interaction between these two factors so the Importance Rating plays a similar role when reading for comprehension and skim reading.

Wrap up skipping had a main effect of Task Type and Importance Rating. Participants skipped the final word of a sentence more when skim reading compared to comprehension reading and they also skipped more when the sentences were rated as lower in importance. This suggests that the readers are trying to efficiently selectively process the important and unimportant information, as they are showing increased skipping of the wrap up region of unimportant sentences compared to the important sentences.

Wrap up gaze duration showed a main effect of Task Type and Importance Rating and, furthermore, there was also had an interaction between Task Type and Importance Rating (see Figure 4.5). In Figure 4.5, we can clearly see that higher ratings result in longer gaze durations for first-pass gaze durations. It is also apparent from Figure 4.5, where there is a steeper slope observed for reading for comprehension compared to the flatter slope for skim reading. This shows the interaction between Task Type and Importance Rating is because in skim reading the importance rating does not seem to have such a large impact as it does when reading for comprehension. We observe the same results for wrap up total reading time: A main effect of Task Type and Importance Rating and also had an interaction between Task Type and Importance Rating (see Figure 4.6). Again we see a steeper slope for reading for comprehension compared to the flatter slope for skim reading, suggesting that the importance of the sentence has a larger impact when we read for comprehension. This could be due to the fact that in skim reading there is a higher level of word skipping and, therefore, the wrap up regions are more likely to be skipped. Wrap up effect reflect higher level processing at the sentence level and when skim reading, this may be happening at a less detailed level.

Finally, wrap up go-past times had a main effect of Task Type (longer times in Comprehension compared to Skimming), but no effect of the Importance Rating. This is probably due to there being a reduction in rereading in skim reading, therefore in the go-past time measures, in skim reading, they will be revisited very rarely. Interestingly there is no effect of importance in wrap up go-past times. This suggests that the reader takes all the information from the important sentences on first pass through the sentence and they do not tend to revisit it later.

Table 4.3

Means and Standard Deviations for Global Eye Movement Measures in Experiment 3.

Task Type	Sentence Gaze Duration (ms)	Total Sentence Reading Time (ms)	Wrap Up Skipping Probability (%)	Wrap Up Gaze Duration (ms)	Wrap Up Total Reading Time (ms)	Wrap Up Go Past Time (ms)
Comprehension	2631 (1886)	3558 (1914)	53.93 (49.85)	246 (113)	290 (152)	435 (384)
Skimming	1493 (993)	1825 (983)	62.01 (48.54)	207 (70)	221 (86)	277 (182)

Table 4.4

Fixed Effect Estimates for Global Eye Movement Measures in Experiment 3.

	Sentence Gaze Duration (ms)			Total Sentence Reading Time (ms)		
	Estimate	Std error	*t* value	Estimate	Std error	*t* value
(Intercept)	6.92	0.11	63.05 *	7.12	0.08	86.36 *
Task Type	-0.46	0.12	-3.84 *	-0.63	0.06	-11.05 *
Important Rating	0.10	0.03	3.36 *	0.18	0.02	8.37 *
Task Type * Importance Rating	0.00	0.03	-0.02	-0.02	0.01	-1.32
	Wrap Up Skipping Probability			Wrap Up Gaze Duration (ms)		
	Estimate	Std error	*z* value	Estimate	Std error	*t* value
(Intercept)	0.83	0.20	4.19 *	5.33	0.02	254.13 *
Task Type	0.36	0.17	2.05 *	-0.15	0.02	-8.53 *
Important Rating	-0.13	0.04	-2.92 *	0.03	0.01	3.00 *
Task Type * Importance Rating				-0.03	0.01	-1.99 *
	Wrap Up Total Reading Time (ms)			Wrap Up Go Past Time (ms)		
	Estimate	Std error	*t* value	Estimate	Std error	*t* value
(Intercept)	5.43	0.02	223.15 *	5.65	0.03	174.09 *
Task Type	-0.25	0.02	-11.25 *	-0.37	0.03	-11.98 *
Important Rating	0.05	0.01	3.82 *	-0.02	0.02	-0.88
Task Type * Importance Rating	-0.06	0.01	-3.74 *	-0.02	0.02	-0.98

*$z > |1.96|$

*$t > |1.96|$

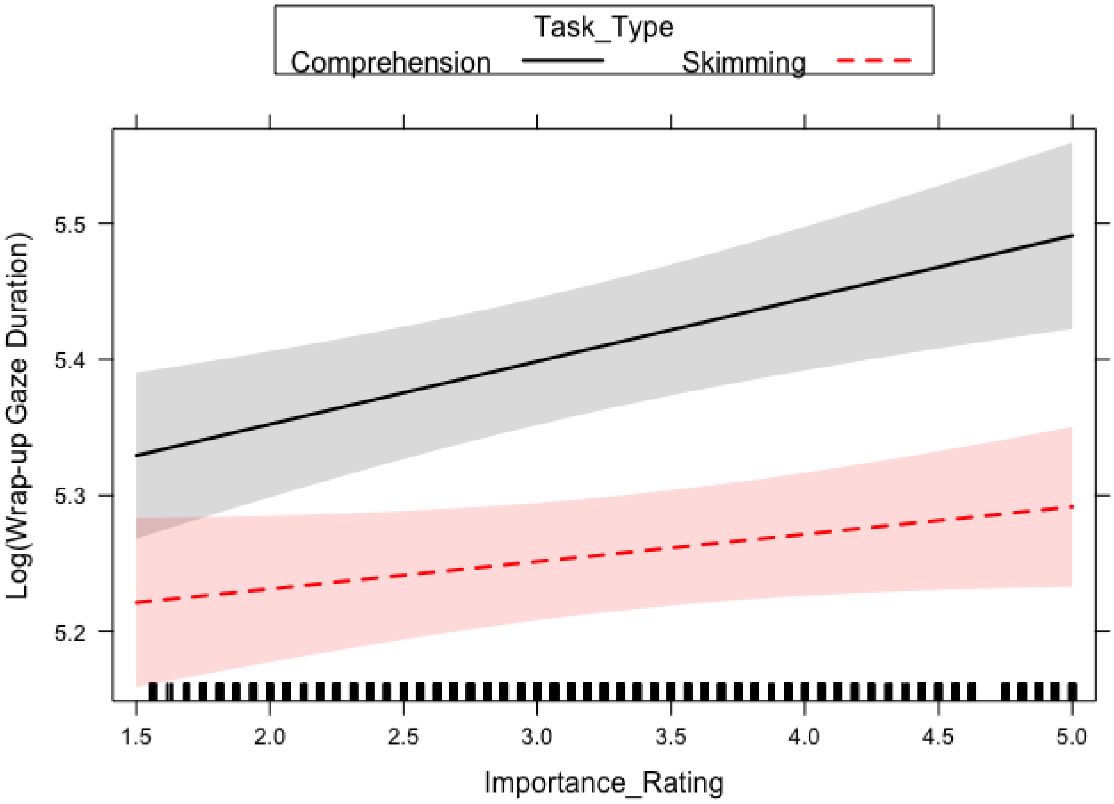


Figure 4.5 Task Type x Importance interactions for Wrap Up Gaze Duration in Experiment 3. The lines on the graph represent the different tasks. A 95-percent confidence interval (the shaded region) is drawn around the estimated effect.

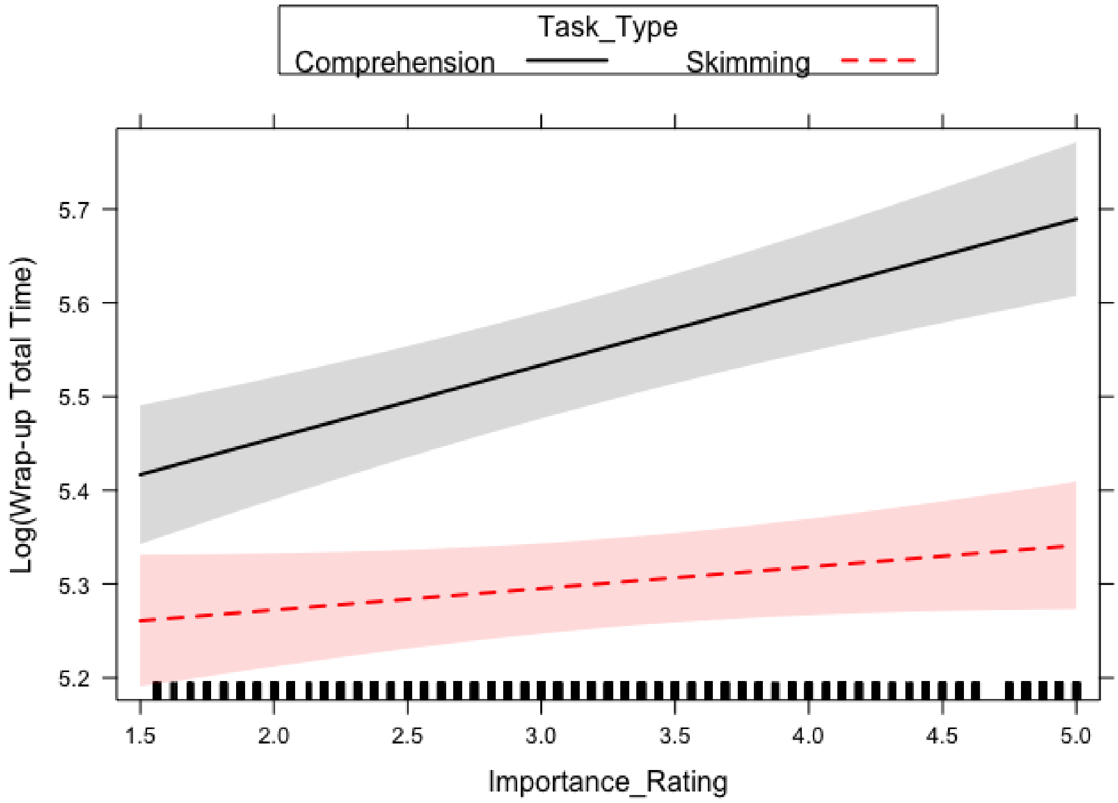


Figure 4.6 Task Type x Importance interaction for Wrap Up Total Time in Experiment 3. The lines on the graph represent the different tasks. A 95-percent confidence interval (the shaded region) is drawn around the estimated effect.

How does Skim Reading affect Comprehension?

In Experiment 2 each edited Wikipedia article had every sentence within it rated for its general importance to the meaning of the whole passage by sixteen independent participants who did not take part in the eye tracking study. The two most important and two least important sentences were selected and had comprehension questions created about them.

We ran LMMs using the lme4 package in R (2009) to explore the impact of two independent variables which were included as fixed factors: Importance (High Importance, Low Importance) and Task Type (Comprehension, Skimming) (see Table 4.5 for means and Table 4.6 for fixed effect estimates). Participants and items were included as random effects variables. A 'maximal' structure was initially specified for the random variables (Barr et al., 2013), but the model failed to converge. If a model did not converge, it was reduced by first removing the random effect correlations and then by removing the interactions between slopes and finally by successively removing the random slopes explaining the least variance until the maximal converging model was identified. In the reported analyses the intercept was allowed to vary for items, and for participants. Model comparisons were run and the interaction between Importance and Task Type was removed removed from the model because it did not contribute significantly to the fit of the data. For the fixed factors successive differences contrasts were used such that the intercept corresponds to the grand mean and the fixed factor estimate for a categorical factor can be interpreted as the difference between the two conditions.

There was a main effect of Task Type where accuracy was significantly lower when the text was being skim read than when it was read for comprehension. This replicates previous research suggesting that comprehension is impaired when skim reading (Carr, 2010; Dyson & Haselgrove, 2000; Just & Carpenter, 1987; Masson, 1982).

Additionally, we used the same methodology for analysing the comprehension question results as Duggan and Payne (2009). They used Signal Detection Theory (SDT) measures to explore participants' comprehension of the text. SDT has been used in a wide range of visual cognition tasks (Macmillan & Creelman, 2005). SDT measures take into account the proportion of trials where participants respond correctly (termed 'hits') as well as the proportion of trials where participants erroneously respond 'yes' (termed 'false alarms'). Duggan and Payne (2009) used the SDT measure *d'* to examine overall response accuracy, with higher values indicating better overall response accuracy. They also used the SDT measure *c* to measure the response criterion or bias. Higher values of the response criterion indicate a tendency to respond 'no', suggesting that participants are 'biased' towards more conservative responses (i.e., they only respond 'yes' when there is a strong reason to do so). Lower values of the criterion indicate a tendency to respond 'yes', suggesting that participants are biased towards more liberal responses (i.e., they are willing to respond 'yes', even when there is only a weak reason to do so). We used these measures and examined them using a 2 (Importance: High Importance, Low Importance) x 2 (Task Type: Comprehension, Skimming) within subjects ANOVA (see Table 4.5 for means). For *d'* there was a main effect of Task Type ($F(1,31)=10.38$, $p<.001$). The participants' comprehension of the text decreased when they were skim reading. There was also a marginal main effect of Importance ($F(1,31)=3.97$, $p=0.06$), which suggests that the participants were to a degree engaged in an adaptive satisficing strategy because they had improved accuracy for comprehension questions relating to the most important information. There was no significant interaction between Importance and Task Type ($F(1,31)=0.31$, $p=0.58$). When examining the bias (*c*) there were no significant differences between the measures (all *F*s were smaller than 2.9, all *p*s were larger than .1). This shows that there was no bias when responding to the comprehension questions, i.e. participants were no more likely to respond yes or no.

Table 4.5

Behavioural results containing accuracy, sensitivity and criterion in Experiment 3. Standard deviation in parentheses.

Task Type	Importance	Accuracy Percentage	*d'*	*c*
Comprehension	High Importance	91 (5)	3.03 (0.67)	-0.29 (0.35)
	Low Importance	90 (5)	2.63 (0.74)	-0.32 (0.41)
Skimming	High Importance	87 (7)	2.9 (0.67)	-0.22 (0.35)
	Low Importance	84 (6)	2.4 (0.64)	-0.4 (0.39)

Table 4.6

Fixed effect estimate for comprehension question accuracy in Experiment 3.

	Estimate	Std. Error	*z* value
Intercept	3.21	0.21	15.08 *
Importance	-0.29	0.26	-1.11
Task Type	-0.66	0.11	-5.99 *

*$z>|1.96|$

General Discussion

Experiment 3 demonstrated that skim reading has a pronounced influence on reading behaviour. As observed in the eye movement measures, fixations were shorter on average for the skim reading condition, replicating the findings of Just and Carpenter (1987). However, the most interesting finding here is in relation to the impact that hyperlinks have on reading behaviour. In Experiment 1C we found that hyperlinks had only a limited impact upon on reading behaviour, apart from increased re-reading of the low frequency, hyperlinked words. We did replicate the finding that when reading for comprehension, hyperlinks are not a hindrance to reading. However, we did not replicate the increased re-reading of the low frequency, hyperlinked words. A suggestion for why we did not replicate this finding may be due to the fact that our go-past times measures were quite insensitive due to relatively rare occurrences of re-reading. Moreover, we had four comprehension questions after each article in Experiment 3, whereas in Experiment 1C there was only one comprehension question after each article. This subtle difference may have changed the nature of the task. A more likely reason is the fact that the amount of text to read was doubled, Experiment 3 contained forty edited Wikipedia articles compared to only twenty in Experiment 1C. The participants having to read such a large amount of text may have discouraged re-reading behaviour. There was an overall reduction in the amount of regressions on the target words in Experiment 3 when compared to Experiment 1C. In Experiment 1C, 20.50% of target words had regressions, compared to 14.11% of target words in Experiment 3. However, the key finding that hyperlinks have no pronounced detrimental effects of reading behaviour remained the case.

During skim reading, not only were the hyperlinked words less likely to be skipped, they were more likely to be the words that were fully processed. This is shown by the presence of a frequency effect being observed in the hyperlinked words for skim reading, but not for the unlinked words. We suggest that the reason for the lack of a frequency effect in the unlinked words is because they were not being fully processed. This lack of a frequency effect is similar to observations in visual search of text (i.e. searching for a target word in a passage of text) when the reader is also presumed to not be lexically processing the words (Rayner & Raney, 1996).

In most of the global sentence reading measures, we found a main effect of Task Type, where there were longer first pass gaze durations on the sentence when it was read for comprehension compared to skim reading. This suggests that the readers spent longer on the sentences when reading for comprehension. We also found a main effect of the importance ratings, where the higher the rating, the longer the gaze durations. This means that the readers did spend longer on sentences rated as more important than those rated as less important. Furthermore, in wrap up gaze duration and total reading time there was an interaction between task type and importance rating. Higher importance ratings resulted in longer gaze durations for first-pass gaze durations and longer total time on the wrap up region. However, in skim reading the importance rating did not seem to have such a large impact as it does when reading for comprehension. This suggests that when skim reading the reader has less opportunity to establish what is less and more important and therefore cannot utilise importance as successfully as when reading for comprehension.

We observed that comprehension accuracy declined when the participants were skim reading, but we also found that they performed somewhat better on the comprehension questions regarding the sentences that were rated as more important. However, this effect was only marginal, perhaps as a result of a ceiling effect of high overall accuracy. Still, this could suggest that the participants were prioritising the more important information effectively. We found that the sentences rated as more important contained more hyperlinks on average. Participants may have been using the links as anchors or signals throughout the text if the links denote the most important information. Typographical signals have been shown to result in the reader paying more attention to the signalled content (Lorch, Lorch & Klusewitz, 1995) and often in improved memory for the signalled text (Cashen & Leicht, 1970; Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch et al, 1995). It has also been previously shown that hyperlinks can assist in helping learners retain information (Nikolova, 2004), perhaps because they are working as a typographical signal.

Experiment 3 confirms that participants do read faster when skim reading hypertext and also that, when skim reading, comprehension was impaired compared to comprehension reading. However, the presence of hyperlinks had an impact on skim reading. Participants were less likely to skip linked words when skim reading and, when participants did land on linked words, the reader processed them fully, as evidenced by the significant frequency effect observed in linked target words in the skimming condition. Conversely, participants were less likely to fixate the unlinked words when skim reading. If the participants did fixate the unlinked words they did not seem to be processing them to the same degree as they processed the linked words. This is seen by the lack of a frequency effect on the unlinked words during skim reading.

If we take both the findings from the eye tracking and the comprehension results together we can suggest that readers could be engaging in an adaptive satisficing strategy, obtaining a speed-comprehension trade-off which is optimal for the task at hand. Participants may have wanted to read quickly while still retaining as much comprehension as possible. From these findings we suggest that participants used the hyperlinks as markers for the presence of important information and used them in a strategy to skim read through the text in the most efficient way possible.

In terms of Web design and the creation of hypertext documents the key lesson here is that if readers are skim reading on the Web as other researchers have suggested (Liu, 2005; Morkes & Nielsen, 1997), then the words that are chosen as hyperlinks have to be taken seriously. These are the words that the reader will be processing and the reader may be using hyperlinks as a signal for the most important information in the page. This would mean that when skim reading the words chosen as hyperlinks become very important to the reader. Our findings show that during skim reading, the reader is more likely to fully process the hyperlinked words and might be using them as an anchor for finding the most important information in the page. However, this is not a recommendation to include more hyperlinks in webpages. Superfluous linking may be damaging to this strategy. If a Wikipedia article contains hyperlinks to articles that are completely irrelevant to the main topic, this would not be useful for the reader. From signalling research we know that making a keyword or phrase distinct in the text means that readers pay more attention to the emphasised content when reading (Lorch, Lorch & Klusewitz, 1995) and that this can result in better memory for the emphasised text (Cashen & Leicht, 1970; Crouse & Idstein, 1972; Fowler & Barker, 1974; Lorch et al, 1995). However, simply emphasising text is not enough to encourage recall of the text. The reader needs to know what the typographical cue represents and be able to utilise it (Lorch et al., 2011). If the cue is meaningless to the readers' goal then the reader will not utilise the signalling (Golding & Fowler, 1992). Another issue relating to the signalling cue being useful for the task at hand is that there is the issue of "over-signalling" that reduces the effectiveness of typographical cues. If a large amount of the text has some form of signal, or in this case if there are a lot of hyperlinks present in the text, then the importance of the hyperlinked words may not seem as useful for signalling importance. Emphasis only works well, if you emphasise only the important parts of the content (Lorch et al., 1995).

Chapter 5

A Novel Methodology for Creating an Interactive Non-Linear Space in an Eye Tracking Environment

Introduction

In the previous experiments in this thesis, we have not allowed the reader the ability to click and navigate through a Web environment. We did this in order to have well-controlled experiments exploring the visual impact of hyperlinks separately from the impact of navigation. However, in order to know if the experimental results from the previous experiments hold true in a realistic Web environment, we need to also explore the same research questions while allowing the reader to click and navigate through this environment. However, there are several issues that mean this is a difficult task. Firstly, there was no existing software that could be used to accomplish this goal. The EyeLink 1000 eye tracker, used for the eye tracking experiments in this thesis, cannot simply track the reader's eyes while displaying content in a Web browser. If we want to create an interactive website, it needs to be displayed in a way that is compatible with the EyeLink system. Essentially, the webpages need to be displayed as individual images with locations in pixel co-ordinates for the reader to click and navigate. Therefore, the aim was to create subsections of Wikipedia (which we refer to as hubs), that contain enough articles that the reader could navigate through the Web environment without running out of articles so that the experience feels like a real Web environment.

To create a realistic Web browsing environment that had the required experimental control requires a large amount of design and planning. There were a number of issues that needed to be addressed in order to create this experiment and those are discussed in this chapter in detail.

Traditional Eye Tracking

Traditional eye movement and reading experiments tend to follow a similar pattern of design and analysis. This can be witnessed in the previous chapters in this thesis. A set of stimuli is created and then displayed in a pseudo random order. The readers will read all of the stimuli presented to them and they have no interactive options as they can only read in the pre-specified order and then answer comprehension questions about the stimulus that has just been presented. This is relatively straightforward to create in current eye tracking software such as Experiment Builder. However, once the experimental design becomes more complex this procedure can be insufficient and a new methodology needs to be created. In this thesis, a new methodology is devised based on maintaining the positive points that the traditional methodology offers and we try to build on this to make a new methodology to explore reading on the Web.

The positive aspects of traditional eye tracking and reading research include:

Accuracy and linked to cognitive processes: The EyeLink 1000 eye-tracker, used for all eye-tracking experiments in this thesis, has high speed and accuracy. It records a sample every millisecond and participants are calibrated to within a third of a degree of visual angle, which roughly translates to the size of a single character. Moreover, this very detailed track record of eye movement data gives us a good indication of moment-to-moment cognitive processes during reading (See Introduction, Liversedge & Findlay, 2000; Rayner, 1998, 2009).

Well tested and established methodology: There are a number of eye movement measures (e.g. first fixation duration, go-past times, etc. see Introduction) that have been developed to investigate and understand reading behaviour (Rayner, 1998) that are well understood and can help us represent not only the presence of effects, but also the time course of effects.

Easy to use software for experiment creation and data analysis: The current eye tracking software for the Eyelink 1000 such as Experiment Builder and Data Viewer are adequate to create and analyse simple, static reading experiments. Interest areas and experimental design are easy to establish and implement and the eye tracking data is simple to output.

However, there are also some drawbacks to the traditional methodology/software:

Complexity ceiling: When the experiment becomes more complex the Experiment Builder software on its own is not adequate. It is designed to make the creation of simple, single target word, reading experiments. However, in order to represent an environment that resembles the Web, and can be interactively interacted with, additional software is required. This software needs to create the images and interest areas on-line and integrate additional clicking functionality in order to create a realistic, interactive website environment.

Analysis issues: A large amount of traditional eye tracking measures can also be used in the analysis of more complex passages of text. However, they need to be modified to work in this new scenario. For example, most of the ways data can be output from Data Viewer results in simple, first-pass reading measures for the purposes of target word analysis. However, when reading a website, participants may scan the page before engaging in reading, thus resulting in interest areas on the entire page being marked as re-reading (because they have supposedly been skipped during first passage). Trials where this occurs render all first-pass measures void, reducing the power of traditional reading measure analysis.

Given these good and bad points about eye tracking research, we need to keep the positive points, while addressing the negative aspects. Therefore, for this thesis a new methodology was devised, with software created specifically to address the required complexity and data analysis issues discussed above. Essentially, an additional stage took place before the experiment was put into Experiment Builder such that we could insert an extra level of complexity into the experimental design, while maintaining the accuracy of traditional eye tracking methodology.

New Methodology

The new methodology required us to first consider the problems with implementing our experimental design into the current eye tracking software. The current software does not allow a website like environment to be created easily. Although code can be implemented to be able to click a region and it presents an appropriate image linked to that region, it is difficult to keep track of previous pages and practically impossible to have functionally like a website that includes making any previously visited links purple. Essentially, we need to add additional functionality that is not covered by default in Experiment Builder. These are:

Clickable regions for navigation: The location of each word and size needs to be identified and monitored such that when a word is linked the clickable region is known and can be utilised. This allows a user to click the word and navigate to a linked page.

Interest areas: Each word location needs to be known such that fixations within interest areas can be monitored. In Experiment Builder, interest areas can be created, but they tend to be created by hand, drawing each one on the stimulus image. This is time consuming (especially as the number of stimuli increases) and human error could easily occur. For a clickable Web environment, a number of interest areas need to created, ideally one for each word. The interest area file needs to also contain the details about each word such that during analysis each word can be distinguished and its variables are known, such as whether it is linked or not and whether it is a target word.

Visited links: When navigating on the Web we commonly click blue links that turn purple when they are visited links. The program needs to be able to keep track of the links that have been clicked and the revisited pages and the visited links need to change in colour from blue to purple. This adds to the ecological validity of the website experience.

Manageable hubs with no dead links: Each hub needs to have links in the articles that link to other, separate articles. All links need to be functional (i.e. they need to have a location that they link to when clicked) and we need to program in a way to check the web environment is complete in that it has no dead links that lead to nowhere. If a link is clicked and it has no page linked to it then the program might crash.

The Web is an expansive, non-linear space: The Web is large and expansive and if the reader is going to feel like they are actually browsing the Web, they need the environment to feel large and expansive. There needs to be enough pages and links present in those pages that the user feels like they are browsing Wikipedia, without the need to re-make Wikipedia with the 39,468,450 pages it currently has (as of June

2016). Due to the nature of this research, target words need to be inserted into the articles which could be very labour intensive.

Part 1: Software to make images and interest areas

The first task was to create custom software that can be used to create Wikipedia articles with defined and constrained target words that can be manipulated. The text within a Wikipedia article then needed to be turned into images that can be displayed by Experiment Builder with matching interest areas. Two versions of each article were created, one with all blue links and one with all purple links, representing the visited links (see Figure 5.1).

The custom software was created in C# and was named The Clicking Gemerator. It imports data from a specifically created input document (Figure 5.2) and generates images, interest areas and files necessary for creation of the subsequent Experiment Builder experimental set up.

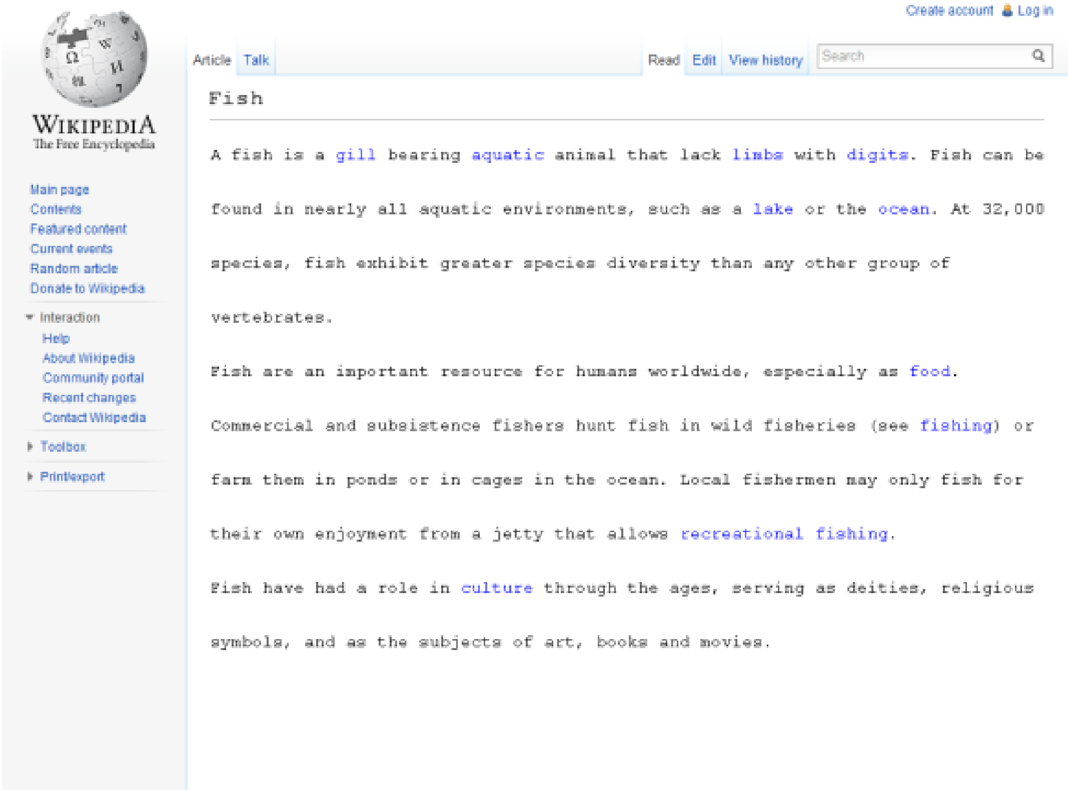
This input document (See Figure 5.2) contains information for generating each individual Wikipedia article. Each row corresponds to single article. Each page needs the information listed in Table 5.1. The text that is shown in each article is defined in the "TEXT" column. In order to manipulate the target words and what words are linked, we defined our own markup for target words and whether a word is linked and what corresponding page will be presented when that link is clicked on. This input file is imported into the Clicking Gemerator program (Figure 5.3)

In Figure 5.4, we see the example of the article labelled "Fish". Triple hash signifies the start and end of markup, similar to < and > in HTML. In the example in Figure 5.4 we see the text "A fish is a ###**fish_gill**###gill". This signifies that the word *gill* is a linked word and the link goes to the page "fish_gill" which is specified as a unique identifier in the "PAGE_NAME" column. A paragraph break is signalled by the markup "##NEWLINE##". Finally, target words are signalled with the markup "##A##", "##B##", "##C##" or "##D##", dependent on how many target words are in a particular article. In this example there are two: "##A##", "##B##". The target words are defined by the following columns of the input file: "X_TARGETS", "X_TARGET_LINKS" and "X_TARGET_PAGES". The X can either be A, B, C or D dependent on the counterbalancing condition. These columns represent the target word, with the character given in place of X referring to whether that word should be linked and what page the target word links to when it is linked.

This input is then imported into the Clicking Gemerator program and the markup is interpreted resulting in the creation of images and interest area files from the text input. The interest area files are generated by creating an interest area around each word and marking up each interest area in the file with what the word is, whether it was linked or not and whether it was a target word. It also includes the sentence number and word number in the article (see Figure 5.5).

The program also creates some input files that are put into Experiment Builder. One file is a list of all the links in the experiment and links to the image that has the visited version of the page where all the links are purple (see Figure 5.6). The second is a region list that lists out the co-ordinates of every linked word so that when a mouse cursor is inside one of these co-ordinates and is clicked, Experiment Builder can know to present the appropriate corresponding linked page the participant wanted to navigate to (see Figure 5.7). These co-ordinates are also used to display the purple, visited links.

The visited links are worked out from a mixture of these two files that define what words are links, the co-ordinates for those linked words and the image of the visited linked version of that article. In Experiment Builder, a variable is created to keep track of what articles have been visited and if the article has been visited before. If so, Experiment Builder can use the co-ordinates and the link to the visited version of the article to show the visited link in purple.



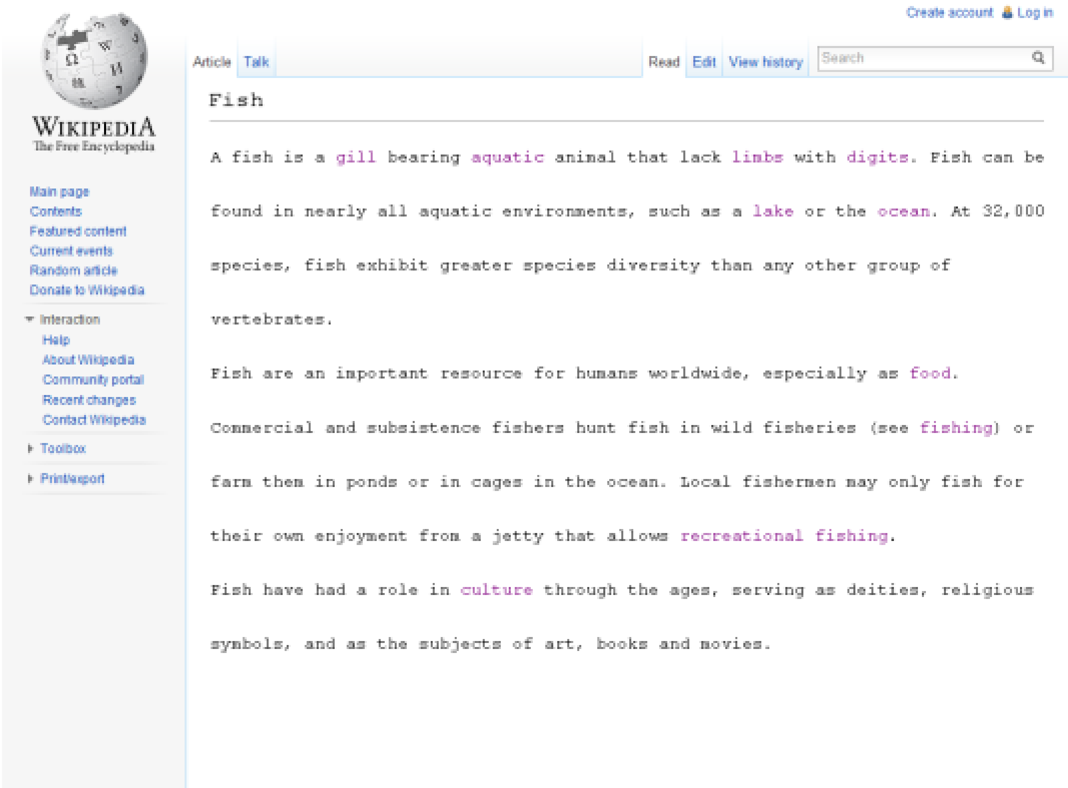


Figure 5.1 Unvisited and visited links an article. The top article is the unvisited article where all the links are blue. The bottom article is the visited version of the article where all the links are purple.

Table 5.1

All of the columns in the input file that is imported into the Clicking Gemerator and a description of what the role of each variable is.

Column name	Description
PAGE NAME	Unique identifier
TITLE	The title displayed at the top of the article
SITE	The hub the article belongs to
IS ENTRY POINT	If TRUE then this is the first page in the hub
TEXT	The text displayed in the article including code for signalling links and target words
A_TARGETS	What target words to display for counterbalancing group A
A_TARGET_LINKS	Whether the target words are linked (TRUE) or unlinked (FALSE) for counterbalancing group A
A_TARGET_PAGES	The pages each target word should link to for counterbalancing group A
B_TARGETS	What target words to display for counterbalancing group B
B_TARGET_LINKS	Whether the target words are linked (TRUE) or unlinked (FALSE) for counterbalancing group B
B_TARGET_PAGES	The pages each target word should link to for counterbalancing group B
C_TARGETS	What target words to display for counterbalancing group C
C_TARGET_LINKS	Whether the target words are linked (TRUE) or unlinked (FALSE) for counterbalancing group C
C_TARGET_PAGES	The pages each target word should link to for counterbalancing group C
D_TARGETS	What target words to display for counterbalancing group D
D_TARGET_LINKS	Whether the target words are linked (TRUE) or unlinked (FALSE) for counterbalancing group D
D_TARGET_PAGES	The pages each target word should link to for counterbalancing group D
QUESTION_TEXT	The comprehension question for the article if it has one or not (NONE)

EXPECTED_ANSWER	The expected answer for the comprehension question, either true (T) or false (F)
NUMBER_OF_TARGETS	Number of target words in the article
TARGET_ONE	First target word pair (high frequency_low frequency)
TARGET_TWO	Second target word pair (high frequency_low frequency)
TARGET_THREE	Third target word pair (high frequency_low frequency)
TARGET_FOUR	Fourth target word pair (high frequency_low frequency)

Figure 5.2 Input file that is created in Excel and imported into the Clicking Generator. One row corresponds to a single article.

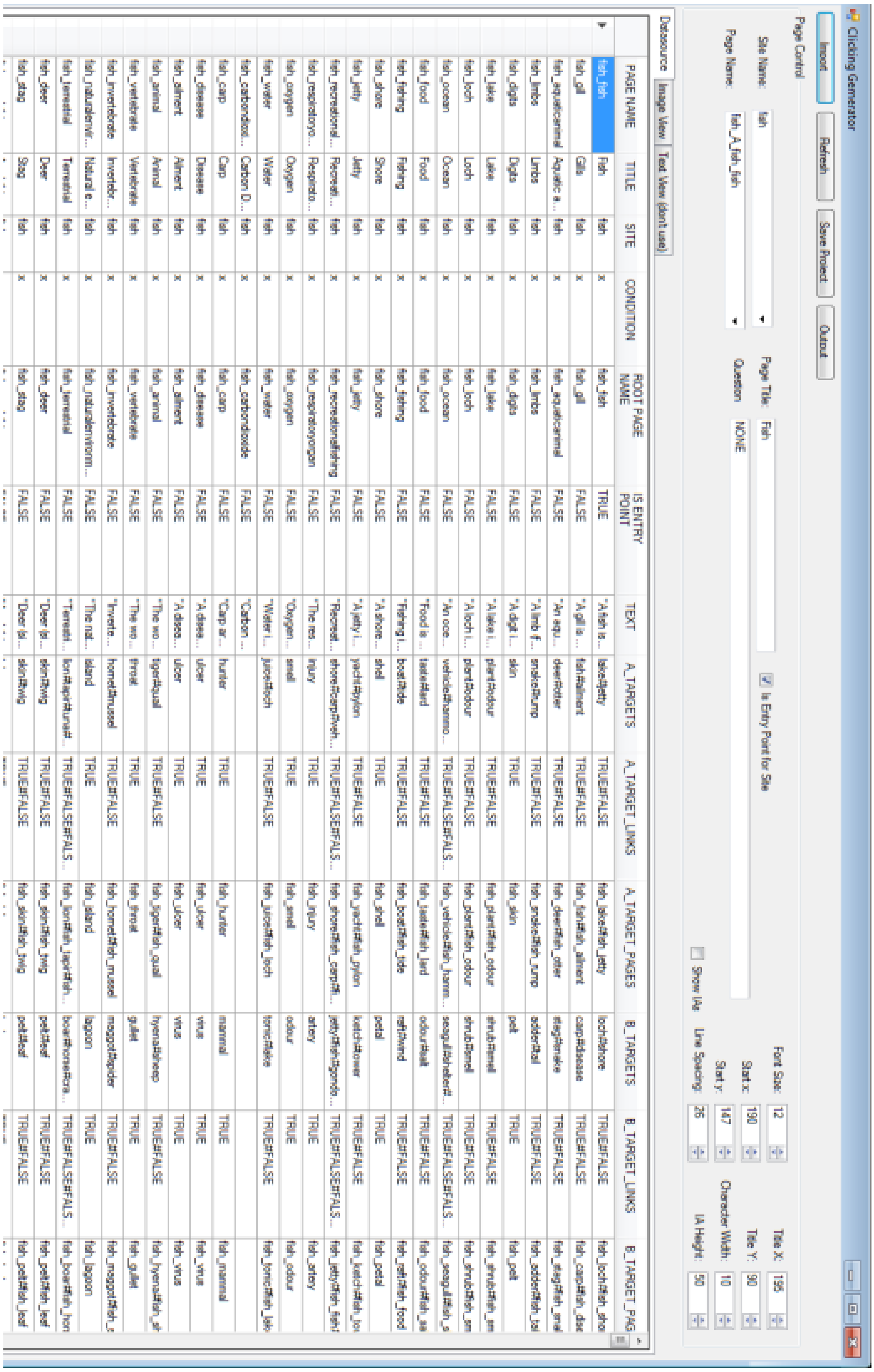


Figure 5.3 The interface for the Clicking Generator. The input file has already been imported and there are a number of variables that can be manipulated such as font size, interest area height, starting location (in pixel co-ordinates) and line height for the text to be displayed.

A fish is a **###fish_gill###**gill bearing **###fish_aquaticanimal###**aquatic animal that lack **###fish_limbs###**limbs with **###fish_digits###**digits. Fish can be found in nearly all aquatic environments, such as a **##A##** or the **###fish_ocean###**ocean. At 32,000 species, fish exhibit greater species diversity than any other group of vertebrates**##NEWLINE##**. Fish are an important resource for humans worldwide, especially as **###fish_food###**food. Commercial and subsistence fishers hunt fish in wild fisheries (see **###fish_fishing###**fishing) or farm them in ponds or in cages in the ocean. Local fishermen may only fish for their own enjoyment from a **##B##** that allows **###fish_recreationalfishing###**recreational **###fish_recreationalfishing###**fishing**##NEWLINE##**.Fish have had a role in **###fish_culture###**culture through the ages, serving as deities, religious symbols, and as the subjects of art, books and movies.

Figure 5.4 An example of the "TEXT" column in the input file. This includes the text that will be displayed on the article page including custom markup that will be interpreted by the Clicking Gemerator program in order to understand what words are linked and where to display target words.

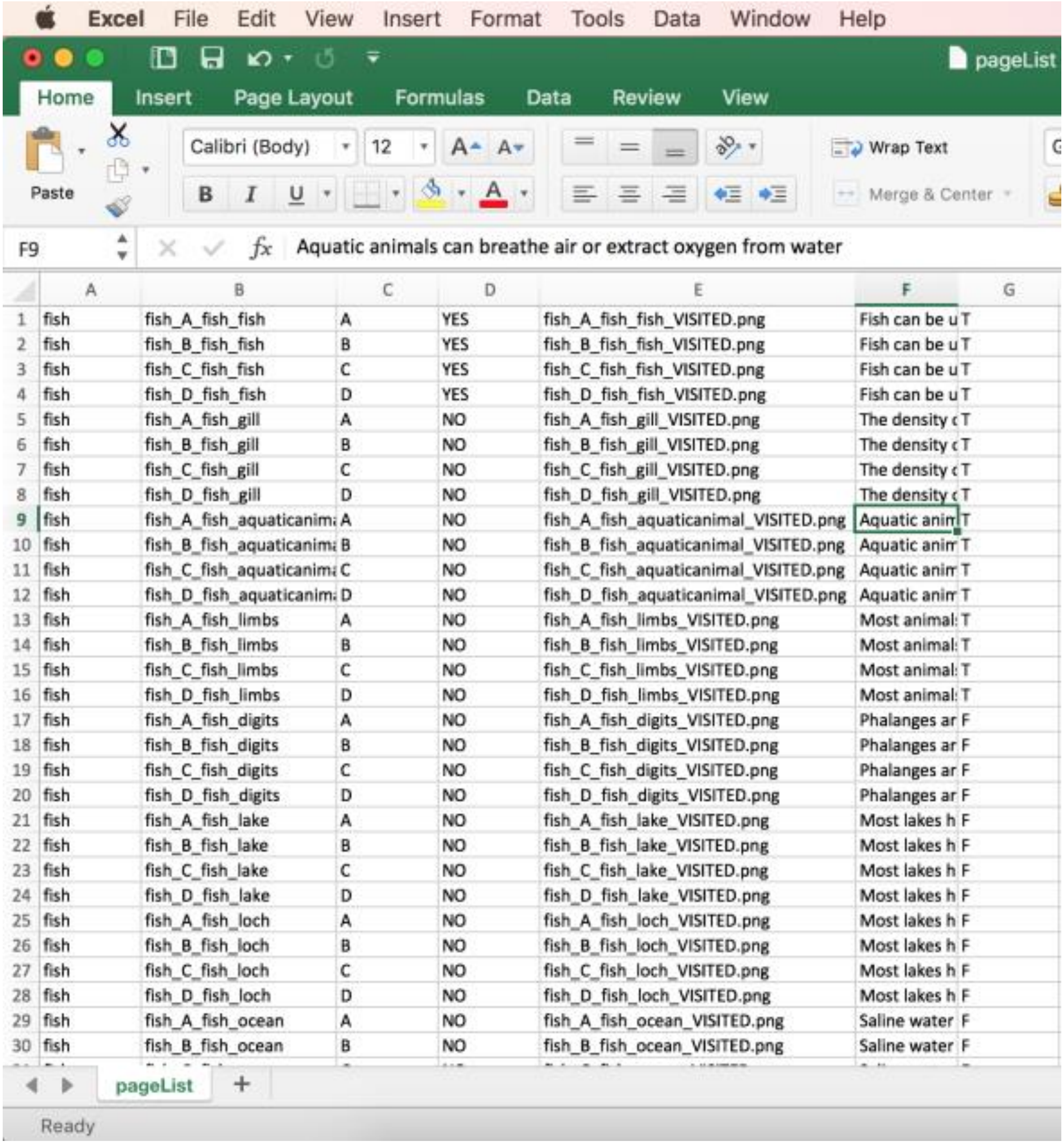
fish_A_fish_fish.ias

File Path : ~/Downloads/fish_A_fish_fish.ias

fish_A_fish_fish.ias

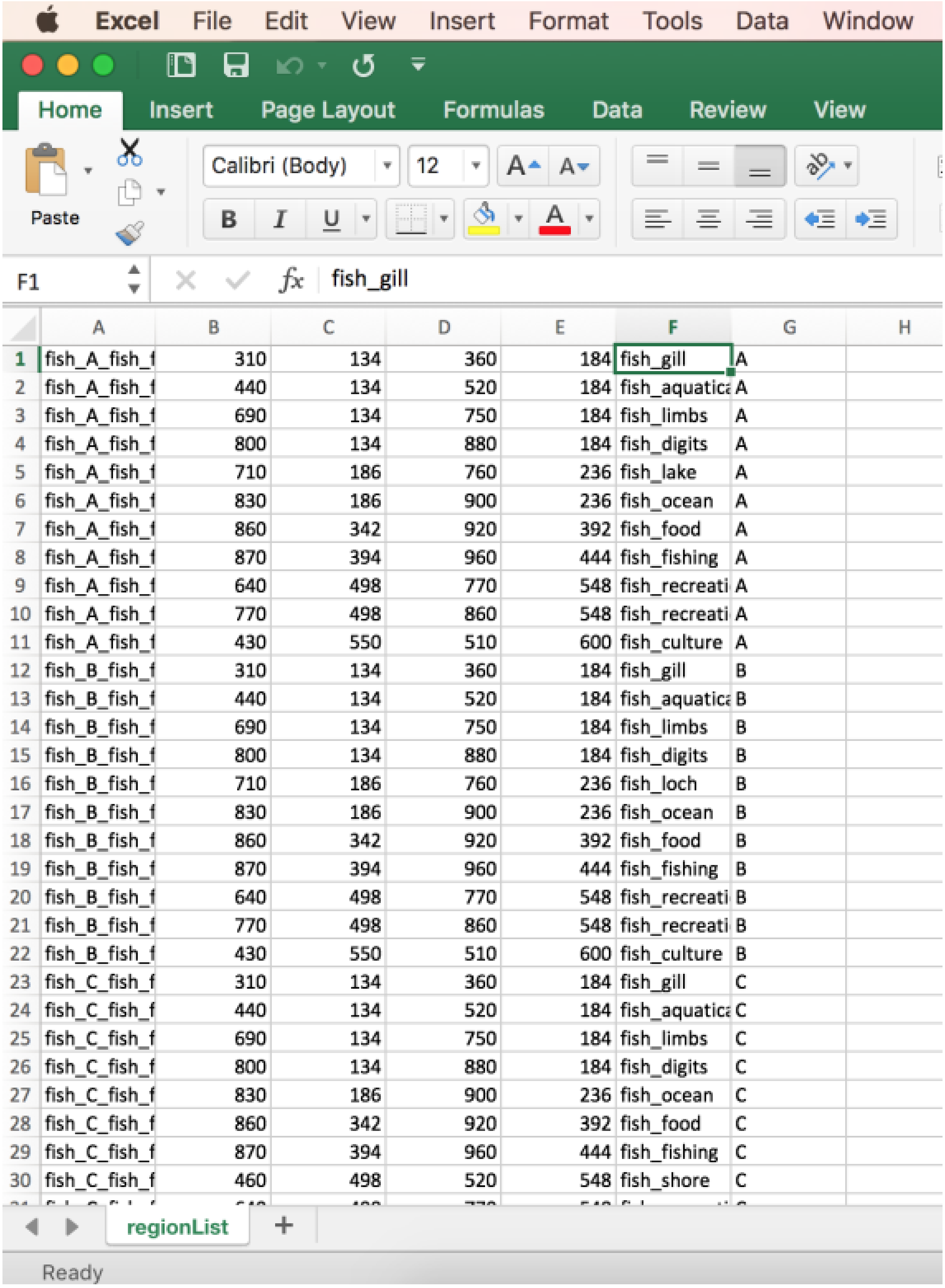
```
# EyeLink Interest Area Set created on 06-21-15 by the wikiGemerator
RECTANGLE	0	190 134 210 184 0_1_A
RECTANGLE	1	210 134 260 184 0_2_fish
RECTANGLE	2	260 134 290 184 0_3_is
RECTANGLE	3	290 134 310 184 0_4_a
RECTANGLE	4	310 134 360 184 0_5_gill_LINKED
RECTANGLE	5	360 134 440 184 0_6_bearing
RECTANGLE	6	440 134 520 184 0_7_aquatic_LINKED
RECTANGLE	7	520 134 590 184 0_8_animal
RECTANGLE	8	590 134 640 184 0_9_that
RECTANGLE	9	640 134 690 184 0_10_lack
RECTANGLE	10	690 134 750 184 0_11_limbs_LINKED
RECTANGLE	11	750 134 800 184 0_12_with
RECTANGLE	12	800 134 880 184 0_13_digits._LINKED
RECTANGLE	13	880 134 930 184 1_14_Fish
RECTANGLE	14	930 134 970 184 1_15_can
RECTANGLE	15	970 134 1000	184 1_16_be
RECTANGLE	16	190 186 250 236 1_17_found
RECTANGLE	17	250 186 280 236 1_18_in
RECTANGLE	18	280 186 350 236 1_19_nearly
RECTANGLE	19	350 186 390 236 1_20_all
RECTANGLE	20	390 186 470 236 1_21_aquatic
RECTANGLE	21	470 186 610 236 1_22_environments,
RECTANGLE	22	610 186 660 236 1_23_such
RECTANGLE	23	660 186 690 236 1_24_as
RECTANGLE	24	690 186 710 236 1_25_a
RECTANGLE	25	710 186 760 236 1_26_lake_TARGET_LINKED
RECTANGLE	26	760 186 790 236 1_27_or
RECTANGLE	27	790 186 830 236 1_28_the
RECTANGLE	28	830 186 900 236 1_29_ocean._LINKED
RECTANGLE	29	900 186 930 236 2_30_At
RECTANGLE	30	930 186 1000	236 2_31_32,000
RECTANGLE	31	190 238 280 288 2_32_species,
RECTANGLE	32	280 238 330 288 2_33_fish
RECTANGLE	33	330 238 410 288 2_34_exhibit
RECTANGLE	34	410 238 490 288 2_35_greater
RECTANGLE	35	490 238 570 288 2_36_species
RECTANGLE	36	570 238 670 288 2_37_diversity
RECTANGLE	37	670 238 720 288 2_38_than
RECTANGLE	38	720 238 760 288 2_39_any
```

Figure 5.5 Example of an interest area file creating by the Clicking Gemerator.



	A	B	C	D	E	F	G
1	fish	fish_A_fish_fish	A	YES	fish_A_fish_fish_VISITED.png	Fish can be u	T
2	fish	fish_B_fish_fish	B	YES	fish_B_fish_fish_VISITED.png	Fish can be u	T
3	fish	fish_C_fish_fish	C	YES	fish_C_fish_fish_VISITED.png	Fish can be u	T
4	fish	fish_D_fish_fish	D	YES	fish_D_fish_fish_VISITED.png	Fish can be u	T
5	fish	fish_A_fish_gill	A	NO	fish_A_fish_gill_VISITED.png	The density c	T
6	fish	fish_B_fish_gill	B	NO	fish_B_fish_gill_VISITED.png	The density c	T
7	fish	fish_C_fish_gill	C	NO	fish_C_fish_gill_VISITED.png	The density c	T
8	fish	fish_D_fish_gill	D	NO	fish_D_fish_gill_VISITED.png	The density c	T
9	fish	fish_A_fish_aquaticanim:	A	NO	fish_A_fish_aquaticanimal_VISITED.png	Aquatic anim	T
10	fish	fish_B_fish_aquaticanim:	B	NO	fish_B_fish_aquaticanimal_VISITED.png	Aquatic anim	T
11	fish	fish_C_fish_aquaticanim:	C	NO	fish_C_fish_aquaticanimal_VISITED.png	Aquatic anim	T
12	fish	fish_D_fish_aquaticanim:	D	NO	fish_D_fish_aquaticanimal_VISITED.png	Aquatic anim	T
13	fish	fish_A_fish_limbs	A	NO	fish_A_fish_limbs_VISITED.png	Most animal:	T
14	fish	fish_B_fish_limbs	B	NO	fish_B_fish_limbs_VISITED.png	Most animal:	T
15	fish	fish_C_fish_limbs	C	NO	fish_C_fish_limbs_VISITED.png	Most animal:	T
16	fish	fish_D_fish_limbs	D	NO	fish_D_fish_limbs_VISITED.png	Most animal:	T
17	fish	fish_A_fish_digits	A	NO	fish_A_fish_digits_VISITED.png	Phalanges ar	F
18	fish	fish_B_fish_digits	B	NO	fish_B_fish_digits_VISITED.png	Phalanges ar	F
19	fish	fish_C_fish_digits	C	NO	fish_C_fish_digits_VISITED.png	Phalanges ar	F
20	fish	fish_D_fish_digits	D	NO	fish_D_fish_digits_VISITED.png	Phalanges ar	F
21	fish	fish_A_fish_lake	A	NO	fish_A_fish_lake_VISITED.png	Most lakes h	F
22	fish	fish_B_fish_lake	B	NO	fish_B_fish_lake_VISITED.png	Most lakes h	F
23	fish	fish_C_fish_lake	C	NO	fish_C_fish_lake_VISITED.png	Most lakes h	F
24	fish	fish_D_fish_lake	D	NO	fish_D_fish_lake_VISITED.png	Most lakes h	F
25	fish	fish_A_fish_loch	A	NO	fish_A_fish_loch_VISITED.png	Most lakes h	F
26	fish	fish_B_fish_loch	B	NO	fish_B_fish_loch_VISITED.png	Most lakes h	F
27	fish	fish_C_fish_loch	C	NO	fish_C_fish_loch_VISITED.png	Most lakes h	F
28	fish	fish_D_fish_loch	D	NO	fish_D_fish_loch_VISITED.png	Most lakes h	F
29	fish	fish_A_fish_ocean	A	NO	fish_A_fish_ocean_VISITED.png	Saline water	F
30	fish	fish_B_fish_ocean	B	NO	fish_B_fish_ocean_VISITED.png	Saline water	F

Figure 5.6 One of the two additional files created by the Clicking Gemerator. This file lists all of the links in each article, one row per link. It also contains the name of the image that displayed all of the links as visited. This file is used by Experiment Builder to display the visited links.



	A	B	C	D	E	F	G	H
1	fish_A_fish_f	310	134	360	184	fish_gill	A	
2	fish_A_fish_f	440	134	520	184	fish_aquatica	A	
3	fish_A_fish_f	690	134	750	184	fish_limbs	A	
4	fish_A_fish_f	800	134	880	184	fish_digits	A	
5	fish_A_fish_f	710	186	760	236	fish_lake	A	
6	fish_A_fish_f	830	186	900	236	fish_ocean	A	
7	fish_A_fish_f	860	342	920	392	fish_food	A	
8	fish_A_fish_f	870	394	960	444	fish_fishing	A	
9	fish_A_fish_f	640	498	770	548	fish_recreati	A	
10	fish_A_fish_f	770	498	860	548	fish_recreati	A	
11	fish_A_fish_f	430	550	510	600	fish_culture	A	
12	fish_B_fish_f	310	134	360	184	fish_gill	B	
13	fish_B_fish_f	440	134	520	184	fish_aquatica	B	
14	fish_B_fish_f	690	134	750	184	fish_limbs	B	
15	fish_B_fish_f	800	134	880	184	fish_digits	B	
16	fish_B_fish_f	710	186	760	236	fish_loch	B	
17	fish_B_fish_f	830	186	900	236	fish_ocean	B	
18	fish_B_fish_f	860	342	920	392	fish_food	B	
19	fish_B_fish_f	870	394	960	444	fish_fishing	B	
20	fish_B_fish_f	640	498	770	548	fish_recreati	B	
21	fish_B_fish_f	770	498	860	548	fish_recreati	B	
22	fish_B_fish_f	430	550	510	600	fish_culture	B	
23	fish_C_fish_f	310	134	360	184	fish_gill	C	
24	fish_C_fish_f	440	134	520	184	fish_aquatica	C	
25	fish_C_fish_f	690	134	750	184	fish_limbs	C	
26	fish_C_fish_f	800	134	880	184	fish_digits	C	
27	fish_C_fish_f	830	186	900	236	fish_ocean	C	
28	fish_C_fish_f	860	342	920	392	fish_food	C	
29	fish_C_fish_f	870	394	960	444	fish_fishing	C	
30	fish_C_fish_f	460	498	520	548	fish_shore	C	

Figure 5.7 One of the two additional files created by the Clicking Gemerator. This file contains all of the links present (one row per link) and their pixel co-ordinates in each article. This file is used to detect whether a mouse click is within a linked region and also to assist with showing visited links in the correct location.

Part 2: Code to manage hubs and visited links

The input needs to be checked to ensure that there are no dead-end links that could crash the experiment. Each link needs to have a unique identifier and any link needs a corresponding article to exist. This is important for tracking which pages people have already seen and making sure the visited links appear when appropriate. An R script was created to go through the input file and check whether every link in the 'X_TARGET_PAGES' column exists in the 'PAGE_NAME' column, which is the unique identifier for each page. If there are any links to pages that do not exist, the script flags up any of these instances of broken links. These were usually simple typographical errors that were easily corrected.

Part 3: Combine in Experiment Builder to get a complex, interactive Web browsing environment

Taking the images and interest areas and putting them into Experiment Builder is the final step towards building an interactive Web environment. We defined earlier that we would need four issues addressed if we wanted a well-controlled, interactive Web environment. These were:

Clickable regions for navigation: We have a list of all of the regions of each linked word and these can be fed into Experiment Builder and the experiment program can be created to respond to a mouse click in these regions.

Interest areas: We have created our own interest area files that contain the location for every word in the display. These interest area files also contain additional information about every word, i.e., whether it is linked, whether it is a target word and its sentence number and word number in the article.

Visited links: This issue requires a few steps to solve, in order to include visited links in the experiment. Firstly, we need to be able to monitor which pages have been visited and display those links in purple. Therefore, we need the location of all links and the ability to list all the pages the reader has already visited. Secondly, we need to have a way of displaying the articles including purple links.

Solving the first problem is simple, given the fact we already have a list of links and the regions they occupy. The only additional programming required in the Experiment Builder experimental file is to include a function that adds each article

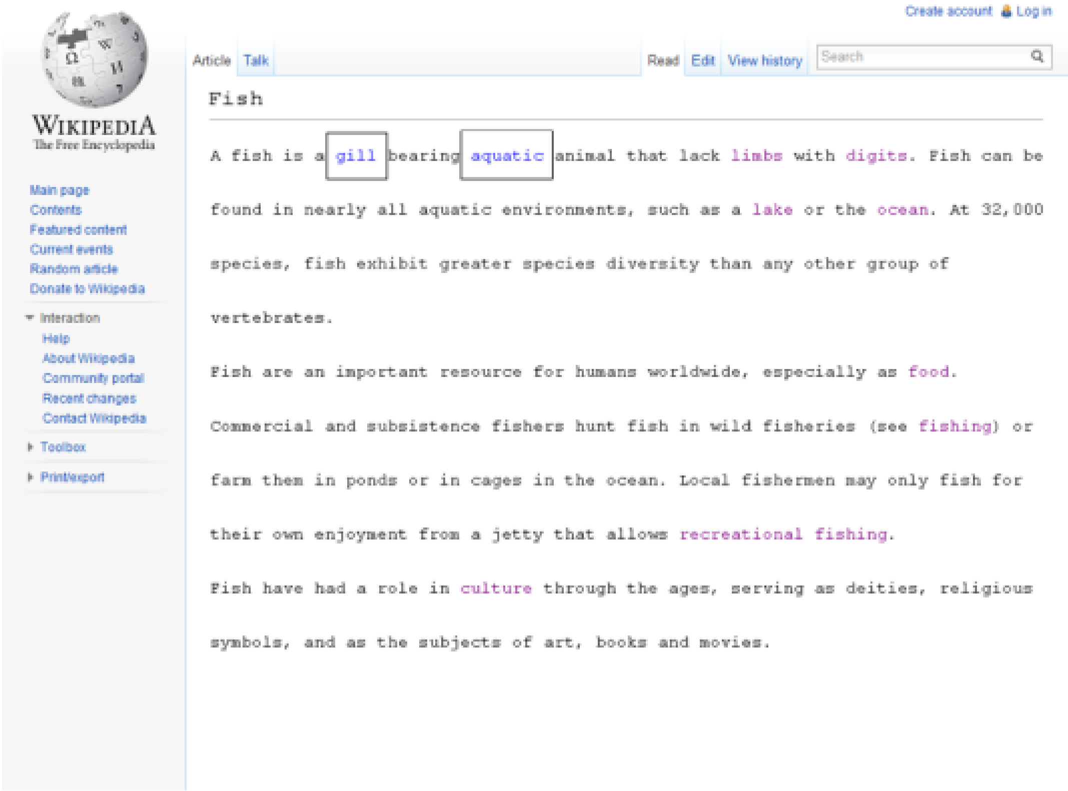


Figure 5.8 An example of how the visited and unvisited links are created. The visited page (with all purple links) is the default page. The linked regions of the unvisited page containing blue links is shown on top of the visited page unless that particular link has been visited. In this example, we can see that the regions for the linked words 'gill' and 'aquatic' have the unvisited linked words laid on top of the default visited image (black border is only present to illustrate the edges of the link region).

that is visited to a list. When a new article is visited/generated that list is checked to know whether to display the link as blue (unvisited) or purple (visited).

The second problem of actually displaying the visited links is slightly more complicated. From the Clicking Gemerator, we generated two versions of each article, one where all the linked words were blue and one where all the linked words are purple. Experiment Builder runs using Python and there is the ability to insert custom classes into the experiments. In this case, custom classes were created to display the visited version as the default image for the display of each article. The link co-ordinates are known from the region file (see Figure 5.7) that was generated by the Clicking Gemerator. Experiment Builder is instructed in the custom Python class to display the linked regions of the unvisited version of the page on top of unvisited version of the article. If the page has been visited before, Experiment Builder is instructed to not display the unvisited region for that particular link (see Figure 5.8).

Manageable hubs with no dead links: This issue was resolved by checking that all page links existed as actual pages. This was completed via the creation of the R script mentioned in: Part 2: Code to manage hubs and visited links.

Part 4: The final issue: The Web is an expansive, non-linear space

Finally, we need to address the, non-software based issue with the following experiment:

The Web is an expansive, non-linear space: This issue is essentially a time-based problem. There was a need to create enough articles that the reader could navigate and read articles with target words, but still had the freedom and range of articles to still feel like they are in an expansive space. This was addressed by creating four hubs rather than one giant website. This way the reader could have four unique starting pages that all readers would have and then the hubs were made to only go 4 levels deep from these starting pages.

The reason for limiting the number of levels is to limit the expansiveness of each hub. The top level of the hub is the starting page and any link on that page links to a page with its own links, which in term link to pages with their own links (see Figure 5.9). Quickly, the amount of pages that need to be created can expand to an exponential amount. Therefore, the number of levels was restricted and at level four most of the links must link back to articles that already exist in the hub. Through this process, the creation of additional articles can be avoided. On average, the first page had ten links and the pages those linked to would also attempt to have ten links, but after that the number of links was dependent on the article length. Some dead-end pages (pages with no links where the reader would have to press the back button to go to the previous article) were also included to limit how expansive each hub could be. There were 843 articles created in total across four hubs. In terms of target words, out of the 843 wiki pages, 326 contained target words and these were mostly focused on the top three levels of the hubs. Some pages had multiple target words: 191 pages contained 1 target word; 124 contained 2 target words and 11 contained 4 target words. There were 472 target word pairs used, but only one target word per experimental sentence. Each article ranged between 1 and 14 sentences long; on average there were 5.44 sentences in the articles. Some of the dead-end articles contained a single sentence to reduce time spent of them when there was no links or target words that would be of interest to analyse. Each reader needed to look at ten unique pages in each hub to end the experiment so that each participant had read 40 articles. This is in line with the previous experiments in this thesis.

The process of creating this stimuli set was a long process, but I believe very worthwhile in the end. The Web environment felt expansive enough for the time the reader spent interacting with it and this was necessary for the task at hand. Although the process to create this experiment was hugely time-consuming, it is an extremely useful tool that can be re-used in the future to explore how people read on the Web, while still being able to retain a more than adequate amount of experimental control on the stimuli.

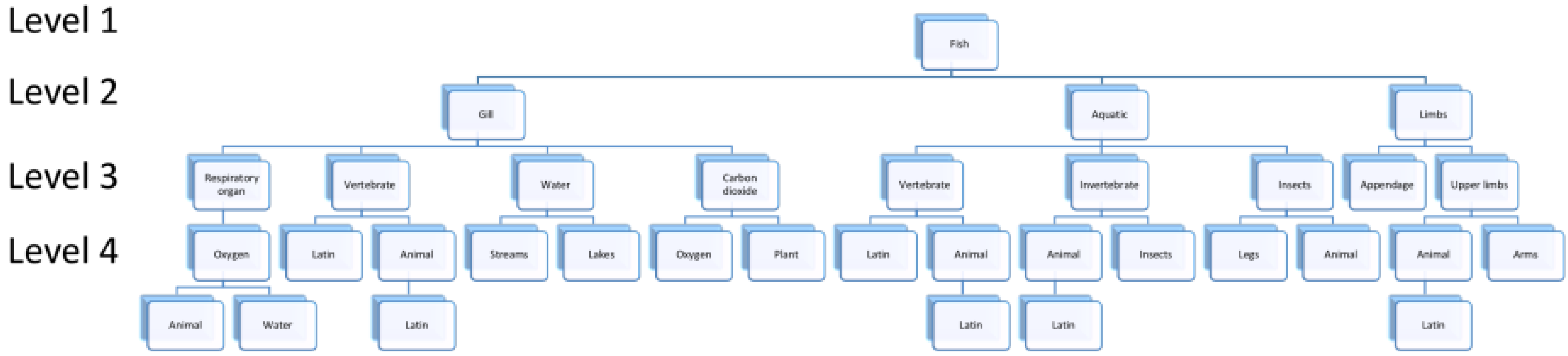


Figure 5.9 A simplified example of a hub and the number of levels. You can clearly see how from a single page of 'Fish' the number of articles needed dramatically increases at each level. This is why when level 4 is reached, the majority of links should link back to other articles that already exist in the hub and the creation of new articles is avoided. This simplified example does not include all level 4 links for the purposes of clarity.

Chapter 6

Clicking and Reading Hypertext

Introduction

In this chapter we consider how the interconnected and interactive nature of Web browsing and navigation may influence reading behaviour on the Web and especially regarding the decision making task of clicking on hyperlinks. The task of reading on the Web is different to more traditional reading for comprehension, partly because of the addition of hyperlinks. Issues discussed in Chapter 2 are the fact that hyperlinks are salient and carry high-level information (i.e. the suggestion of linked content), but another issue we have yet to explore is the impact of readers needing to make decisions regarding when to leave a particular page and which hyperlink to click on to move to a different page. This decision-making process is likely to require additional processing resources and so some have argued that it leads to an increase in cognitive load (DeStefano & LeFevre, 2007). This Chapter will also allow us to examine whether our findings reported in the previous chapters were influenced by our participants not being able to click on hyperlinks, something they would be able to do during normal reading on the Web.

Navigating through a Non-linear Space

One of the main differences between reading plain text and hypertext is the fact that hypertext is non-linear and has no strict route through the information. Conklin (1987) suggests that a reader could easily become 'lost in hyperspace' when trying to navigate through a website due to the mass interlinking of webpages. McKnight, Dillon and Richardson (1990) suggested that the unknown scope of the hypertext document could lead to incorrect assumptions about the scope of the documents' content and result in a poor reading strategy. In a linear text it is much easier to see the scope of the document and browse through the content. Dillon, Richardson and McKnight (1993) argue that if the user does not know how the information is organised, it makes it more difficult to find specific information. Paper-based documents, such as books, tend to have a convention for how the information is organised, such as index pages and contents pages, which catalogue the location of topics and convey the overall organisation of the whole text. This links back to research on signalling cues, where index/contents pages can serve as organisational cues that can be used to assist the reader in processing the text (see the Signalling section in Chapter 1).

It is not just the issue of getting lost in a hypertext environment that we need to consider; there is also the issue of the large amount of choices and decisions that need to be made. Elm and Woods (1985) suggest that users may be overwhelmed and disorientated by the sheer amount of choice offered by a complex, large network of information. The users may not understand the structure of the system and what potentially exists in the hypertext document. McDonald and Stevenson (1996) argue that although a large linear text can also be confusing for a reader, there are a number of discourse cues such as page numbers, contents listings and headings that the reader can take advantage of. Non-linear hypertext lacks a lot of these types of cues. The same text presented as a linear document may cause no issues to the reader, but in its hypertext format, it might lead to navigational problems where the reader is confused by the non-linear structure. In the next section, we explore the previous research on decision making and cognitive load to investigate the impact decision making and clicking on hyperlinks might have on cognitive load and if this would impact on reading behaviour.

Decision Making and Cognitive Load

There is an on-going debate about whether in-text hyperlinks hinder reading. Carr (2010) suggests that hyperlinks within the text are a distraction and hinder comprehension of the text. He argues that having to evaluate hyperlinks and navigating a path through them is a demanding task that substantially increases readers' cognitive load and thereby weakens their ability to comprehend and retain what they are reading (Carr 2010, pp.126).

Carr’s argument is based on research investigating the cognitive load of hypertext on users (e.g. Miall & Dobson, 2001), which suggests that comprehension increased when participants read plain text compared to when they read hyperlinked text. However, as noted by Miall and Dobson, their study is somewhat limited in being able to generalise to other forms of hypertext, including reading webpages. The text used by Miall and Dobson (2001) was a piece of literary fiction that had been converted to hypertext and hyperlinks were added to it. The text had not originally been created to be displayed in a hypertext format, making the hyperlinked document quite artificial. This artificial hypertext document may be the reason for the increase in cognitive load, in turn making it difficult to generalise these results to reading on the Web. This being said, other research does corroborate with Carr's (2010) suggestion that extra cognitive demands are associated with having to make decisions about whether to follow hyperlinks.

Some researchers have explored working memory and the concept of cognitive load and its impact on reading hypertext. DeStefano and LeFevre (2007) conducted a review of cognitive load in relation to reading hypertext. They argued that the extra task demands of reading hypertext causes an increased cognitive load to the readers in comparison to linear text. Because the readers have to make decisions about which hyperlinks to follow, additional cognitive demands are placed on working memory. Working memory models suggest that working memory is limited and only a finite amount of information can be stored at any one time (Baddeley & Logie, 1999). DeStefano and LeFevre (2007) remark that this aspect of the working memory model supports the assumption that this increased cognitive load may result in reduced performance for reading hypertext. In comparison, linear text features no hyperlinks so the reader follows the text linearly and is not required to also conduct a decision making task of how to navigate the text.

Recently, Scharinger, Kammerer and Gerjets (2015) measured both the EEG and pupil size of readers engaging in a task that closely simulated hypertext reading and link selection. They found evidence of increased load on executive functions when the reader had to perform hyperlink-like selection in the form of a significant increase in pupil size as well as a significant decrease of alpha frequency band power (these are both used as an index of cognitive load). This suggests that there is additional cognitive load placed on readers when having to make hyperlink-like selections, which could hinder reading and comprehension of the text.

It is not just the decision of whether or not to click a hyperlink that could increase cognitive load. The reader's decision to follow a hyperlink and explore different content could interrupt on-going comprehension processes. Comprehension involves the creation and development of situation models, which are complex mental representations that the reader instantiates in order to integrate statements from the text they are reading into their knowledge (Kintsch, 1988). For example, Dee-Lucas and Larkin (1995) found that hyperlinks in text distract users by interrupting information processing. While reading, users may stop to click on hyperlinks in the middle of text content, thus interrupting their cognitive processing and leaving the reader with a fragmented representation of the text content. Because of the nonlinear nature of hypertext, when a reader is reading text on one topic on a webpage, if they choose to click a hyperlink, it takes them to another webpage. This new webpage may contain content that is unrelated to the content they have just come from on the previous webpage. This could cause disruption to the reader's development of a situation model and could result in the readers' comprehension of the text being reduced.

Hyperlinks are visually salient and important navigational tools during reading of hypertext, as they represent a link to other content on the Web. We need to consider how the reader reads the text when it contains clickable hyperlinks. In this thesis, we have explored the issues of the saliency of hyperlinks, the impact of hyperlinks on the importance ratings of sentences and the differential impact of tasks effects associated with reading for comprehension versus skim reading. However, in all these cases the reader could not click and navigate the environment. This was because we wanted to maintain experimental control by simplifying the experience as much as possible in order to explore the impact of all of these factors, without yet introducing the added complexity of navigation and clicking. In the current experiment, we will be running a similar manipulation to that seen in Chapter 4, where the task was manipulated (reading for comprehension or skim reading). The target words within the text were also manipulated to be either high or low word frequency and were displayed either in blue (linked) or black (unlinked) or purple if the links have been (visited). We created an environment as close to a Web environment as possible while retaining as much experimental control as possible. The reader was able to read webpages and then choose a link to click, to navigate to another page. In this experiment, we were able to explore the impact of hyperlinks and reading on the Web in a more ecologically valid environment. We have already explored the impact of the above factors without the ability to click the links; we can now compare the findings from this experiment to the previous chapters to explore the full impact of clicking and navigating links on reading behaviour. Indeed, whereas the data collected in this Chapter can be analysed in multiple ways answering a multitude of theoretical questions, for the purpose of this thesis and compatibility with previous chapters, we will limit the reported analyses to eye movement behaviour on the target words. For similar reasons we will verbally encourage our subjects to first finish reading or skimming the page before clicking on the hyperlink of their choosing as this will allow us to more directly compare the findings in this Chapter where clicking is possible to the findings of the previous chapters where participants also read or skimmed the entire page.

Experiment 4: How Does Clicking and Navigating Hypertext Affect Reading Behaviour and Comprehension and How Does Skim Reading Interact with this?

The present experiment expands on the previous experiments, by including the ability for the reader to click on hyperlinks and navigate in the hypertext environment. We also explore the effect of task on reading and navigating hypertext, expanding our research on skim reading seen in Chapter 4.

In Chapter 4, we observed that when skim reading, readers placed an importance on hyperlinked words, as they were less likely to be skipped compared to unlinked words. And when the linked words were fixated, they were more likely to be fully processed, shown by the presence of a frequency effect for linked words that was not present for unlinked words during skim reading. This finding suggests that during skim reading, the reader tries to engage in an efficient reading strategy by assuming that hyperlinked words denote where important information might lie in the text. Therefore, the reader focuses on the hyperlinked words and those are the words that are more likely to be fixated and fully processed. However, this result was found when readers were reading what resembled hyperlinked text in the form of a

Wikipedia page, but in which they were not allowed to click or navigate these pages. The pages were simply presented to the participants for them to read and once they had finished reading the text on that page, a new random page of text was presented. The current experiment included the additional factor of clicking and navigating and investigated how this impacts on readers' behaviour when reading hypertext.

By first exploring the impact of reading hypertext without the complexity of clicking and navigating in the previous chapters, we can differentiate between the effect of reading hyperlinked text and the added effect of the decision making task of clicking and navigating.

From the previous work in this thesis, we predicted that we would observe similar findings to those in Chapter 4 where we found differences, in eye movement measures, between reading for comprehension and skim reading. When skim reading, we found reduced skipping of linked words. When the linked words were fixated, they were fully processed, evidenced by a word frequency effect, where low frequency words received longer fixations. The presence of this effect indicates the linked words were fully processed as lexical characteristics of the word were impacting upon fixation durations. However, we did not find a word frequency effect for unlinked words when skim reading. This suggested that unlinked words were not fully processed during skim reading and that the reader was not putting the same priority into reading them as they were for the linked words. In Chapter 3, we discussed how readers think of links as important and having links present in the sentence increases the perceived importance of that sentence. So it would make sense that, when skim reading text, the reader will focus on the important parts of the text to develop an efficient strategy to read the text in a reduced time. Links being present in the text is a highly salient cue to where important information might be in the text. We predicted we would observe all the same effects in the current experiment. An additional issue to consider is how the users' eye movements will change when they have to choose and click a link in the page to navigate. We might expect to observe inflated fixation durations for the linked words in total reading times where the reader may spend longer on the linked words to evaluate which link to click to navigate to another page.

Methodology

Participants

32 native English speakers (15 male, 17 female) with an average age of 20.03 years participated in exchange for payment (£9). All had normal or corrected-to-normal vision and no known reading disabilities. None of the participants tested took part in any other Experiments in this thesis.

Apparatus

Identical apparatus to Experiment 1A.

Stimuli

The stimuli in Experiment 4 consisted of edited Wikipedia articles. These were articles created for this experiment and the reader could click the hyperlinks in the page and navigate to other edited Wikipedia articles (Figure 6.1). Participants could follow any hyperlinks they wished to click on and because of this environment, the number of target words observed by each participant varied dependent on the pages they choose to view. Target words were embedded within sentences and inserted into the Wikipedia articles. The Wikipedia pages contained between zero and four target words. All target words were 4-7 characters in length with an average of 5.12 characters and the high/low frequency pairs were matched on word length. The high frequency words had an average log transformed HAL frequency of 9.62 and the low frequency words had an average log transformed HAL frequency of 6.02 (according to the norms collected in the HAL corpus, Burgess & Livesay, 1998). There was a significant difference in frequency between the high and low word frequency stimuli, $t(471)=49.24$, $p<0.001$. For more information about the creation of this experiment see Chapter 5 for the complete methodology.

The experiment took place as four sessions with different starting pages. The first two sessions were read for comprehension and the last two sessions were skim read. The participant was instructed to read for comprehension or to skim read the text (dependent on the session) and to then choose any hyperlink they wished to follow. If the participant wished they could also go back as many pages as they wished with a designated back button. Participants could use this back button at any

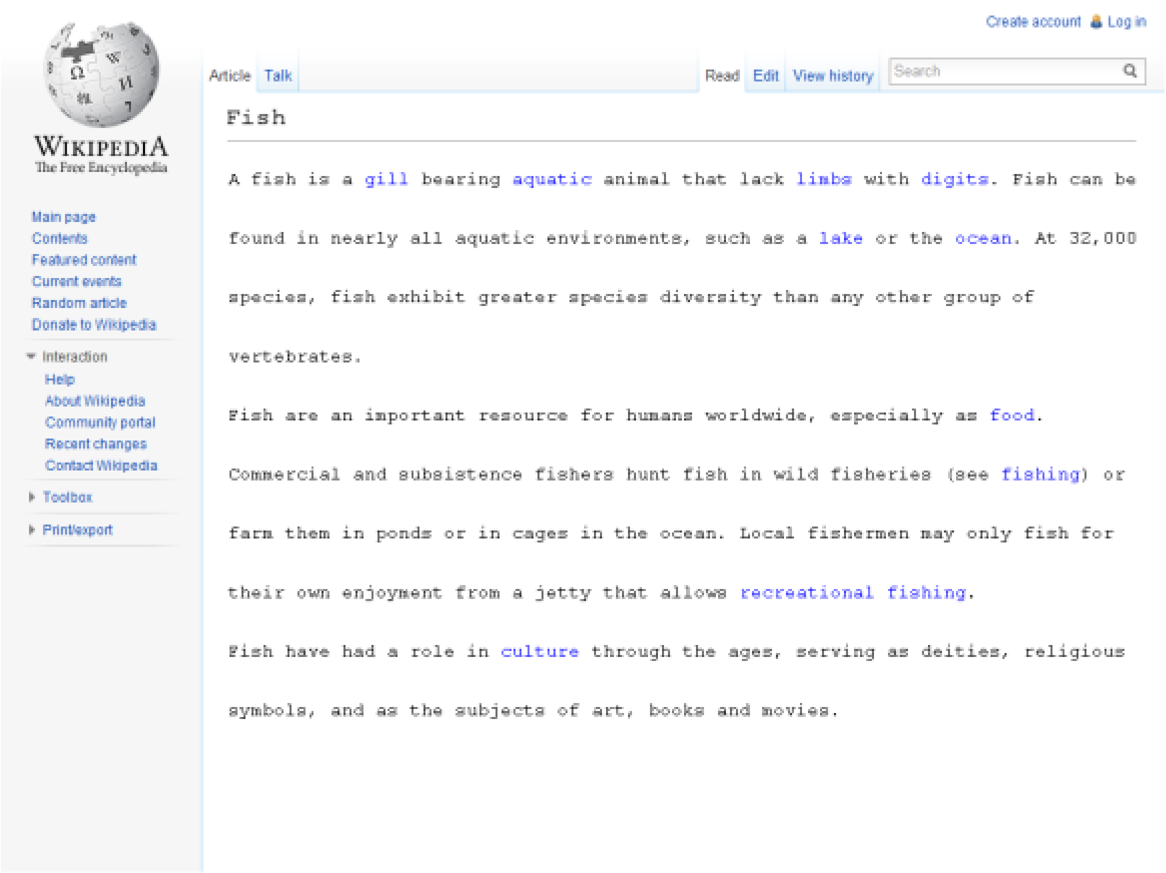


Figure 6.1 Example stimulus from Experiment 4.

time to re-read a page, follow a different link on a previous page or simply to navigate away from a page that was a dead-end containing no hyperlinks in the text to click. The participant was told they may occasionally have to answer comprehension questions after reading some of the articles so they were encouraged to read all of the text on the page before choosing a link to click. Due to the large structure of this experiment and the fact that participants could choose any hyperlinks they wished to follow, we constrained the experiment by ending each session after the participant had visited ten unique pages in each session.

In total there were 8 conditions in a 2 (Task Type: Normal, Skimming) x 2 (Word Type: Linked, Unlinked) x 2 (Word Frequency: High, Low) within participants' design. The participants were instructed to read the text on the screen either for comprehension or to be skim read. As in Chapter 4, we did not counterbalance the Task Type because the normal reading blocks may have been influenced by first having to skim read. If participants are first asked to skim read, it may become difficult to then read "normally" afterwards and this would affect our data because we would not observe normal reading behaviour for those conditions. Participants were not told they were going to be skim reading until just before that half of the experiment was due to begin, so as not to influence the first part of the experiment which was to be read normally. At a target word level, the target words within these articles were either displayed in blue or black to denote if the word was a hyperlink or not (and would be displayed in purple if that link had been visited). The display was 73 cm from the participant's eye and at this distance three characters equal about 1° of visual angle.

Procedure

Participants were given an information sheet and a verbal description of the experimental procedure and informed that they would be reading passages on a monitor while their eyes were being tracked. They were told to read for comprehension in the first half of the experiment and that they were to respond to comprehension questions presented after randomly chosen trials. The participants' head was stabilised in a head/chin rest to reduce head movements that could adversely affect the quality of the calibration of the eye tracker. At the beginning of each trial the participant had to look at a fixation point on the screen. When the eye tracker registered a stable fixation on the fixation point, the sentence was displayed ensuring that the first fixation fell at the beginning of the text. This is to be certain that the reader is starting at the beginning of the passage and not starting the trial by picking up information from later in the text. When participants finished reading they selected a link within the text that they wished to follow, or pressed a button on the keyboard that they were told corresponds to the "back button" on a browser which would go back to the page they previously visited. Participants could go back as many pages as they wished and could click any hyperlink of the page they were on. Comprehension questions were presented to the participants if that page had a comprehension question attached to it, on average participants were presented with comprehension questions on 45% of trials. The comprehension questions were related to the text on the article that had just been read, were simple and required a true or false response. The comprehension questions were presented to ensure the participants were definitely reading and comprehending the text displayed to them and also to measure the level of comprehension across the tasks. Participants responded to the questions by pressing the appropriate response on the screen with the mouse cursor. The appropriate next page that had been selected by the participant would then subsequently appear. If the participant had visited a page with a comprehension question before, the question was not displayed again. The experiment lasted approximately 90 minutes.

Results

Firstly, before exploring the eye movement data, the comprehension question accuracy needs to be addressed. The comprehension question accuracy was 62% (Reading for comprehension: 66%; Skim reading: 60%). Comparing these scores to the previous experiment in Chapter 4 (Reading for comprehension: 91%; Skim reading: 86%), the current experiment has much lower accuracy scores. This could be interpreted in one of two ways. Either the act of navigating has a serious impact on comprehension or another difference between the current experiment and the one in Chapter 4 has caused this reduction in comprehension. For example, in the Chapter 4 there were four comprehension questions shown after each article. For the current experiment, however, there was only one comprehension question shown after 45% of trials, on average. The reader may have given more attention to the task of reading and navigating and found the comprehension questions an unusual aside to the actual task. This is considered further in the general discussion.

The data were cleaned before the eye tracking analysis and this process was identical to the previous eye tracking experiments in this thesis (resulting in the removal of 5.89% of the total dataset). When calculating the eye movement measures, data that were more than 2.5 standard deviations from the mean for a participant within a specific condition were removed (<1% of dataset). Data loss affected all conditions similarly. The means for all of the eye movement measures for Experiment 4 are listed in Table 6.1.

How does Skim Reading affect the Way we Read Hypertext in a Navigable Environment?

We focused our analysis on target word analysis so that we could compare how individuals read in the clickable hypertext environment compared to the "unclickable" environment in the previous experiments in this thesis. We paid particular attention to the comparison with the data presented in Chapter 4, where we explored the same factors (task, hyperlinking and frequency of the target word) but not in a clickable, navigable environment, compared to the current experiment. In that experiment, when the target word was fixated, in 93.91% of the cases it received only a single fixation. Therefore, we limited the fixation duration analyses to when there was a single fixation on the target word. In the current experiment the target word was fixated only once 54.95% of the time. Due to this fact we include a wider range of fixation measures to explore the data, compared to those completed in Chapter 2.

Additionally, in traditional eye movement and reading research the total time incorporates all re-reading that might occur. However, in the current experiment, when the participants are choosing which hyperlinks to click, they may re-read these links, but not for traditional reasons to re-read the text such as problems with integrating the meaning of words and sentences. In the current experiment, the reader may re-read links in order to select and decide which links to click to navigate away from that article to a new article. Therefore, the total time on the target word not only represents re-reading, a measure of processing difficulty associated with the specific word, and the high-level information that content is linked to the target word, but also the additional time to make the decision to click a hyperlink or not. It is very difficult to pull apart the re-reading time and the decision time in the eye movement measures, but we can assume that traditional reading will take place for the unlinked target words, and any inflated times observed for the linked target words could be due to re-reading and time devoted to processing the high-level information that content is linked to the target word (as in Chapter 2) and the decision making related to clicking (unique to the current Experiment).

A series of linear mixed-effects models (LMMs) using R (2009) were used to examine the eye movement measures which explored the impact of three variables. These three independent variables were included as fixed factors: Task Type (Comprehension, Skimming), Word Type (Linked, Unlinked) and Word Frequency (High, Low). Participants and items were included as random effects variables. A 'maximal' structure was initially specified for the random variables (Barr et al., 2013), but the model failed to converge. If a model did not converge, it was reduced by first removing the random effect correlations and then by removing the interactions between slopes and, finally, by successively removing the random slopes explaining the least variance until the maximal converging model was identified.

For skipping probability, only the intercept was allowed to vary for the participants and the items. For first fixation duration the intercept for items was allowed to vary. The intercept was also allowed to vary for the participants, as were the slopes for Word Frequency and Task Type but without the interaction between the two and correlations also needed to be removed. For single fixation duration the intercept for items needed to be removed for the model to converge. Therefore, only the intercept for the participant variable was allowed to vary, as were the slopes for Word Frequency and Task Type but without the interaction between the two and correlations needed to be removed. For gaze duration, only the intercept was allowed to vary for the items variable. Besides the intercept, the slope obtained for Task Type and Word Frequency and their interaction and the slopes of Word Frequency and Word Type but without the interaction between those two were allowed to vary for the participant variable. The interaction between Task Type and Word Type was removed from the gaze duration model because it did not contribute significantly to the fit of the data. For go past time, the intercept for the items variable had to be removed to allow the model to converge. Therefore, only the intercept was allowed to vary for the items variable. Also, the intercept and the slope for Task Type, Word Frequency and Word Type but without their interactions were allowed to vary for the participant variable. Finally, for total reading time the intercept and the slopes obtained for Task Type, Word Frequency and Word Type were allowed to vary for the participant variable, but the interactions and correlations needed to be removed to allow the model to converge. For the fixed factors, successive differences contrasts were used such that the intercept corresponded to the grand mean and the fixed factor estimate for a categorical factor could be interpreted as the difference between the two conditions. All the models were run on log-transformed fixation durations due to skewness in the raw data. The means are presented in Table 6.1 and the output of the LMM's in Table 6.2.

Table 6.1

Means and Standard Deviations Experiment 4. Standard deviation in parentheses.

Task Type	Word Type / Word Frequency	Skipping Probability (%)	First Fixation Duration (ms)	Single Fixation Duration (ms)	Gaze Duration (ms)	Go-Past Time (ms)	Total Reading Time (ms)
Comprehension	Linked/High	12 (32)	197 (54)	201 (54)	273 (137)	1065 (4215)	542 (389)
	Linked/Low	15 (36)	212 (64)	224 (67)	305 (169)	1728 (4954)	572 (370)
	Unlinked/High	19 (39)	207 (66)	211 (64)	256 (111)	322 (341)	311 (179)
	Unlinked/Low	18 (39)	215 (64)	219 (64)	275 (128)	443 (1320)	322 (192)
Skimming	Linked/High	20 (40)	205 (63)	207 (67)	240 (105)	388 (970)	437 (310)
	Linked/Low	14 (34)	228 (68)	239 (65)	293 (120)	461 (938)	501 (343)
	Unlinked/High	19 (39)	202 (60)	204 (56)	227 (87)	245 (147)	245 (110)
	Unlinked/Low	15 (36)	204 (67)	200 (61)	217 (89)	247 (284)	229 (109)

Table 6.2

Fixed Effect Estimates for Skipping Probability Percentage of the Target Word and the Fixation Times on the Target Word in ms for Experiment 4.

	Skipping Probability			First Fixation Duration (ms)		
	Estimate	Std. Error	*z* value	Estimate	Std. Error	*t* value
Intercept	-2.01	0.17	-12.173 *	5.30	0.02	304.52 *
Word Frequency	-0.08	0.19	-0.392	0.06	0.02	3.07 *
Word Type	0.27	0.19	1.374	-0.02	0.02	-0.91
Task Type	0.10	0.18	0.574	0.00	0.02	-0.17
Word Frequency x Word Type	0.19	0.50	0.381	-0.05	0.06	-0.84
Word Frequency x Task Type	-0.53	0.35	-1.519	0.01	0.04	0.16
Word Type x Task Type	-0.41	0.35	-1.166	-0.10	0.04	-2.73 *
Word Frequency x Word Type x Task Type	0.68	0.70	0.974	-0.04	0.07	-0.58
	Single Fixation Duration (ms)			Gaze Duration (ms)		
	Estimate	Std. Error	*t* value	Estimate	Std. Error	*t* value
Intercept	5.32	0.02	278.02 *	5.48	0.02	240.42 *
Word Frequency	0.07	0.03	2.59 *	0.12	0.03	4.16 *
Word Type	-0.04	0.03	-1.61	-0.11	0.03	-4.06 *
Task Type	-0.02	0.03	-0.6	-0.11	0.02	-4.59 *
Word Frequency x Word Type	-0.12	0.07	-1.66	-0.18	0.08	-2.18 *
Word Frequency x Task Type	-0.01	0.05	-0.16	0.00	0.05	-0.04
Word Type x Task Type	-0.11	0.05	-2.16 *	-0.11	0.05	-2.24 *
Word Frequency x Word Type x Task Type	-0.10	0.10	-1.01			
	Go Past Time (ms)			Total Reading Time (ms)		
	Estimate	Std. Error	*t* value	Estimate	Std. Error	*t* value
Intercept	5.62	0.04	142.56 *	5.76	0.03	172.34 *
Word Frequency	0.19	0.06	2.95 *	0.08	0.04	1.91 .
Word Type	-0.29	0.07	-4.42 *	-0.49	0.04	-11.08 *
Task Type	-0.25	0.06	-4.32 *	-0.21	0.04	-5.84 *
Word Frequency x Word Type	-0.31	0.15	-2.09 *	-0.23	0.12	-1.89 .
Word Frequency x Task Type	-0.08	0.10	-0.79	-0.04	0.07	-0.53
Word Type x Task Type	0.09	0.10	0.87	-0.08	0.07	-1.17
Word Frequency x Word Type x Task Type	-0.09	0.23	-0.37	-0.08	0.14	-0.58

*$z > |1.96|$

*$t > |1.96|$

For skipping probability, there was no effect of any of the fixed factors. Our previous experiment in Chapter 4 showed increased word skipping of unlinked words, especially of unlinked words when skim reading, yet we did not replicate this effect in the current experiment. On average there was less skipping in this experiment than in Chapter 4, which could have caused the differences in the findings (Average skipping probability: Current experiment: 17%; Experiment 3: 55%). The only difference between these two studies is the addition of being able to navigate, indicating that less skipping occurs in passages of text when it is navigable. This is discussed further in the general discussion.

For first fixation duration and single fixation duration, there was a main effect of Word Frequency where low frequency words received longer fixation durations. This was qualified by an interaction between Word Type and Task Type (see Figure 6.2). When reading for comprehension, we observed no difference in fixation times for Word Type (first fixation duration: Estimate=-0.03, SE=0.03, t=-1.25; single fixation duration: Estimate=-0.02, SE=0.04, t=-0.63), but when skim reading, the unlinked words received significantly shorter fixation durations (first fixation duration: Estimate=0.06, SE=0.03, t=2.29; single fixation duration: Estimate=0.08, SE=0.03, t=2.68). This suggested that when skim reading, unlinked words were judged as less important and, therefore, the reader spent less time on them.

For gaze duration, there was a main effect of Word Frequency, where low frequency words received longer fixations. There was a main effect of Task Type, where skim reading resulted in shorter fixation durations. There was also a main effect of Word Type, where unlinked words received shorter fixation durations. These main effects were qualified by a two-way interaction between Word Type and Task Type (see Figure 6.2) which was similar to the two-way interaction we observed in first fixation duration and single fixation duration where there is no significant effect of Word Type when reading for comprehension (Estimate=0.04, SE=0.04, t=0.93), but there was an effect of Word Type when skim reading. The unlinked words received much shorter fixation durations when skim reading (Estimate=0.15, SE=0.04, t=3.92) in comparison to the linked words. This indicates that when skim reading, readers were more likely to skim over the unlinked words compared to the linked words.

For gaze duration, there was also a two-way interaction between Word Frequency and Word Type (see Figure 6.3). There was a frequency effect for linked

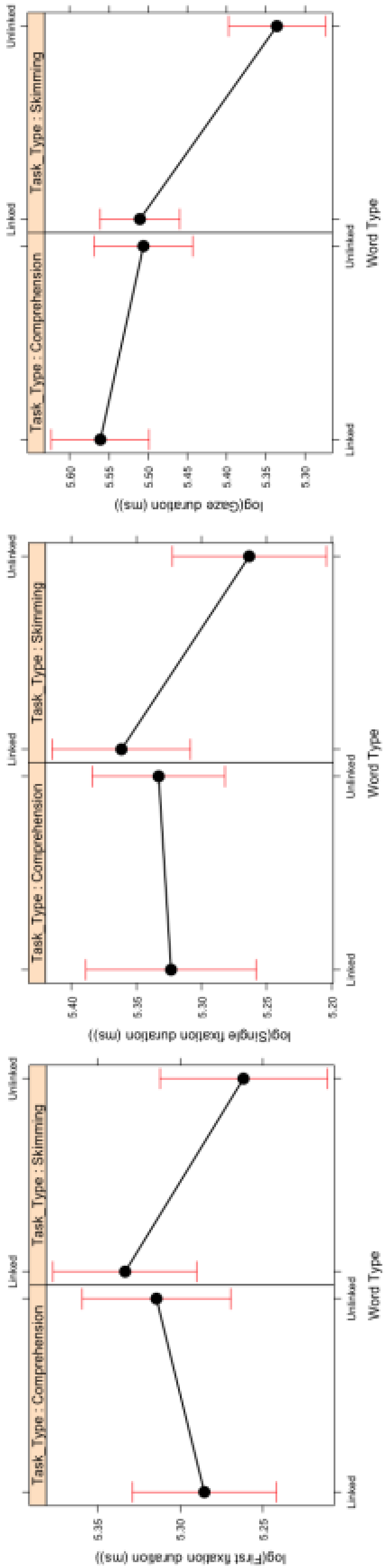


Figure 6.2 Two-way interaction between Word Type and Task Type for Experiment 4. Means and standard error bars for first fixation duration, single fixation duration and gaze duration.

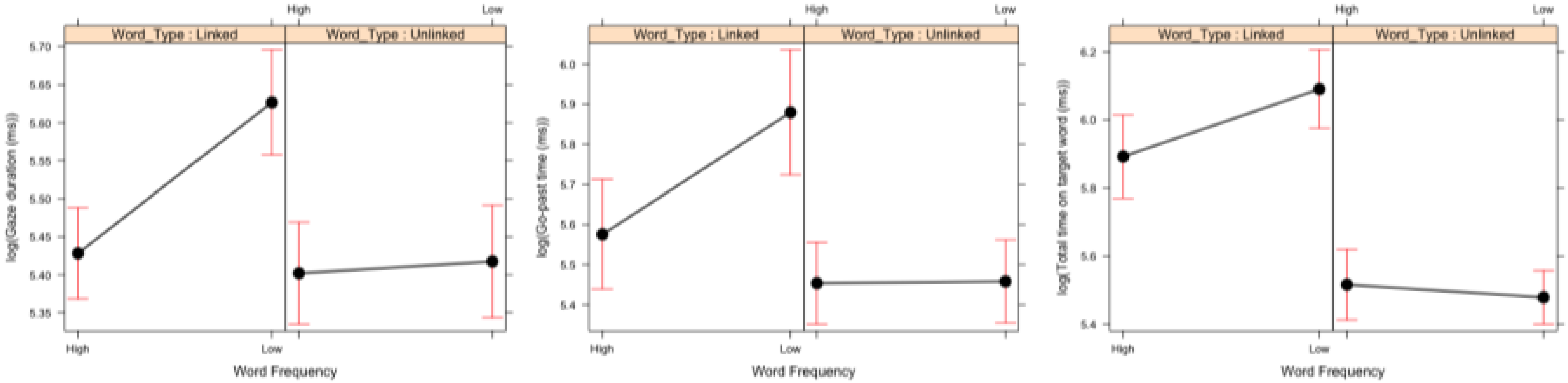


Figure 6.3 Two-way interaction between Word Frequency and Word Type for Experiment 4. Means and standard error bars for gaze duration, go-past time and total time.

words (Estimate=-0.21, SE=0.05, *t*=-4.20), with low frequency words receiving longer gaze durations. This frequency effect, however, was not present for unlinked words (Estimate=-0.06, SE=0.05, *t*=-1.14), suggesting that the unlinked target words were not lexically processed to the same degree as the linked target words.

For go past times, there was a main effect of Task Type, where there were shorter fixation times during skim reading compared to reading for comprehension. There was also main effect of Word Frequency, where low frequency words were fixated for longer than high frequency words and a main effect of Word Type where the unlinked words received shorter fixation durations. These latter 2 main effects were qualified by a two-way interaction between Word Frequency and Word Type (see Figure 6.3). A frequency effect is present when the word is linked (Estimate=-0.25, SE=0.10, *t*=-2.39), with low frequency words receiving longer fixations. However, the frequency effect was not present for the unlinked words (Estimate=0.02, SE=0.07, *t*=0.29). Again, this suggested that when the target word is unlinked it was not being fully lexically processed.

For total reading time on the target word, there was a main effect of Task Type, where the target words received shorter fixations in the skim reading condition compared to reading for comprehension. There was also a main effect of Word Type, where the unlinked words received shorter fixations compared to hyperlinks. And there was a marginal main effect of Word Frequency, with a suggestion of low frequency words receiving longer fixation times than high frequency words. These effects were qualified by a marginal two-way interaction between Word Frequency and Word Type (see Figure 6.3), again suggesting a frequency effect present when the word is linked (Estimate=-0.53, SE=0.09, *t*=-5.62), with low frequency words receiving longer fixations. This frequency effect was not present for the unlinked words (Estimate=-0.06, SE=0.07, *t*=-0.87). Again, this suggested that when the target word is unlinked it was not being fully lexically processed. This is the two-way interaction between Word Frequency and Word Type that was also present in gaze duration and go-past times.

We have already mentioned that in this experimental scenario, total reading time on the target word interest area in all likelihood also represents the decision time when deciding which link to click to navigate to the next page. We see very inflated fixation times for total time for the linked words compared to the unlinked word (513 ms on average for linked words and 277 ms on average for unlinked words) meaning that readers were spending much more time on the linked words after first-pass reading. This was because linked target words were being revisited, as part of the decision making process, in order to choose which link to navigate and click on.

Finally, we measured trial duration (for means for Table 6.3). This is a global measure so we only included the independent variable of Task Type (Comprehension, Skimming) as a fixed factor. The intercept was allowed to vary for the items variable. For the participants' variable the intercept, as well as the slope for Task Type was allowed to vary. We found a main effect of Task Type where the trial duration (See Table 6.4) was significantly longer when it was read for comprehension compared to when it was skim read.

Table 6.3

Means and Standard Deviations for Trial Duration in Experiment 4. Standard deviation in parentheses.

Task Type	Trial Duration (seconds)
Comprehension	39.04 (14.98)
Skimming	21.97 (7.88)

Table 6.4

Fixed Effect Estimates for Trial Duration in seconds for Experiment 4.

	Estimate	Std. Error	*t* value
Intercept	10.21722	0.04337	235.57
Task Type	-0.59884	0.03444	-17.39

*$t > |1.96|$

General Discussion

Experiment 4 is the final step in this thesis of exploring how people read on the Web. In comparison to previous chapters, in the current experiment readers had the ability to freely navigate the hypertext environment. Readers read the text presented to them in edited Wikipedia pages and could then click on links they wanted to navigate to. This is a novel experiment in that it is one of the first to explore eye movements and reading in a Web-like environment that is well-controlled, allows for clicking, and with eye movements recorded with millisecond level accuracy. This experiment is also very important in that it attempts to replicate the previous work in this thesis, but in a 'real' Web environment, where the reader can click and navigate as well as read.

In the early measures we see a reduction in the time spent on unlinked words when skim reading (see Figure 6.2). We also find a frequency effect for the linked words, but not for the unlinked words (see Figure 6.3). This replicates our findings in Experiment 3. We found that there was no frequency effect for the unlinked words suggesting they weren't being fully processed. However, the linked words did show a frequency effect suggesting the reader placed some importance on the linked words (as witnessed in Experiment 2) and thus fully processed them.

The later measures of go past and total time showed inflated fixations on the low frequency linked words compared to the other conditions. If we look at Figure 6.3 we can clearly see that linked, low frequency words have inflated fixation times. This is a replication of the findings from Chapter 2, Experiment 1C. In the current experiment, the low frequency, linked words caused the readers to make multiple, first-pass fixations (as evidenced by the effect on gaze duration), and increased re-reading fixations (as evidenced by the effect on go-past times and total time). The effect of go-past and total time suggests low frequency linked words are considered different to high frequency linked words. When reading passages and having to make a decision about what link to click, the low frequency linked words are fixated for more time, suggesting readers need to evaluate low frequency links more in order to decide whether they are worth navigating to. This is compared to high frequency links that cover more general, easier to understand concepts, requiring less evaluation time by the reader. This further leads us to believe that not only are linked words important to the reader, but they can differ in their importance and ease of evaluation in navigable hypertext. In Chapter 2, Experiment 1C, we also found no impact of links on skipping rates, which we also observe in the current experiment.

Bringing these findings together, it appears that linked words are being processed to a greater degree than unlinked words. This suggests that readers could use links to infer where important information might be in the text and then use the links to guide where they spend their time processing the text. It is of particular interest that we can replicate the findings from Experiment 1C and Experiment 3 in a navigable environment. This is proof that our previous experiments are valid even without the ability to click and navigate. Although we replicate key findings from previous chapters, we will now explore the differences between the previous and current experiments.

In the current experiment, we found that no effects in skipping probability related to the task the reader was completing. Previously, Experiment 3 showed increased word skipping of unlinked words when skim reading (Experiment 3: Unlinked words/Skim reading: 66%; Linked words/Skim reading: 50%), yet we did not replicate this effect in the current experiment. On average, there was considerably less skipping in this experiment than in Experiment 3 (Average skipping probability: Current experiment: 17%; Experiment 3: 55%), which could have caused the differences in our findings. The main difference between these two studies is the addition of being able to navigate and click links, indicating that less skipping occurs in passages of text when it is navigable. And it is important to note that the materials being read were different in both experiments, which could have resulted in one experiment presenting considerable more difficult material than the other experiment. This could have impacted the skipping rates. The stimuli in the current experiment contained a large number of articles and covered a wide range of topics, so perhaps the content was more difficult for some readers. This leads us onto the next issue, the differences in comprehension questions between the two experiments.

In the current study we find that the accuracy for the comprehension questions was much lower to that observed in Chapter 4. The comprehension question accuracy was 62% (Reading for comprehension: 66%; Skim reading: 60%). This is compared to Chapter 4 where the comprehension question accuracy was much higher (Reading for comprehension: 91%; Skim reading: 86%). Again, this could reflect an overall difference in difficulty between the materials presented in the two experiments. The manipulation of target word variables was the same, the main experimental difference was that participants had the ability to click and navigate through hypertext and that there were less comprehension questions in the current experiment. If the reason for reduced accuracy is due to the act of navigating, then this is a serious issue for reading on the Web. Previous research has discussed the impact of increased cognitive load. It would appear that navigation, as a task, might require significant processing resources, to such an extent that high-level comprehension of the text is sacrificed. Previous research has suggested that having to evaluate hyperlinks and navigating a path through them is a demanding task that substantially increases readers' cognitive load (Carr, 2010; DeStefano & LeFevre, 2007; Scharinger, Kammerer & Gerjets, 2015). Alternatively, the reduction of comprehension accuracy could be due to the design of the current experiment. For example, in the experiment in Chapter 4, there were four comprehension questions shown after each article. For the current experiment there was only one comprehension question shown after 45% of trials, on average. The comprehension accuracy might have been very high for the experiment in Chapter 4 because, as they received four questions on each article, they prioritised comprehension as the most important task aspect of that experiment. Therefore, they devoted the majority of processing resources to getting the questions correct. In the current experiment, participants may have prioritised navigation as the most important task, devoting processing resources away from comprehension. Furthermore, the scarcity of questions, relative to Experiment 3 may have results in participants being less motivated to get the questions correct.

To explore these two interpretations, we can investigate whether there is evidence for increased cognitive load. We found no difference in skipping probability in the current experiment, whereas we did witness higher skipping rates for unlinked words when skim reading in Chapter 4. At first there could be a concern that the readers were not engaging in skim reading behaviour, where word skipping is often increased. However, we did find consistent effects in the eye movement measures that are in line with our previous work in Chapter 4 looking at skim reading. From this perspective, it is more likely that having to navigate and click on links increases cognitive load, occupying processing resources, causing a general reduction in comprehension.

Not only does the act of having to make a decision on whether or not to click a link and what link to click on lead to an increase in cognitive load, the disruptive nature of clicking a link could impact on the ability to comprehend the text. Dee-Lucas and Larkin (1995) found that hyperlinks in text distract users by interrupting information processing. Readers may click on hyperlinks in the middle of text content, interrupt their cognitive processing and leave the reader with a fragmented representation of the text. Due to the nature of the Web, websites such as Wikipedia are non-linear environments. Each article is distinct and the articles do not link together in a linear story where the reader can build a complete discourse representation of all of the text. Due to this, the readers may fail to build a complete discourse of each article and this could reduce comprehension accuracy. Summarising, there are number of possible explanations for the difference in overall skipping rates and accuracy on the comprehension questions between Chapter 4 and the current findings. A definitive answer to the question of which one of these factors (or combinations) is responsible for the differences would require a replication of the experiments with or without the possibility to click but with the same materials and with the same comprehension questions.

The final issue to discuss is related to the rather large go-past times for linked words compared to unlinked words, especially when reading for comprehension (linked high: 1065 ms linked low: 1728 ms; unlinked high:322 ms, unlinked low:443 ms). We have discussed how total time also includes the additional time it takes to revisit a link and decide whether to click that link or not. This is perhaps why we see larger total fixation durations for the linked words compared to the unlinked words. However, the go-past time measure also has inflated times for linked words. This is due to occasions when the reader would quickly scan the page to see which links were present and then they would go to the top of the page and read the text from the beginning. On these few occasions the go-past could be very inflated for linked words, especially so when reading for comprehension. When reading for comprehension we found the reader takes significantly longer on the trial (comprehension: 39 seconds; skim reading: 22 seconds) and thus it takes longer to get past the linked target word that was glanced at the beginning of the trial, before the reader engaged in careful reading. We do not see such a large inflation of the linked target words in go-past times for skim reading because when skim reading the reader can reach the bottom of the page and finish the article in a much quicker time. However, even with this additional noise in the dataset, we still clearly observe the consistent effect of inflated fixation times for low frequency, linked words in the later measures (as in Chapter 2).

In summary, we find that being able to navigate has a substantial effect on how readers process text in a hypertext environment. By comparing the findings from Experiment 1C and Experiment 3 to the current experiment (Experiment 4) we can explore the impact navigating and clicking has on factors such as the task the reader is engaged in (reading for comprehension or skim reading), whether a target word is linked or not and the frequency of the target word. We replicate a number of effects seen in Chapter 2 and 4 in an interactive, Web-like environment which gives us a solid foundation to make suggestions of the actual impact of links and reading on the Web. We observe that the reader places importance on linked target words and spend longer on them, especially the low frequency, linked target words. Additionally, by including the ability to click and navigate we find new findings unique to that additional task. We find a reduction in skipping and comprehension question accuracy in the current experiment compared to that observed in Chapter 4. This suggests that having to navigate had a significant impact on the comprehension of the text. This could be due to the additional cognitive load of having to choose a link and navigate. However, we are fully aware that to confirm this claim an additional experiment with the same materials and comprehension questions either with or without the possibility to click would be necessary.

In terms of Web design and the creation of hypertext environments, from the current experiment we see the impact that navigating can have on reading behaviour and comprehension. We witness positive effects of having links in the text in that the linked words are fixated for longer, especially when they are low frequency words, but there is no hindrance of having links in the text on general reading behaviour, as we have shown throughout this thesis. However, once the ability to click and navigate is included in the task, comprehension severely reduces. This suggests that clicking and navigating has a direct impact on reading comprehension. Having links within text on the Web does not seem to have a negative impact on reading behaviour and can even help the reader efficiently find important information. But the act of navigating is a complex task that seems to increase the readers cognitive load. Therefore, when designing webpages we must take this into account. However, further research needs to be conducted to explore navigating and its impact and if it impacts all levels of comprehension and discourse representation or if this task is unique. In the current experiment the comprehension questions where shown after reading an article and this is perhaps an alien task to do. If in future research, we explore comprehension in an even more ecologically valid way, such as comprehension testing after the task of navigating and reading has ended, this would enlighten us to the impact of navigating and clicking on comprehension and tell us if it truly has such a negative impact as that observed in the current experiment.

Chapter 7

General Discussion

The motivation of this thesis was to explore how we read on the Web and what makes it different from reading plain text. We have collected a large amount of data of how people read from the foundations up by starting at a basic level of exploring the impact of a single salient word in a sentence (as a prelude to examining the processing of hyperlinks). Then we expanded our experimental environment and created edited Wikipedia articles to study how individuals read hypertext in a controlled environment. We also rated each sentence in the text to explore the impact of hyperlinks on the rated importance of the sentences and explored the impact of task effects, such as asking individuals to skim read the text. Finally, we created an interactive, but well-controlled, Web environment for participants to read and navigate through. This approach was methodologically novel, and allowed us to fully explore the impact of navigating and having links in the text while reading as well as laying the foundations for future eye tracking investigations of interactive environments. Reading on the Web is a complicated issue to address and by using this stepwise approach to research, it helped us understand the different factors and their individual and combined impact on reading on the Web.

Findings, Implications and Contributions

Chapter 2: Do Hyperlinks have a Negative Influence on Reading When Reading for Comprehension?

Chapter 2 examined the following research questions:

- *Experiment 1A: Does displaying a word in a different colour to the rest of the text change reading behaviour?*
- *Experiment 1B: Does displaying multiple words in a different colour to the rest of the text change reading behaviour?*

- *Experiment 1C: Does a hyperlinked word change reading behaviour? does hyperlinking interact with lexical processing?*

In Chapter 2, Experiments 1A and 1B explored whether salient, coloured words impact reading behaviour. In Experiment 1A, we manipulated the colour of a single target word. In Experiment 1B, we manipulated multiple coloured words in a sentence to investigate if coloured words still have an impact on reading behaviour when there are multiple coloured words embedded in a single sentence. We also manipulated the word frequency to see if colour interacted with word difficulty. In Experiment 1C, we explored whether perceiving the words as hyperlinks influences reading behaviour and if these hyperlinks are processed differently to coloured words in plain text. As such, we moved away from plain text and into an edited Wikipedia environment to simulate a Web experience. The reader could not (yet) click and interact with the pages, only read them, and this was to retain control of the variables being manipulated without already introducing the issue of clicking and navigating which might confound the results. In Experiment 1C, we manipulated whether a target word was displayed as a link or not (it was coloured either in blue, to represent a link, or in black to match the rest of the text) and also its word frequency, whether it was high or low.

To be clear, the rationale behind these three experiments was as follows: In order to examine the impact of links coloured in blue in a hypertext environment, it is important to examine the impact of reading a coloured word in a sentence outside of a hypertext context. This way we can distinguish if any disruption to reading occurred during reading due to the saliency of a coloured word or because the coloured word is a hyperlink.

Experiment 1A demonstrated that a coloured word is less likely to be skipped when it is the only coloured word in the sentence, but that making a word coloured does not negatively impact reading behaviour as measured in fixation durations unless the colour has reduced contrast making it difficult to read, as seen when the target word was grey (or to a lesser extent green).

Experiment 1B replicated Experiment 1A and showed a reduction in skipping of the target word when it was the only coloured word in a sentence. However, this reduction of skipping was not observed when there were multiple coloured words within the sentences. There are two possible suggestions as to why there was a reduction in word skipping for a single coloured word, but not for multiple coloured words in the text. One suggestion is there is an effect of over-signalling, whereby the importance of the coloured words gets reduced when there are too many words coloured randomly and the reader could not utilise the signal as it was meaningless (Goldberg & Kotval, 1999; Lorch et al., 1995). The alternative suggestion is that the saliency of the target word is reduced when there are multiple coloured items present in the sentence (Treisman & Gelade, 1980). A single target word is presumably very salient in a single line of text, but less salient when there are other coloured words in that same line of text.

Experiment 1C demonstrated that there is a difference between processing a word that is coloured during reading a sentence and a linked word in an environment that resembles hypertext. We observed a significant difference in go-past times and total reading times. The linked, low frequency words had longer fixation times in these late eye movement measures which indicated that the reader had difficulty integrating and processing the low frequency word when it was linked. As a result, participants were more likely to reread the preceding content to re-evaluate it. A link on a word implies there is more information behind the link, as we know that links typically can be clicked on and can take us to other, hopefully related, information. When the linked word is a low frequency word the reader may wonder why that word is linked and want to re-evaluate the preceding content. Presumably to make sure that they understood it, or try to decide why it is important.

Although the participants could not click and navigate the links in these experiment, it at least shows that having links shown in blue text does not negatively impact reading behaving. It does increase re-reading when a reader reaches a low frequency, hyperlinked word, but this is not necessarily a negative behaviour. Those links could signal important information that the reader may wish to visit. Because there is no effect on earlier reading measures this suggests that having coloured words in the text does not make reading the text difficult. However, the hyperlinks do signal that the hyperlinked words are special/important in some way, as is also demonstrated in other experiments in this thesis.

Limitations and Future Research

The results presented as part of these three experiments serve as the foundation for future experimentation using the eye movement methodology to examine reading behaviour on the Web. By basing our future research on research on eye movements and reading we can build an understanding of how we read hyperlinked text. Future experimentation in this thesis expands the experimental task towards a more realistic approximation of live Web behaviour. Chapter 6 explored reading behaviour alongside the navigation and decision making elements that hyperlinked text entails. There is also the issue of task effects to consider. The experiments presented here focused only on reading for comprehension, however, we do not always read for comprehension, especially when there is a large amount of information to read in a limited time. Chapter 4 explored this idea in depth by investigating the differences between reading for comprehension and skim reading on the Web. Chapter 2 is the basis for interpreting the results in the following chapters.

Chapter 3: Rating the Importance of Sentences containing Hyperlinks while Reading on the Web

Chapter 3 examined the following research questions:

- *Experiment 2: How does the presence of links affect the importance of unlinked text?*
- *Experiment 2: Are sentences with links rated as more important than sentences without links?*
- *Experiment 2: Does having more links increase importance rating of the sentences?*

In Chapter 3, we explored how readers allocate importance to the different sentences within passages of text (in this case, edited Wikipedia articles). One of the factors this experiment examined was whether the importance rating is modulated by the presence of hyperlinks or whether importance is solely guided by the semantic content of the text.

In order to measure the impact of links we conducted the experiment on two groups of participants. One group of participants viewed passages of text in a Wikipedia environment with links present, the other group saw the same text, but had all of the links removed. The participants had to judge the importance of each of the sentences. By doing this, we were able to separate the impact of hyperlinks from the importance of other textual and content factors. Two additional factors besides the presence or absence of hyperlinks were taken into consideration: length of the sentence and the sentence position on the page.

Chapter 3 demonstrated that there are a number of factors that people use to rate the importance of a sentence. From comparing unlinked sentences in both the experiments with and without links we learnt that when a sentence does not contain links, but it is in a hypertext environment with other sentences that do contain links, it will be rated lower. The sentences containing links are rated as more important and are therefore, by comparison, taking importance away from the sentences without links. This happens regardless of the content of that sentence as we were directly comparing sentences that were visually identical, apart from the presence of hyperlinks in the surrounding sentences for one group of participants.

When we explored just the version of the experiment that contained links and compared sentences that have links versus those that do not have links, we find that those with links are rated higher. This effect was more apparent towards the bottom of the page. This could be because having a link present serves as a boost to importance, especially when a sentence near the bottom of the page. This has less of an impact at the top of the page, where sentences are rated as very important in general. We also explored the impact of the number of links. We found that short sentences with more links that were at the top of page were rated higher, but this drops off as the reader moves down the page. However, long sentences with many links were rated higher, no matter the position of the sentence on the page. This suggests that links do make a sentence more important, especially for long sentences.

Hyperlinks can be used to help signal important information and could be a very clear signal for the reader to help them find the important information. This can be very helpful to readers who do not have the time to fully read and comprehend large pieces of text they encounter. Important sentences containing one or more hyperlinks can be easily recognised due to the saliency of the coloured words and they can prove a very useful typographical cue for the readers.

Limitations and Future Research

Our exploration of the findings in this chapter were fairly exploratory as we do not know if we have captured all of the factors that individuals are using to make importance judgements. In this Chapter we only explore relatively low-level factors (e.g. length of the sentence) that might influence how the importance of a sentence is judged, but we do not take into consideration for instance the actual semantic content of the text. Exploring the influences of the actual text content (e.g. global text difficulty) is outside of the scope of the current thesis, but it is still an important factor that we could be overlooking and which might have a large impact on how important a piece of text is judged. This could be especially true if we consider a readers' task. Some text could be judged as more important than another based on the task at hand. If the reader is searching for a certain piece of information, any text related to that information is going to be judged as more important than a piece of information unrelated to the task, regardless of its sentence length or position on page or how many hyperlinks are present. Previously, Kaakinen and Hyönä (2005) have shown that when the reader had to read from different perspectives, relevant text that was important to the task at hand had increased fixation times and recall in comparison to irrelevant text.

Chapter 4: Skim Reading: Using Adaptive Strategies for Reading on the Web

Chapter 4 examined the following research questions:

- *Experiment 3: How does skim reading affect the reading and comprehension of webpages and what is the influence of hyperlinks on skim reading?*
- *Experiment 3: How does skim reading affect the way we read hypertext?*
- *Experiment 3: How does skim reading affect comprehension?*

In Chapter 4, we explored how links impact on skim reading behaviour and how individuals sample the text and extract information from it in an efficient manner. Skim reading is an efficient way of gaining as much information as possible in the shortest amount of time, while trying not to sacrifice too much comprehension. Links in the text may be used to assist in the strategy of determining which parts of the text contain important information and should be read to gain the desired amount of comprehension.

We manipulated the task the participants engaged in: reading for comprehension or skim reading. The target words within the passages were manipulated to either be black or blue (resembling a hyperlink) and also their lexical difficulty was manipulated by making the target either a highly frequent or a low frequency word. We also tested the participants' comprehension. Between each page of text, the participant was asked comprehension questions which were either related to important or unimportant sentences in the text. These importance ratings were taken from Chapter 3.

Experiment 3 demonstrated that skim reading has a pronounced influence on reading behaviour. As observed in the eye movement measures, fixations were shorter on average for the skim reading condition, replicating the findings of Just and Carpenter (1987). We replicated our findings from Experiment 1C, where we found that links had only a limited impact upon on reading behaviour. We also replicated the finding that when reading for comprehension, links are not a hindrance to reading. Note that we did not replicate the finding from Chapter 2 of increased re-reading of low frequency, linked words. However, our go-past times were not very meaningful in this experiment because re-reading was overall so rare.

When exploring the impact of links on skim reading we found very interesting results. During skim reading, not only were the linked words less likely to be skipped, they were also more likely to be the words that were fully processed. This is shown by the presence of a frequency effect being observed for the linked words when skim reading, but not for the unlinked words. The lack of a frequency effect in the unlinked words during skimming is because they were not fully processed, similar to observations in visual search of text (i.e. searching for a target word in a passage of text) when the reader is also presumed to not lexically processing the words (Rayner & Raney, 1996).

In most of the global sentence reading measures, we found longer first pass gaze durations on the sentence when it was read for comprehension compared to skim reading. This suggests, as we would expect, that the readers spent longer on the sentences when reading for comprehension. We also found that the higher the importance of the sentence, the longer the sentence reading times. In wrap up gaze duration and total reading time, higher importance ratings resulted in longer gaze durations for first-pass and longer total time on the wrap up region. However, in skim reading the importance rating did not seem to have such a large impact as it did when reading for comprehension. This suggests that when skim reading the reader has less opportunity to establish the importance of the sentence and therefore cannot utilise importance as successfully as when reading for comprehension.

From the comprehension accuracy we observed that it declined when the participants were skim reading, probably due to a speed/comprehension trade-off. However, we also found that participants performed marginally better on the comprehension questions regarding the sentences that were rated as more important. This could suggest that the participants were prioritising the more important information effectively. However, this analysis was potentially compromised by ceiling effects in the accuracy scores.

Limitations and Future Research

One limitation of Chapter 4 is due to the inherent nature of reading research. Comprehension questions appear after each trial in order to serve as a check to make sure that participants are actually reading the text. In traditional reading studies these comprehension questions tend to be quite easy and only serve the role of keeping the participant focused on task and motivated so that they actually engage in reading the text. If the questions are too difficult the reader may believe the main task is to give the correct answer on the comprehension questions and instead of engaging in natural reading behaviour, they will perhaps want to re-read the text to make sure they remember all the details. This is not standard reading behaviour and we know from Yarbus (1967) that different tasks will have an impact on our eye movement behaviour. With this is mind the comprehension questions for the present experiment were also quite easy to answer and therefore we observed very high accuracy scores. If the accuracy had been too low for some participants, we would have removed them from the analysis because they may not have been reading the text properly. Due to this fact it makes it difficult for us to observe differences in comprehension accuracy between the important and unimportant questions due to ceiling effects. However, even with this issue we do still show a marginal effect with the important questions showing a higher accuracy than the unimportant sentences. It is difficult to suggest a remedy for this problem because if we were to make the questions more difficult we may not be able to observe natural reading behaviour.

Another concern revolves around the fact that this study is a step forward from a traditional reading experiment towards an ecologically valid reading on the Web study, but it still has some issues in regard to ecological validity. For example, the use of comprehension questions in between trials is questionable in terms of ecological validity because comprehension questions in this form do not occur in normal Web browsing activities. This issue is perhaps more concerning regarding the skim reading task. Those engaged in skim reading online may not be expecting to retain any information if it is not relevant for the users' task. However, because this is one of the first experiments exploring the skim reading of hypertext and the fact that comprehension was a valuable measure to collect, the experiment is still valid. We now have an understanding of skim reading compared to reading for comprehension that we did not have before.

Even though in the current experiments participants only engaged in reading or skim reading behaviour and did not have to make decisions and click any hyperlinks, we obtained significant findings regarding skim reading on the Web which can be built upon in future studies. By taking away the decision making required to navigate through different Web pages, we could therefore focus on how hypertext is read. The results presented here serve as the foundation for future experimentation. Future experimentation will expand our experimental task, which is a simplification of live Web behaviour. In Chapter 6 of this thesis, there is a step towards a more ecologically valid Web environment.

Chapter 5: A Novel Methodology for Creating a Interactive Non-Linear Space

In Chapter 5, we developed new software and a novel methodology for exploring reading behaviour on the Web in a well-controlled environment. This methodology was required to develop an interactive Web environment with good experimental control. Before this thesis there was nothing in existence that could be used for this specific goal and hopefully this methodology and software can be used in the future to create complex hypertext environments to further study how people use and interact with the Web. This methodology was used in Chapter 6.

Limitations and Future Research

Currently, the main limitation is that the creation of stimuli is always going to take a great deal of time. Many articles have to be created to make it feel like a real Web environment that is non-linear and extensive. However, the reusability of this software means the difficulty of creating stimuli may reduce over time.

Chapter 6: Clicking and Reading Hypertext

Chapter 6 examined the following research questions:

- *Experiment 4: How does clicking and navigating hypertext affect reading behaviour and comprehension and how does skim reading interact with this?*

Chapter 6 is the final step in this thesis of exploring how people read on the Web. In comparison to previous chapters, in Chapter 6 readers did have the ability to freely navigate the hypertext environment. Readers could read the text presented to them in edited Wikipedia pages and then click on links they wanted to navigate to. This is a novel experiment in that it is one of the first to explore eye movements and reading in a Web environment that is well-controlled and with eye movements recorded with millisecond level accuracy. This experiment is also very important in that it attempts to replicate the previous work in this thesis, but in a 'real' Web environment, where the reader can click and navigate as well as read.

In the early measures we see a reduction in the time spent on unlinked words when skim reading. We also find a frequency effect for the linked words, but not for the unlinked words, replicating our findings in Chapter 4. We found that there was no frequency effect for the unlinked words suggesting they weren't being fully processed. However, the linked words did show a frequency effect suggesting the reader placed some importance on the linked words (as witnessed in Chapter 3) and thus fully processed them.

In the later measures of go past and total time there were inflated fixations on the low frequency, linked words. This is a replication of the findings from Chapter 2, Experiment 1C. We also found no impact of links on skipping rates, which also replicates Experiment 1C. The effect of go-past and total time suggests low frequency linked words are considered different to high frequency linked words. When reading passages and having to make a decision about what link to click, the low frequency linked words are fixated for more time, suggesting readers need to evaluate low frequency links more in order to decide whether they are worth clicking on in a later stage.

Bringing these findings together, it appears that linked words were being processed to a greater degree than unlinked words. This suggests that readers could use links to infer where important information might be in the text and then use the links to guide where they spend their time processing the text. It is of particular interest that we can replicate the findings from Chapter 2 and 4 in a navigable environment. This is proof that our previous experiments are valid even without the ability to click and navigate. The current experiment adds the additional complexity of being able to click and navigate and thus we have additional findings related to this fact that differ from our previous research. Although we replicate key findings from previous chapters, we will now explore the differences between the previous and current experiments.

In Chapter 6 we find that the accuracy for the comprehension questions was much lower to that observed in Chapter 4. The comprehension question accuracy was 62% compared to Chapter 4 where the comprehension question accuracy was much higher at 89%. The main differences were that the materials being read in Chapter 6 were different from the materials being read in the preceding chapters, that participants had the ability to click and navigate through hypertext, and that there are less comprehension questions in the current experiment. If the reason for reduced accuracy is due to the act of navigating, then this is a serious issue for reading on the Web. It would appear that navigation, as a task, requires significant processing resources, to such an extent that high-level comprehension of the text is sacrificed. Previous research has suggested that having to evaluate hyperlinks and navigating a path through them is a demanding task that substantially increases readers' cognitive load (Carr, 2010; DeStefano & LeFevre, 2007; Scharinger, Kammerer & Gerjets, 2015).

Not only does the act of having to make a decision on whether or not to click a link and what link to click on lead to an increase in cognitive load, the disruptive nature of clicking a link could impact on the ability to comprehend the text. Dee-Lucas and Larkin (1995) found that hyperlinks in text distract users by interrupting information processing. Readers may click on hyperlinks in the middle of text content, interrupt their cognitive processing and leave the reader with a fragmented representation of the text. Due to the nature of the Web, Wikipedia is a non-linear environment. Each article is distinct and the articles do not link together in a linear story where the reader can build a complete discourse representation of all of the text. Due to this the readers may fail to build a complete discourse of each article and this could reduce comprehension accuracy.

However, it is important to note that to fully understand whether the reduced skipping rates and reduced accuracy scores observed in Chapter 6 compared to the previous chapters are due to the extra processing load of having to select links for clicking, it would be necessary to repeat the experiment either allowing to click on the links or not but with the same materials and the same comprehension questions.

In summary, we find that being able to navigate has an effect on how readers process text in a hypertext environment. By comparing the findings from Chapter 2 and 4 to the current experiment we can explore the impact navigating and clicking has on factors such as the task the reader is engaged in (reading for comprehension or skim reading), whether a target word is linked or not and the frequency of the target word. We replicate a number of effects seen in Chapter 2 and 4 in an interactive, Web-like environment that gives us a solid foundation to make suggestions of the actual impact of links and reading on the Web. We observe that the reader places importance on linked target words and spend longer on them, especially the low frequency, linked target words. Additionally, by including the ability to click and navigate we find new findings unique to that additional task. We find a reduction in skipping and comprehension question accuracy in the current experiment compared to that observed in Chapter 4. This suggests that having to navigate had a significant impact on the comprehension of the text. This could be due to the additional cognitive load of having to choose a link and navigate. However, this would need to be firmly established in a follow-up experiment which could prove that the observed differences in skipping rates and accuracy scores were not due to differences in the materials begin read or the comprehension questions that were asked but indeed were due to the additional task of selecting links to click on.

Limitations and Future Research

One limitation of Experiment 4 is due to the nature of reading research, as we already indicated. Comprehension questions appear after around 45% of trials in order to serve as a check to make sure that participants are actually reading the text. If the questions are too difficult, the reader may believe the main task is to give the correct answer on the comprehension questions and, instead of engaging in natural reading behaviour, they will perhaps want to re-read the text to make sure they remember all the details. With this is mind, the comprehension questions for the present experiment were very easy to answer. However, we still observed quite low accuracy compared to traditional reading experiments, and this could have been due to the complexity of the task of reading and then having to click and navigate the Web environment. It is difficult to suggest a remedy for this issue of having comprehension question present in reading experiments. We want to observe natural reading behaviour, encourage comprehension and have an ecologically valid reading experience. It is often difficult to have all of these things. For this experiment, the fact we have lower accuracy then we would normally expect is a very interesting finding and if we did not have comprehension question we would not have been aware of the comprehension problems experienced by the readers. We sacrifice a small amount of ecological validity for theoretical gain. For example, if we wanted to examine what level of comprehension is being hindered by the task of clicking and navigating, in future research we could explore the comprehension of simple facts versus more complex, inference based comprehension.

Another limitation, or at least an idea for future research, is that the reader was not given a specific goal to accomplish during this reading task. The reader was simply instructed to read and navigate freely. This can be said to represent normal reading behaviour on the Web because the reader was told they could click whatever they wanted to. However, browsing a restricted Wikipedia hub of articles without purpose is quite unusual. A lot of Web users would be either searching for specific information or browsing a Website because they are interested in a certain topic, so this browsing behaviour could be said to be still somewhat artificial. However, the present experiment represents the final step of this thesis in understanding how we read hyperlinked text and future research can explore more naturalistic Web reading behaviour now that the factors have been explored in this thesis in a more simplified experimental setting.

Future Directions

The experiments in this thesis have provided some answers to important questions. However, they have also led to many more questions, as is typical for research in fields that have not been thoroughly investigated yet. One stream of research that would be interesting to explore would be how people search for information on the Web. People search for information on the Web every day and although we explore reading for comprehension and skim reading in this thesis, searching for information is another facet of using the Web that deserves exploring. Another stream of research would be the impact of building a coherent mental representation of information in a non-linear Web environment. If someone is browsing multiple pages and gaining information on a topic, what is the impact of being able to navigate and how does it impact on building a representation of the information and their comprehension of the text. This is a much higher level issue than the issues explored in this thesis. Further research into this is necessary to understand how hypertext can impact on mental representations and comprehension.

Also, this thesis only explored the online encyclopaedia, Wikipedia. This environment was chosen because most people have experience with it and the text is displayed in passages with links embedded in the text. In other Web environments there might be different effects and influencing factors to consider. On news websites links are often used either as navigation or as a link to source information. These links may be processed in a different way to those read on Wikipedia. There is also the issue of social media and reading on the Web. Social media feeds such as Facebook and Twitter are quickly updated and displayed as individual posts in a list. This is again a very different format of reading on the Web and has its own interesting theoretical questions. How we read, process and interact with the Web is a large question and current research has only just scratched the surface.

Final Conclusions

The research presented in this thesis shows that there is a need to examine how people read on the Web. Reading on the Web is a complex task and the Web is only becoming larger and more relevant in our everyday life over time. With more people using the Web every year, young and old, we need to understand how people use, read, and interact with the Web in order to build a better Web for the future.

To summarise, in Experiment 1A we found that salient words are less likely to be skipped when they are the only word that is salient in the sentence. However, in Experiment 1B, when multiple words are salient this effect is no longer present due to over signalling or the fact that the words are not as salient when there is more than one of them.

In Experiment 1C we found that participants were more likely to reread the preceding content when they encounter low frequency, linked words. This suggests that the reader may wonder why that word is linked and want to re-evaluate the preceding content, presumably to make sure that they understood it, or try to decide why it is important.

In Experiment 2 we found that having links in sentences reduces the importance of sentences without links in the same text. We also found that sentences with links are rated higher than those with links, especially towards the bottom of the page. We also found that long sentences with many links were rated higher and short sentences at the top of the page were rated higher. All together we witness that links increase the importance of sentences and this can interact with other factors such as position on the page and sentence length.

In Experiment 3 we found that during skim reading the unlinked words were being skipped more often and when they were fixated they were not fully processed, as shown by the lack of frequency effect for the unlinked words when skim reading. This suggests that linked words are thought to be more importance and salient and therefore the reader fixates and processes them. We also found that the higher the importance of the sentence, the longer the sentence reading times. As for the comprehension accuracy, it declined when the participants were skim reading, however, participants performed marginally better when the comprehension questions were about the sentences that were rated as more important. This could suggest that the participants were prioritising the more important information effectively.

In Experiment 4 we replicate Experiment 1C and find increased rereading of the previous content when the reader encounters a low frequency, linked words. We also replicate Experiment 3. We found a frequency effect for the linked words, but not for the unlinked words, suggesting the unlinked words were not being fully processed but the linked words were. This suggests the readers placed an importance on the linked words. We observe the increase in importance for sentences with linked words in Experiment 2 and this would link together with our findings showing that readers place an importance on linked words and will fixate them and fully process them in comparison to unlinked words, which they may consider less important. We also witness a large reduction on comprehension for Experiment 4, that could be due to the readers having to navigate through the text and the act of navigating could increase cognitive load.

This thesis has provided a novel approach to understanding how people read on the Web. It has taken the theories from eye movements and reading research and takes a stepwise approach to move from the controlled experiments on which these theories are based to a more ecologically valid approach, exploring reading on the Web. We have created new methodologies and laid the foundations for future research to continue to understand how we read, process and interact with hypertext on the Web.

References

Anderson, R. C. (1982). Allocation of attention during reading. In A. Flammer & W. Kintsch (Eds.), *Discourse Processing* (pp. 292–313). Amsterdam: North-Holland Publishing Company. doi:http://dx.doi.org/10.1016/S0166-4115(08)62699-0

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439–461. doi:10.3758/BF03214334

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189–208. doi:10.1016/S0021-9924(03)00019-4

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multicomponent model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge University Press.

Balota, D., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–90. doi:10.1016/0010-0285(85)90013-1

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning. II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, *42*(1), 88–110. doi:10.1016/0749-5978(88)90021-0

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Bartlett, F. C. (1932). A theory of remembering. In *Remembering: A study in experimental and social psychology*.

Benway, J. P., & Lane, D. M. (1998). Banner blindness: web searchers often miss “obvious” links. *Interworking*, *1*(3). Retrieved from http://www.internettg.org/newsletter/dec98/banner_blindness.html

Bransford, J., & Johnson, M. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726. doi:10.1016/S0022-5371(72)80006-9

Brett, P. (1998). Using multimedia : A descriptive investigation of incidental language learning. *Computer Assisted Language Learning*, *11*(2), 179–200. doi:10.1076/call.11.2.179.5684

Brooks, L. W., Dansereau, D. F., Spurlin, J. E., & Holley, C. D. (1983). Effects of headings on text processing. *Journal of Educational Psychology*, *75*(2), 292–302. doi:10.1037//0022-0663.75.2.292

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance* (pp. 53–78). Oxford, UK: Oxford University Press.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time n a basic word recognition task: Moving on from Kucera and Francis. *Behavior Research Methods, Instruments, & Computers*, *30*, 272–277. doi:10.3758/BF03200655

Buswell, G. T. (1922). *Fundamental reading habits: a study of their development*. Chicago, USA: The University of Chicago Press.

Campbell, C. S., & Maglio, P. P. (1999). Facilitating navigation in information spaces : Road-signs on the World Wide Web. *International Journal of Human-Computer Studies*, *50*(4), 309–327. doi:10.1006/ijhc.1998.0246

Carr, N. G. (2010). *The Shallows*. New York: W. W. Norton.

Carver, R. P. (1984). Rauding theory predictions of amount comprehended under different purposes and speed reading conditions. *Reading Research Quarterly*, *19*(2), 205–218. doi:10.2307/747363

Cashen, V., & Leicht, K. (1970). Role of the isolation effect in a formal educational setting. *Journal of Educational Psychology*, *61*(6), 484–486. doi:10.1037/h0030286

Charney, D. (1994). The impact of hypertext on processes of reading and writing. In S. . Hilligoss & C. . Selfe (Eds.), *Literacy and computers* (pp. 238–263). New York, NY: Modern Language Association of America.

Conklin, J. (1987). Hypertext : An introduction and survey. *Computer*, *20*(9), 17–41. doi:10.1109/MC.1987.1663693

Crouse, J., & Idstein, P. (1972). Effects of encoding cues on prose learning. *Journal of Educational Psychology*, *83*(4), 309–313. doi:http://dx.doi.org/10.1037/h0033574

Dee-Lucas, D., & Di Vesta, F. J. (1980). Learner-generated organizational aids: Effects on learning from text. *Journal of Educational Psychology*, *72*(3), 304–311. doi:10.1037//0022-0663.72.3.304

Dee-Lucas, D., & Larkin, J. H. (1988). Novice rules for assessing importance in scientific texts. *Journal of Memory and Language*, *27*(3), 288–308. doi:10.1016/0749-596X(88)90056-3

Dee-Lucas, D., & Larkin, J. H. (1995). Learning from electronic texts: Effects of interactive overviews for information access. *Cognition and Instruction*, *13*(3), 431–468. doi:10.1207/s1532690xci1303_4

DeStefano, D., & LeFevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*, 1616–1641. doi:10.1016/j.chb.2005.08.012

Dillon, A., Richardson, J., & McKnight, C. (1993). Space-the final chapter or why physical representations are not semantic intentions. In C. Mcknight, A. Dillon, & J. Richardson (Eds.), *Hypertext. A psychological perspective* (pp. 169–191). Chichester: Ellis Horwood Ltd.

Doctorow, M., Wittrock, M., & Marks, C. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology*, *70*(2), 109–118. doi:10.1037/0022-0663.70.2.109

Drieghe, D. (2008). Foveal processing and word skipping during reading. *Psychonomic Bulletin & Review*, 1–17.

Duggan, G. B., & Payne, S. J. (2009). Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied*, *15*(3), 228–242. doi:10.1037/a0016995

Dyson, M. C., & Haselgrove, M. (2000). The effects of reading speed and reading patterns on the understanding of text read from screen. *Journal of Research in Reading*, *23*(2), 210–223. doi:10.1111/1467-9817.00115

Ehrlich, S., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655. doi:10.1016/S0022-5371(81)90220-6

Elm, W. C., & Woods, D. D. (1985). Getting lost: a case study in interface design. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 927–929). SAGE Publications.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. doi:10.1037/0033-295X.112.4.777

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.

Eylon, B., & Reif, F. (1984). Effects of Knowledge Organisation on Task Perfromance. *Cognition and Instruction*, *1*(1), 5–44. doi:10.1207/s1532690xci0101_2

Fisher, D. F., & Shebilske, W. L. (1985). There is more that meets the eye than the eye mind assumption. In R. Groner, G. McConkie, & C. Menz (Eds.), *Eye movements and human information processing*. Amsterdam: Elsevier.

Fitzsimmons, G., & Drieghe, D. (2011). The influence of number of syllables on word skipping during reading. *Psychonomic Bulletin & Review*, *18*, 736–741. doi:10.3758/s13423-011-0105-x

Fitzsimmons, G., & Drieghe, D. (2013). How Fast can Predictability influence Word Skipping during Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1054–1063. doi:10.1037/a0030909

Fitzsimmons, G., Weal, M. J., & Drieghe, D. (2013). On Measuring the Impact of Hyperlinks on Reading. In *5th Annual ACM Web Science Conference* (pp. 65–74). Paris, France. doi:10.1145/2464464.2464501

Forster, K. I., & Dickinson, R. G. (1976). More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F1,F2,F′, and min F′. *Journal of Verbal Learning and Verbal Behavior*, *15*(2), 135–142.

Fowler, R. L., & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology*, *59*(3), 358–364. doi:10.1037/h0036750

Frase, L., & Schwartz, B. (1979). Typographical cues that facilitate comprehension. *Journal of Educational Psychology*, *71*(2), 197–206. doi:10.1037/0022-0663.71.2.197

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210. doi:10.1016/0010-0285(82)90008-1

Gagl, B. (2016). Blue hypertext is a good design decision: No perceptual disadvantage in reading and successful highlighting of relevant information. *PeerJ*. doi:10.1017/CBO9781107415324.004

Galitz, W. O. (1997). *The essential guide to user interface design: An introduction to GUI design principles and techniques* (3rd ed.). New York, NY: John Wiley & Sons.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *16*(3), 430–45. doi:10.1037//0278-7393.16.3.430

Gilchrist, I. D., Brown, V., & Findlay, J. M. (1997). Saccades without eye movements. *Nature*, *390*, 130–131. doi:10.1038/36478

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

Golding, J., & Fowler, S. (1992). The limited facilitative effect of typographical signals. *Contemporary Educational Psychology*, *13*, 99–113. doi:10.1016/0361-476X(92)90052-Z

Haberlandt, K., & Graesser, A. (1989). Processing of new arguments at clause boundaries. *Memory & Cognition*, *17*(2), 186–193. doi:10.3758/BF03197068

Henderson, J. . M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 417–429. doi:10.1037/0278-7393.16.3.417

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. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier.

Holley, C. D., Dansereau, D. F., Evans, S. H., Collins, K. W., Brooks, L., & Larson, D. (1981). Utilizing intact and embedded headings as processing aids with nonnarrative text. *Contemporary Educational Psychology*, *6*(3), 227–236. doi:10.1016/0361-476X(81)90003-5

Huey, E. B. (1968). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press.

Hyönä, J. (1995a). An eye movement analysis of topic-shift effect during repeated reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(5), 1365–73. doi:10.1037/0278-7393.21.5.A

Hyönä, J. (1995b). Do irregular letter combinations attract readers' attention?

Evidence from fixation locations in words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 68–81. doi:10.1037/0096-1523.21.1.68

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: evidence from adult readers' eye fixation patterns. *Learning and Instruction*, *14*(2), 131–152. doi:10.1016/j.learninstruc.2004.01.001

Hyönä, J., Lorch, R., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*(1), 44–55. doi:10.1037//0022-0663.94.1.44

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception Psychophysics*, *40*(6), 431–439. doi:10.3758/BF03208203

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. doi:10.1016/S0042-6989(99)00163-7

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

Juhasz, B. J., & Pollatsek, A. (2011). Lexical influences on eye movements in reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 873–894). Oxford, UK: Oxford University Press.

Juola, J. F., Ward, N. j., & McNamara, T. (1982). Visual search and reading of rapid serial presentations of letter strings, words, and text. *Journal of Experimental Psychology: General*, *111*, 208–227. doi:10.1037/0096-3445.111.2.208

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. doi:10.1037/0033-295X.87.4.329

Just, M. A., & Carpenter, P. A. (1987). *The Psychology of Reading and Language Comprehension*. Newton, MA: Allyn and Bacon.

Just, M. A., Carpenter, P. A., & Masson, M. E. J. (1982). *What eye fixations tell us about speed reading and skimming. Eye-lab Technical Report Carnegie-Mellon University, Pittsburgh.*

Kaakinen, J. K., & Hyona, J. (2005). Perspective Effects on Expository Text

Comprehension: Evidence From Think-Aloud Protocols, Eyetracking, and Recall. *Discourse Processes*, *40*(3), 239–257. doi:10.1207/s15326950dp4003_4

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(6), 1561–6. doi:10.1037/a0020693

Kim, K. (2001). Implications of user characteristics in information seeking on the World Wide Web. *International Journal of Human-Computer Interaction*, *13*(3), 323–340.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. doi:10.1037/0033-295X.95.2.163

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W., & Dijk, T. Van. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. doi:http://dx.doi.org/10.1037/0033-295X.85.5.363

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73. doi:10.1016/j.jml.2011.03.002

Lemarié, J., Eyrolle, H., & Cellier, J.-M. (2006). Text-signaling devices and their effects on reading and memory processes. *Computers in Human Behavior*, *22*(6), 1096–1115. doi:http://dx.doi.org/10.1016/j.chb.2006.02.013

Lemarié, J., Lorch, R. F., Eyrolle, H., & Virbel, J. (2008). SARA: A Text-Based and Reader-Based Theory of Signaling. *Educational Psychologist*, *43*(1), 27–48. doi:10.1080/00461520701756321

Leyland, L., Kirkby, J. A., Juhasz, B. J., Pollatsek, A., & Liversedge, S. P. (2013). The influence of word shading and word length on eye movements during reading. *The Quarterly Journal of Experimental Psychology*, *66*(3), 471–486. doi:10.1080/17470218.2011.599401

Ling, J., & Van Schaik, P. (2004). The effects of link format and screen location on visual search of web pages. *Ergonomics*, *47*(8), 907–921. doi:10.1080/00140130410001670417

Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation*, *61*(6), 700–712. doi:10.1108/00220410510632040

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14. doi:10.1016/S1364-6613(99)01418-7

Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology*, *75*(3), 402–412. doi:10.1037//0022-0663.75.3.402

Lorch, R. F. (1989). Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review*, *1*(3), 209–234. doi:10.1007/BF01320135

Lorch, R. F., & Chen, A. H. (1986). Effects of number signals on reading and recall. *Journal of Educational Psychology*, *78*(4), 263–270. doi:10.1037//0022-0663.78.4.263

Lorch, R. F., Lemarié, J., & Grant, R. a. (2011). Three Information Functions of Headings: A Test of the SARA Theory of Signaling. *Discourse Processes*, *48*(3), 139–160. doi:10.1080/0163853X.2010.503526

Lorch, R. F., & Lorch, E. P. (1985). Topic structure representation and text recall. *Journal of Educational Psychology*, *77*(2), 137–148. doi:10.1037//0022-0663.77.2.137

Lorch, R. F., & Lorch, E. P. (1996a). Effects of Headings on Text Recall and Summarization. *Contemporary Educational Psychology*, *21*(3), 261–278. doi:10.1006/ceps.1996.0022

Lorch, R. F., & Lorch, E. P. (1996b). Effects of organizational signals on free recall of expository text. *Journal of Educational Psychology*, *88*(1), 38–48. doi:10.1037//0022-0663.88.1.38

Lorch, R. F., Lorch, E. P., & Inman, W. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology*, *85*(2), 281–290. doi:10.1037/0022-0663.85.2.281

Lorch, R. F., Lorch, E. P., & Klusewitz, M. (1995). Effects of typographical cues on reading and recall of text. *Contemporary Educational Psychology*, *20*, 51–64. doi:10.1006/ceps.1995.1003

Lorch, R. F., Lorch, E. P., & Matthews, P. D. (1985). On-line processing of the topic structure of a text. *Journal of Memory and Language*, *24*(3), 350–362.

doi:http://dx.doi.org/10.1016/0749-596X(85)90033-6

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Cambridge: Cambridge University Press.

Masson, M. E. J. (1982). Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 400–417. doi:10.1037//0278-7393.8.5.400

Masson, M. E. J. (1983). Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition*, *11*(3), 262–74. doi:10.3758/BF03196973

Masson, M. E. J. (1986). Comprehension of rapidly presented sentences: The mind is quicker than the eye. *Journal of Memory and Language*, *25*(5), 588–604. doi:10.1016/0749-596X(86)90013-6

Matin, E. (1974). Saccadic suppression: a review and an analysis. *Psychological Bulletin*, *81*(12), 899–917. doi:10.1037/h0037368

Mayer, R., Dyck, J., & Cook, L. (1984). Techniques that help readers build mental models from scientific text: Definitions pretraining and signaling. *Journal of Educational Psychology*, *76*(6), 1089–1105. doi:10.1037//0022-0663.76.6.1089

McConkie, G., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578–586. doi:10.3758/BF03203972

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: 1. The location of initial eye fixations on words. *Vision Research*, *28*(10), 1107–1118. doi:10.1016/0042-6989(88)90137-X

McConkie, G. W., & Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (pp. 137–162). Newark, DE: International Reading Association.

McDonald, S., & Stevenson, R. J. (1996). Disorientation in hypertext: the effects of three text structures on navigation performance. *Applied Ergonomics*, *27*(1), 61–8. doi:10.1016/0003-6870(95)00073-9

McKnight, C., Dillon, A., & Richardson, J. (1990). A comparison of linear and hypertext formats in information retrieval. In R. McAleese & C. Green (Eds.), *HYPERTEXT: State of the Art* (pp. 10–19). Oxford: Intellect.

Meyer, B. J. . (1984). Text dimensions and cognitive processing. In H. Mandl, N. . Stein,

& T. Trabasso (Eds.), *Learning and comprehension of text*. Hillsdale, NJ: Erlbaum.

Miall, D. S., & Dobson, T. (2001). Reading Hypertext and the Experience of Literature. *Journal of Digital Information*, *2*(1), 1–20. Retrieved from http://journals.tdl.org/jodi/index.php/jodi/article/view/35/37

Morkes, J., & Nielsen, J. (1997). Concise, SCANNABLE, and Objective: How to write for the Web. Retrieved from http://www.nngroup.com/articles/concise-scannable-and-objective-how-to-write-for-the-web/

Morris, R. K., Rayner, K., & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 268–281. doi:10.1037/0096-1523.16.2.268

Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *The Quarterly Journal of Experimental Psychology*, *40*(4), 693–718. doi:10.1080/14640748808402294

Nelson, T. H. (1965). A file structure for the complex, the changing, and the indeterminate. In *ACM 20th National Conference* (pp. 84–100). doi:10.1145/800197.806036

Nielsen, J. (1999). When bad design elements become the standard. Retrieved from http://www.nngroup.com/articles/when-bad-design-elements-become-the-standard/

Nielsen, J. (2006). F-shaped pattern for reading web content. Retrieved from http://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/

Nikolova, O. R. (2004). Effects of visible and invisible hyperlinks on vocabulary acquisition and reading comprehension for high-and average-foreign language achievers. *Apprentissage Des Langues et Systèmes d' Information et de Communication*, *07*, 29–53. doi:10.4000/alsic.2279

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology Human Perception and Performance*, *10*(2), 250–257. doi:http://dx.doi.org/10.1037/0096-1523.10.2.250

Office for National Statistics. (2015). *Internet Access 2015 Households and Individuals*. *Office for National Statistics*. London. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2015-08-06

Office for National Statistics. (2016). *Internet users in the UK : 2016*. *Office for National Statistics*. London. Retrieved from https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2016

Pearson, R., & van Schaik, P. (2003). The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task. *International Journal of Human-Computer Studies*, *59*, 327–353. doi:10.1016/S1071-5819(03)00045-4

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 767–79. doi:http://dx.doi.org/10.1037/0096-1523.24.3.767

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, USA: Springer.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675. doi:10.1037/0033-295X.106.4.643

Pollatsek, A., Bolozky, S., Well, A. D., & Rayner, K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and Language*, *14*, 174–180. doi:10.1016/0093-934X(81)90073-0

Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1142–1158. doi:http://dx.doi.org/10.1037/0096-1523.25.4.1142

Potter, M. C., Kroll, J. F., & Harris, C. (1980). Comprehension and memory in rapid sequential reading. In *Attention & Performance VIII* (pp. 395–418).

R Core Development Team. (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–88. doi:10.1007/s00426-008-0173-3

Rayner, K. (1978). Eye movements in reading and information processing.

Psychological Bulletin, *85*(3), 618–60. doi:http://dx.doi.org/10.1037/0033-2909.85.3.618

Rayner, K. (1979). Eye guidance in reading: fixation locations within words. *Perception*, *8*(1), 21–30. doi:10.1068/p080021

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461

Rayner, K., & Bertera, J. H. (1979). Reading Without a Fovea. *Science*, *206*(4417), 468–469. doi:10.1126/science.504987

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception Psychophysics*, *58*(5), 734–747. doi:10.3758/BF03213106

Rayner, K., & McConkie, G. (1976). What guides a reader's eye movements? *Vision Research*, *16*(8), 829–837. doi:10.1016/0042-6989(76)90143-7

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*(2), 245–248. doi:10.3758/BF03212426

Rayner, K., Schotter, E. R., & Masson, M. E. J. (2015). So much to read, so little time: How do we read, and can speed reading help? *The Effects of Brief Mindfulness Intervention on Acute Pain Experience: An Examination of Individual Difference*, *1*, 1–86. doi:10.1017/CBO9781107415324.004

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(5), 1188–1200. doi:10.1037/0096-1523.22.5.1188

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *Journal of Experimental Psychology Human Perception and Performance*, *37*(2), 514–528. doi:10.1017/CBO9781107415324.004

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(6), 1290–301. doi:10.1037/0278-7393.30.6.1290

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. doi:10.3758/BF03214555

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled lettres: there is a cost. *Psychological Science*, *17*(3), 192–3. doi:10.1111/j.1467-9280.2006.01684.x

Reader, W. R., & Payne, S. J. (2007). Allocating time across multiple texts: Sampling and satisficing. *Human–Computer Interaction*, *22*(3), 263–298. doi:10.1080/07370020701493376

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–526. doi:10.1017/S0140525X03000104

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21. doi:10.3758/PBR.16.1.1

Sanchez, R. P., Lorch, E. P., & Lorch, R. F. (2001). Effects of Headings on Text Processing Strategies. *Contemporary Educational Psychology*, *26*(3), 418–428. doi:10.1006/ceps.2000.1056

Scanlan, C. (2003). Writing from the Top Down: Pros and Cons of the Inverted Pyramid. Retrieved May 18, 2016, from http://www.poynter.org/2003/writing-from-the-top-down-pros-and-cons-of-the-inverted-pyramid/12754/

Scharinger, C., Kammerer, Y., & Gerjets, P. (2015). Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. *PLoS ONE*, *10*(6), 1–24. doi:10.1371/journal.pone.0130608

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. doi:10.1037/0033-295X.84.1.1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. doi:10.1037/0033-295X.84.2.127

Shrestha, S., & Lenz, K. (2007). Eye gaze patterns while searching vs. browsing a website. *Usability News*, *9*(1). Retrieved from http://usabilitynews.org/eye-gaze-patterns-while-searching-vs-browsing-a-website/

Simola, J., Kuisma, J., Oörni, A., Uusitalo, L., & Hyönä, J. (2011). The impact of salient advertisements on reading and attention on web pages. *Journal of Experimental Psychology. Applied*, *17*(2), 174–90. doi:10.1037/a0024042

Slatin, J. M. (1990). Reading hypertext: Order and coherence in a new medium. *College English*, *52*(8), 870–883. doi:10.2307/377389

Smith, E. E., & Swinney, D. A. (1992). The role of schemas in reading text: A real-time examination. *Discourse Processes*, *15*(3), 303–316. doi:10.1080/01638539209544814

Spyridakis, J., & Standal, T. (1987). Signals in expository prose: Effects on reading comprehension. *Reading Research Quarterly*, *22*, 285–298. doi:10.2307/747969

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 1–23. doi:10.1167/11.5.5.Introduction

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. doi:10.1037/h0061626

Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science*. John Wiley & Sons.

Treisman, A. M., & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology*, *12*, 97–136. doi:10.1016/0010-0285(80)90005-5

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081–7. doi:10.3758/BF03206376

Van Dijk, T. A. (1980). *Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition*.

van Schaik, P., & Ling, J. (2003). The effect of link colour on information retrieval in educational intranet use. *Computers in Human Behavior*, *19*, 553–564. doi:10.1016/S0747-5632(03)00004-9

von Restorff, H. (1933). Uber die Wirkung von Bereichsbildung im Spurenfeld. *Psychologische Forschung*, *18*, 299–342. doi:10.1007/BF02409636

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, *111*(1), 132–137. doi:10.1016/j.cognition.2008.12.011

White, S. J., & Filik, R. (2009). Reading visually distinctive words in passages of text. In S. P. Liversedge (Ed.), *Abstracts of the Fifteenth European Conference on Eye Movements (ECEM15)* (p. 134). Southampton, UK.

Whiteside, J., Jones, S., Levy, P., & Wixon, D. (1985). User performance with command, menu, and iconic interfaces. *ACM SIGCHI Bulletin*, (April), 185–191.

Wilhite, S. (1986). The relationship of headings, questions, and locus of control to multiple-choice test performance. *Journal of Literacy Research*, *18*(1), 23–40. doi:10.1080/10862968609547554

Wochna, K. L., & Juhasz, B. J. (2013). Context length and reading novel words: An eye-movement investigation. *British Journal of Psychology*, *103*(3), 347–363. doi:10.1111/j.2044-8295.2012.02127.x

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